

23

ISSN 1990-9330

Ю.П. Холюшкин, В.С. Костин,
Е.Е. Витяев, В.В. Мартынович,
С.К. Канн, С.А.Федоров,
А.Н. Нуртдинов,
Ю.А. Загоруйко, О.И. Боровикова

ОЧЕРКИ ПО НАУКОВЕДЕНИЮ (юбилейный выпуск)



Информационные технологии
в гуманитарных исследованиях

2018

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ГОСУДАРСТВЕННАЯ ПУБЛИЧНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ
БИБЛИОТЕКА СО РАН

Ю.П. Холюшкин, В.С. Костин, Е.Е. Витяев, В.В. Мартынович, С.К. Канн,
С.А.Федоров, А.Н. Нуртдинов, Ю.А. Загоруйко, О.И. Боровикова

ОЧЕРКИ ПО НАУКОВЕДЕНИЮ:
Юбилейный выпуск
Монография в журнале

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ГУМАНИТАРНЫХ
ИССЛЕДОВАНИЯХ

Выпуск 23

Новосибирск
2018

УДК 004.9 + 902 + 930.1 + 303

ББК 63.4 + 63.5

Главный редактор

академик РАЕН, д.и.н. Ю.П. Холюшкин

Заместитель главного редактора д.ф.-м.н. Е.Е. Витяев (ИМ СО РАН)

Ответственный секретарь: В.С. Костин (ИЭОПП СО РАН, Новосибирск)

Редколлегия:

Чл.-корр. РАН, академик РАЕН, д.и.н., профессор Л.И. Бородкин (МГУ Москва), д.и.н., профессор В.Н. Владимиров (АГУ, Барнаул), д.и.н. И.М. Гарскова (МГУ, Москва), д.т.н. О.Л. Жижимов (ИВТ СО РАН, Новосибирск), д.и.н. И.В. Журбин (Физико-технический институт УрО РАН, Ижевск), к.т.н. Ю.А. Загорюлько (ИСИ СО РАН, Новосибирск), к.и.н. С.К. Канн (ГПНТБ СО РАН), к.т.н. Н.А. Мазов (ИНГГ СО РАН), д.ф.-м.н., профессор А.Г. Марчук (ИСИ СО РАН, Новосибирск), д.т.н. В.В. Москвичев (ИВМ СО РАН, Красноярск), чл.-корр. РАЕН, д.и.н. А.Н. Садовой (Институт угля и углехимии СО РАН, Кемерово), чл.-корр. РАН, д.ф.-м.н., профессор А.М. Федотов (ИВТ СО РАН, Новосибирск), чл.-корр. РАЕН д.и.н., профессор Ю.Л. Щапова (МГУ, Москва).

Информационные технологии в гуманитарных исследованиях.

И 74. Выпуск 23: Ю.П. Холюшкин, В.С. Костин, Е.Е. Витяев, В.В. Мартынович, С.К. Канн, С.А.Федоров, А.Н. Нуртдинов, Ю.А. Загорюлько, О.И. Боровикова.

Очерки по науковедению: юбилейный выпуск. Новосибирск: ГПНТБ СО РАН, 2018. 118 с.

ISSN 1990-9330

За 20-летний период издания "Информационных технологий в гуманитарных исследованиях" вопросам археологической теории и гуманитарной информатики посвящено 246 статей и 5 монографических изданий.

Наибольший удельный вес и первую позицию по значимости имеют статьи, посвященные разработке и применению математических и статистических методов в археологии. Данному направлению всегда уделялось достаточно большое внимание.

Как отметил авторитетнейший специалист в области исторической информатики, чл.-корр. РАН, профессор Л.И.Бородкин, авторы статей сборника успешно развивают потенциал новосибирской школы, получившей широкую известность в области применения современных методов многомерного статистического анализа в гуманитарных исследованиях.

Им же отмечено, что статьи издания отражают передовые позиции сибирских ученых в разработке тематических электронных ресурсов по археологии, этнологии и истории, создании электронных библиотек и обеспечении доступа к ним с использованием метаданных.

Юбилейный выпуск рассчитан на археологов, историков, библиотечных работников и на широкий круг исследователей, интересующихся информационными технологиями в гуманитарных науках.

УДК 004.9 + 902 + 930.1 + 303

ББК 63.4 + 63.5

© Ю.П. Холюшкин, 2018

© В.С. Костин, 2018

© Е.Е. Витяев, 2018

© С.К. Канн, 2018

© Ю.А. Загорюлько, 2018

ISSN 1990-9330

СОДЕРЖАНИЕ

ОТ РЕДАКТОРА	4
ГЛАВА I. ТЕМАТИКА СТАТИСТИЧЕСКОГО И ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В СБОРНИКЕ "ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ"	7
1.1. Ю.П. Холюшкин, В.С. Костин. Предтечи периодического издания "Информационные технологии в гуманитарных исследованиях".	7
1.2. Ю.П. Холюшкин, В.С. Костин. Программа создания статистических методов на страницах издания "Информационные технологии в гуманитарных исследований"	14
1.3. Ю.П. Холюшкин, В.С. Костин. Концептуальные подходы к созданию статистического пакета для археологов, отражённые на страницах ИТГИ.	25
1.4. Ю.П. Холюшкин, В.С. Костин, Е.Е. Витяев, В.В. Мартынович. Описание последовательностей применения методов (стратегий) интеллектуального анализа данных для решения задач археологии.	34
ГЛАВА II. ИНФОРМАЦИОННЫЕ РЕСУРСЫ В СОДЕРЖАНИИ СБОРНИКА "ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ"	52
2.0. Ю.П. Холюшкин, В.С. Костин. Проблемы автоматизации информационных потоков в археологии	52
2.1. Ю.П. Холюшкин, В.С. Костин. Введение	52
2.2. Ю.П. Холюшкин, В.С. Костин. Система электронного управленческого документооборота (СЭУД)	53
2.3. Ю.П. Холюшкин, В.С. Костин. Система электронного научного документооборота (СЭНД)	54
2.4. В.С. Костин, А.В. Постнов. Полевые отчеты в ИАЭТ СО РАН	57
2.5. Ю.П. Холюшкин, В.С. Костин, В.В. Воробьев. Научный отчет	58
2.6. Ю.П. Холюшкин, В.С. Костин, С.А. Федоров. Электронная библиотека.	60
2.7. Ю.П. Холюшкин, В.С. Костин. Сайт для поддержки, организации и проведения научных конференций	62
2.8. Ю.П. Холюшкин, В.С. Костин, М.Ю. Ильиных. Виртуальный музей	64
2.9. Ю.П. Холюшкин, В.С. Костин, В.М. Семёнов, А.Ю. Подчасов. Виртуальный VRML-музей Института археологии и этнографии СО РАН.	71
2.10. А.Н. Нуртдинов, Ю.П. Холюшкин, В.С. Костин, ГИС в музейных исследованиях	78
2.11. Ю.А. Загорулько, О.И. Боровикова, Ю.П. Холюшкин. Система знаний информационного портала по археологии и этнографии	84
2.12. Ю.П. Холюшкин, В.С. Костин, Р.В. Рыбинцев. Сибирский книжный археологический портал.	90
ГЛАВА III. АНАЛИТИКА БИБЛИОТЕЧНОГО ВЕБ-ПРОСТРАНСТВА	93
3.1. С.К. Канн. Вебометрия электронного ресурса ИТГИ (2013–2018 гг.)	93
3.2. С.К. Канн. Основная проблематика изучения библиотечного веб-пространства. Вебометрия и веб-аналитика (обзор литературы)	97
ЗАКЛЮЧЕНИЕ	106
ЛИТЕРАТУРА	108

После отмеченного в 2008 году 10-летнего юбилея издания "Информационные технологии в гуманитарных исследованиях" (ИТГИ), обращение к историческому наследию, отраженному в разработках и публикациях сектора археологической теории и информатики (САТИ) Института Археологии и Этнографии СО РАН, становилось доброй традицией.

Однако данный ретроспективный анализ материалов единственного в СО РАН научно-теоретического издания ИТГИ обусловлен не только 20-летним юбилеем журнала. Тому есть, как минимум, две серьезные причины.

Во-первых, материалы научной периодики являются самыми мобильными из всех видов научной информации, а их динамика наиболее точно отражает текущее состояние, исследовательские перспективы и публикационные возможности любой научной области.

Во-вторых, обращение к истории издания порождено нерешенностью проблем в археологической науке. В этой связи необходимость осмыслить и беспристрастно оценить собственное прошлое, выявить то ценное и перспективное, что сделано и накоплено наукой, передовым зарубежным опытом, стало настоятельной потребностью. Извлечь же уроки истории нашего издания необходимо хотя бы потому, что с их учетом возможно обоснованное планирование перспектив дальнейшего научного поиска в наше достаточно сложное время.

Периодическое издание "Информационные технологии в гуманитарных исследованиях" – прямое продолжение издания "Методы реконструкций в археологии", издававшегося с 1985 по 1995 гг.

Однако появление сборников с дублирующими наш сборник названиями, грозило вполне предсказуемой возможностью затеряться среди массы изданий подобного рода.

Все это заставляло подумать о новом названии издания, в котором была бы продолжена реализация идеи: публиковать не только "мудрые" статьи, лишь изредка читаемые теоретиками-археологами, а наоборот – сделать такое издание, которое было бы интересно своими практическими приложениями.

Цель предпринятого исследования — определение и анализ массива публикаций по тематике "Археологическая теория и историческая информатика", представленного на страницах анализируемого издания за период с 1998 по 2018 гг. Опубликованные в издании материалы рассматривались в нескольких аспектах:

1. Выявлялись количественные параметры формирования массива публикаций;
2. Выяснялась проблематика статей;
3. Анализировался авторский состав. В рамках такого рассмотрения материалов издания получены следующие результаты.
4. Подведены итоги двадцатилетней деятельности издания.

За двадцать лет редакцией осуществлено издание 22 выпусков серии, которые выходили практически ежегодно, за исключением 1999 и 2001 гг. Несмотря на имеющиеся финансовые трудности, нам удавалось в некоторые годы (2002-2004 и в 2006) издавать по два выпуска.

При этом большая часть выпусков были опубликованы за счет личных средств редактора и редколлегии. Параллельно материалы сборника были размещены на сайте "Sibirica", а в последующем на сайте Отделения ГПНТБ СО РАН. Благодаря этому материалы сборника стали хорошо известны историкам и археологам России и стран СНГ.

Начиная с 10 выпуска, сборник получил статус серийного издания. При этом была создана авторитетная международная редколлегия.

Не будет преувеличением утверждение о том, что серийное издание "Информационные технологии в гуманитарных исследованиях" представляет собой ответственную трибуну для гуманитариев нашей страны, интересующихся проблемами

информатизации науки. В связи с этим были выявлены некоторые статистические характеристики состава авторского сообщества, а также география распределения авторов по регионам страны. Данные показывают, что научное сообщество журнала составили 168 авторов, опубликовавших 251 статью из 7 регионов страны. 87,25% авторов представляли Новосибирский научный центр. По три автора были представлены города Санкт-Петербург, Томск, Москва. Два автора представляли город Омск, 5 представляли Барнаул и один – Вашингтон. Надо отметить высокий уровень соавторства в публикациях сборника, составляющих в целом 60% от числа публикаций. Остальные исследователи выступали в роли единоличных авторов. Среди них 6 исследователей активно выступали в роли первых и вторых соавторов статей.

Квалификационный уровень авторского сообщества нашего издания достаточно высок: 2 академика РАН, 1 чл-корр. РАН, 4 академика РАЕН, 1 член-корр. РАЕН, 14 докторов наук, 27 кандидатов наук, (в целом 75%). Оставшиеся – ученой степени не имеют. В это число вошли аспиранты, соискатели ученой степени, инженеры-программисты, работники системы образования, выпускники вузов. Среди них 30 человек являлись молодыми специалистами.

За 20-летний период в издании "Информационные технологии в гуманитарных исследованиях" вопросам археологической теории и гуманитарной информатики посвящено 246 статей, и 5 монографическим изданий.

Наибольший удельный вес и первую позицию по значимости имеют статьи посвященные разработке и применению математических и статистических методов в археологии. Данному направлению всегда уделялось достаточно большое внимание.

Как отметил авторитетнейший специалист в области исторической информатики, чл.-корр. РАН, профессор Л.И. Бородкин, авторы статей сборника успешно развивают потенциал новосибирской школы, получившей широкую известность в области применения современных методов многомерного статистического анализа в гуманитарных исследованиях.

Им же отмечены, что статьи издания отражают передовые позиции сибирских ученых в разработке тематических электронных ресурсов по археологии, этнологии и истории, создании электронных библиотек и обеспечении доступа к ним с использованием метаданных.

Действительно, тематика этих разделов достаточно разнообразна. Так вопросам применения картографии и ГИС было посвящено 7 статей, 8 статей опубликовано по теме БД и информационные системы и 4 статьи посвящены созданию библиотечных информационных систем и электронных архивов.

Тематика "Интеллектуальные информационные системы" отражена в 11 статьях. «В публикациях этого цикла информационные системы нового поколения авторами видятся как распределенные системы, опирающиеся на множественные базы "осмысленных" данных, содержащие неспецифические общие данные, неспецифические приватные данные, специфические для информационной системы общие и частные данные, спецификации модели мира, предметной области и задачи. Совместимость данных и описаний должна в этих системах обеспечиваться общей методологией, едиными стандартами» [Марчук, 2006]. На основе этой технологии начато создание портала археологических знаний. Вот что об этой разработке писала компьютерная газета:

При помощи компьютерных технологий специалисты смогли создать в Рунете уникальный археологический ресурс (<http://www.sati.archaeology.nsc.ru/sibirica/>), объединяющий в себе около 4200 объектов (учёные, публикации, справочные системы и другие разделы), между которыми прописаны около 16000 связей по 18 критериям [<http://www.nestor.minsk.by/kg/press/2007/11/2305.html>].

Достаточно значимыми по числу статей представлены разделы, посвященные методологии археологической науки и науковедческим исследованиям, а также палеоастрономии.

Таким образом, проблемно-тематическое обозрение статей позволяет утверждать, что серийное издание "Информационные технологии в гуманитарных исследованиях" рельефно фиксирует "эпицентры" исследовательских интересов археологов, математиков, специалистов в области информатики. В основном данное издание помогает ввести в научный оборот новые данные и способствует эффективному приращению и продвижению вперед научного знания в сфере археологической теории и исторической информатики.

Литература

Марчук А.Г. Подход к построению распределенных информационных систем нового поколения // информационные технологии в гуманитарных исследованиях. Вып.10.- Новосибирск, 2006.

Союз технарей и гуманитариев. Пресс-релиз 23.11.2007 // Компьютерная газета. <http://www.nestor.minsk.by/kg/press/2007/11/2305.html>

1.1. Предтечи периодического издания «Информационные технологии в гуманитарных исследованиях»

Одной из характерных черт современного научного познания является проникновение методов и приёмов технических и технологических средств (математики, информатики и др.) в сферу гуманитарного знания, в частности в археологию. Подобные интеграционные процессы привели сегодня к синтезу и дубликации различных дисциплин (математическая лингвистика, биохимия, математическая биология, социометрия, археологическая и историческая информатика и др.).

На пути своего становления и развития археологическая информатика пережила довольно сложную историю. В ходе этой истории отношение к математическим методам в археологии было различным (от безусловной поддержки в "домашинный" период, до открытого скепсиса и неприятия: зачем это нужно, если и без этого можно обойтись). Порой дело доходило до обвинений в бездумном заимствовании методов "новой археологии", вульгаризации истории и выхолащивании духа исторического исследования [Холюшкин, Воронин, 2005: с.5].

Начиная с 60-х годов XX века немногочисленная группа энтузиастов в СССР, несмотря на острые дискуссии, активно занималась поисками преимуществ использования количественных методов в археологической науке. Одним из главных аргументов в пользу применения информационных подходов была идея о том, что археологу нужно не только выступать интуитивным аналитиком первичной информации археологических источников, но и порождать информацию вторичную (производную, интегрирующую исходные сведения) [Холюшкин, Воронин, 2005: с.5].

В истории становления археологической информатики ИАЭТ СО РАН можно выделить пять периодов:

- 1) эмбриональный период – эра больших ЭВМ (1980-1989);
- 2) неформальный (т.е. на общественных началах) – в эру IBM PC (1990 - 1992);
- 3) формационный (1993 -1995);
- 4) период зрелости (1996 - 2006);
- 5) период упадка и уничтожения (2007 - 2012).

Каждый новый этап был обусловлен рядом обстоятельств и причин:

– объективными и субъективными факторами (связанными с развитием электронно-вычислительной техники и технологий обработки всё более возрастающего потока информации);

– отчётливо выраженными тенденциями к интеграции научного познания;

– попытками некоторых археологов более точно фиксировать тенденции и закономерности изучаемых явлений и процессов.

На первом этапе в Институте истории, филологии и философии СО РАН был сформирован неформальный научный коллектив, ядро которого составляли сотрудники сектора информатики социальных процессов (А.Ф. Фелингер, В.Т. Воронин, Е.Д. Гражданников и Д.В. Малев-Ланецкий) и сектора палеолита (Ю.П. Холюшкин).

Целью деятельности этого неформального коллектива была разработка и реализация комплексного проекта в области информатизации научных исследований в археологии, определившего основные направления деятельности будущего сектора археологической теории и информатики: теоретическое, науковедческое и математико-статистическое.

В рамках этого проекта были предприняты и в значительной степени реализованы масштабные исследования. Так А.Ф. Фелингером и Ю.П. Холюшкиным была развернута разработка единой теории социальных процессов, к сожалению, не завершённая из-за преждевременной кончины А.Ф. Фелингера в 1992 году. Совместными усилиями

Е.Д. Гражданникова и Ю.П. Холюшкина были начаты пионерные разработки системной классификации археологической науки, основных археологических понятий и системной периодизации истории. В.Т. Ворониным была разработана теория динамики социальных процессов.

В области статистического анализа обработки археологических материалов были разработаны алгоритмы и программы кластерного анализа, основанные на применении теории информации К. Шеннона и введенных им понятий энтропии и количества информации. Описание этих методов было включено в две монографии [Фелингер, 1985; Деревянко, Фелингер, Холюшкин, 1989].

В монографии "Методы информатики в археологии каменного века" [Деревянко, Фелингер, Холюшкин, 1989] был впервые в СССР предложен термин "археологическая информатика". «Сутью информатики и её стержнем провозглашалась разработка новой технологии научных исследований, вопросы совершенствования и многократного повышения интеллектуального потенциала личности и общества на основе "технических" составляющих интеллектуальных процессов. И вот здесь "ощущается зарождение" некой новой "синтетической" дисциплины. Важной проблемой этой дисциплины (под которой понимается археологическая информатика) является разработка методов и средств фильтрации процессов интеллектуальной деятельности, выделение в них технической составляющей, её отчуждение, исполнение, а затем и синтезирование результатов с гипотезами, выводами, решениями собственно творческой составляющей интеллектуальной деятельности» [Деревянко, Фелингер, Холюшкин, 1989: с. 5].

Также были начаты исследования количественных и качественных характеристик кадровой составляющей сибирской археологической науки.

Завершился этот период изданием трёх монографий:

Воронин В.Т. Динамика социально-экономического развития: анализ временных лагов. Новосибирск: Наука, 1989;

Деревянко А. П., Фелингер А.Ф., Холюшкин Ю.П. Методы информатики в археологии каменного века. Новосибирск, 1989. 268 с.;

Фелингер А.Ф. Статистические алгоритмы в социологических исследованиях. Новосибирск: Наука, СО АН СССР, 1985.

Кроме того, был организован выпуск серии методологических сборников по археологическим реконструкциям (1985, 1988, 1989).

Второй этап совпал с началом микрокомпьютерной революции в России. Персональные компьютеры позволили археологам значительно расширить сферы применения новых информационных технологий. Новый импульс получили исследования по истории культуры, искусства, всеобщей истории, археологии, музееведения. Появление Интернет способствовало развитию международного сотрудничества и обеспечило доступ к ресурсам глобальной сети Интернет.

В этот период продолжались работы по системной классификации археологии, завершившиеся изданием монографии "Системная классификация социологических и археологических понятий" [Гражданников, Холюшкин, 1990]. Эта монография получила положительную оценку профессора МГУ Ю.Л. Щаповой. Так, при рассмотрении теории фрактальных множеств, она отметила, что «этот раздел математики археологам едва ли известен, если не считать почти забытого опыта новосибирских коллег — Е.Д. Гражданникова и Ю.П. Холюшкина. В системной классификации социологических и археологических понятий они первыми в археологии (первыми — бесспорно, насколько удачно — не об этом речь) пытались применить теорию фракталов, назвав ее странным именем — всеобщий фрагментный закон» [Щапова, 2005].

Продолжается в этот период выпуск серии методических сборников. Так в 1991 году вышел очередной сборник "Методы реконструкций в археологии".

В начале третьего этапа (1993), приказом директора академика А.П. Деревянко, при секторе палеолита была сформирована неструктурная группа информатики (руководитель группы д.и.н. Ю.П. Холушкин).

В состав группы вошли:

а) два участника неформального коллектива (к.э.н. В.Т. Воронин и Д.В. Малев-Ланецкий);

б) коллектив опытных специалистов в области математической статистики под руководством к.т.н. П.С. Ростовцева (В.С. Костин, Н.В. Смирнова, Ю.Г. Корнюхин).

Одним из важных результатов периода 1993-1995 гг. является разработка базовых алгоритмов и программ для персональных ЭВМ, предназначенных для использования в археологических исследованиях. Здесь следует выделить алгоритмы и программы по анализу структуры и типологического группирования, основанной на максимизации потерь остаточной дисперсии с последующим их объединением групп в типы (минимизация потерь остаточной дисперсии) (рис.1). Разработчиками являлись к.т.н. П.С. Ростовцев, В.С. Костин, Н.В. Смирнова, Ю.Г. Корнюхин. Этот программный комплекс, созданный в 1994 году, многократно применялся в анализе данных по палеолитическим комплексам [Ростовцев, Корнюхин, Костин, Смирнова, 1994: с. 59-68].

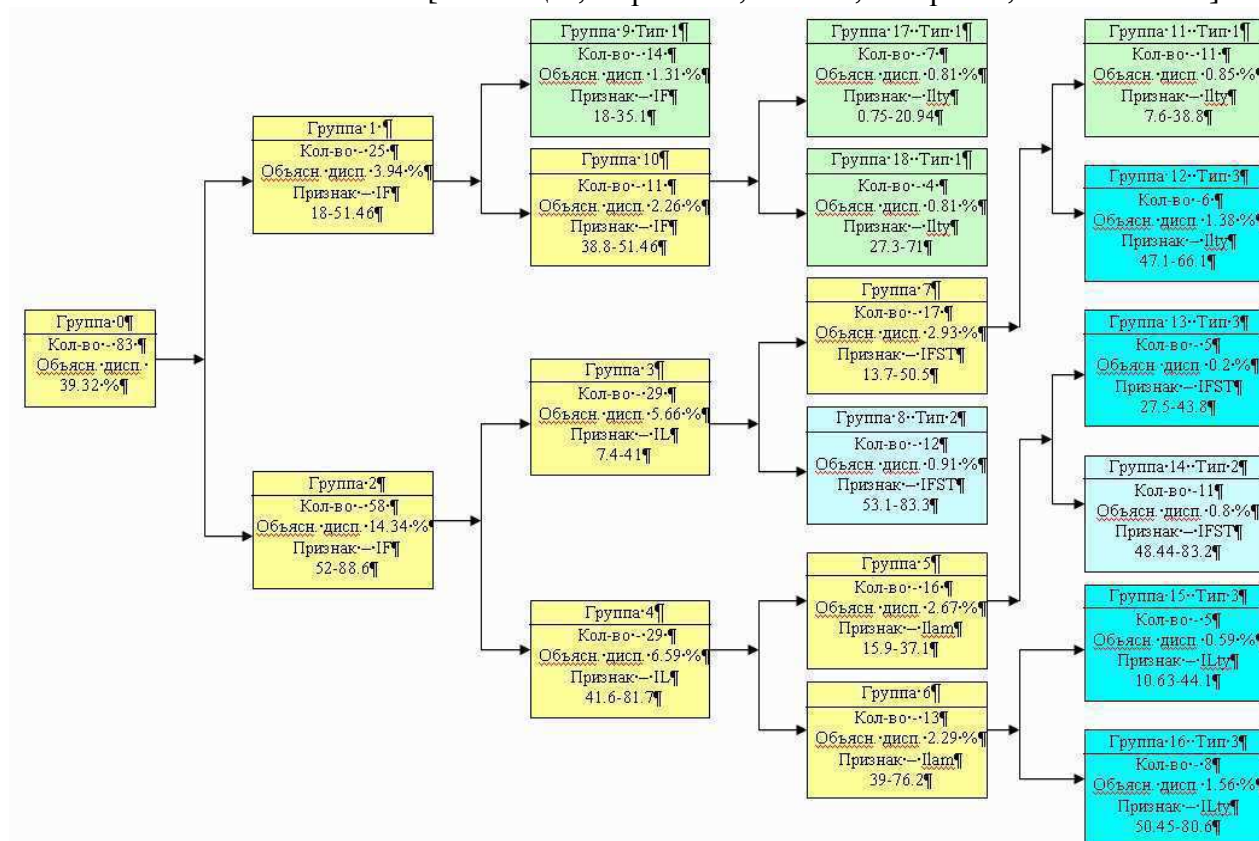


Рис.1. Разбиение множества объектов на кластеры (дихотомия)

в) молодые математики (недавние выпускники Новосибирского госуниверситета Д.Н. Горячев, Е.В. Коптева, В.В. Щипунов, Г.Ю. Жилицкая.

В.В. Щипунов и Е.В. Коптева разработали алгоритмы и программу кластерного анализа, использованные для классификации бифасов [Коптева, Щипунов, 1994: с.55-59].

Г.Ю. Жилицкая (ученица П.С. Ростовцева) разработала вычислительные схемы традиционных алгоритмов статистического анализа данных применительно к потребностям археологических исследований (методы оцифровки, регрессионного, кластерного и факторного анализа, построение доверительных интервалов, методы сравнений эмпирического и теоретического распределений).

Д.В. Горячев разработал компьютерную технологию элиминации масштабной сетки в процессе обработки графических полевых документов (планов и стратиграфических проекций, изображённых на миллиметровой бумаге) [Горячев, 1994: с.100-101].

г) перешедшая в Институт из областного краеведческого музея Л.В. Андерсон разработала БД музея (рис. 2-3). База данных музея была создана на основе лицензионной СУБД Paradox for Windows. В базу данных было введено около 1000 артефактов. Содержательную часть базы подготовила сотрудница музея И.И. Кедрова, программно-техническую часть выполнила инженер-программист сектора Л.В. Андерсон. Среди возможностей, которые получал посетитель, можно отметить, вывод на экран изображений выбранных экспонатов и наиболее значимой информации [Деревянко, Холюшкин, Воронин, Штабной, Воробьев, 1998, рис.3] и просмотр каталога изображений [Деревянко, Холюшкин, Воронин, Штабной, Воробьев, 1998, рис.2].

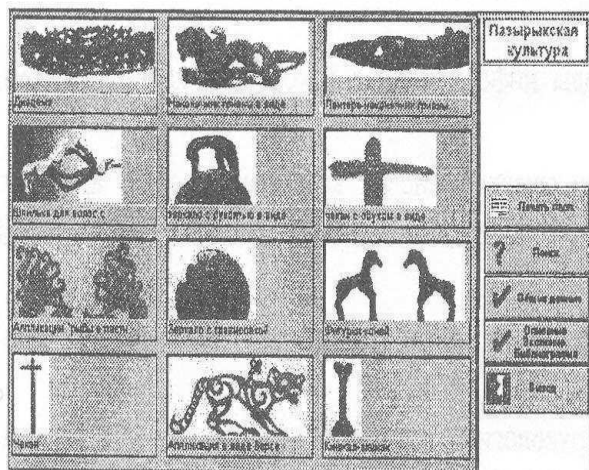


Рис.2. Просмотр коллекции в режиме видеокаталога

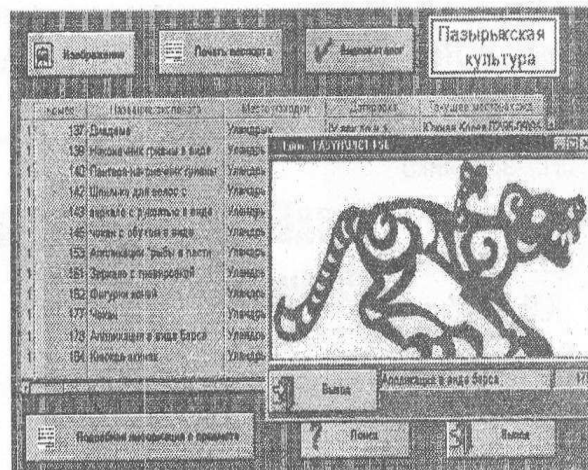


Рис.3. Вывод на экран изображения выбранного экспоната

С приходом в САТИ программистов В.В. Воробьева и К.В. Штабного начаты работы по созданию электронных ресурсов ИАЭТ СО РАН.

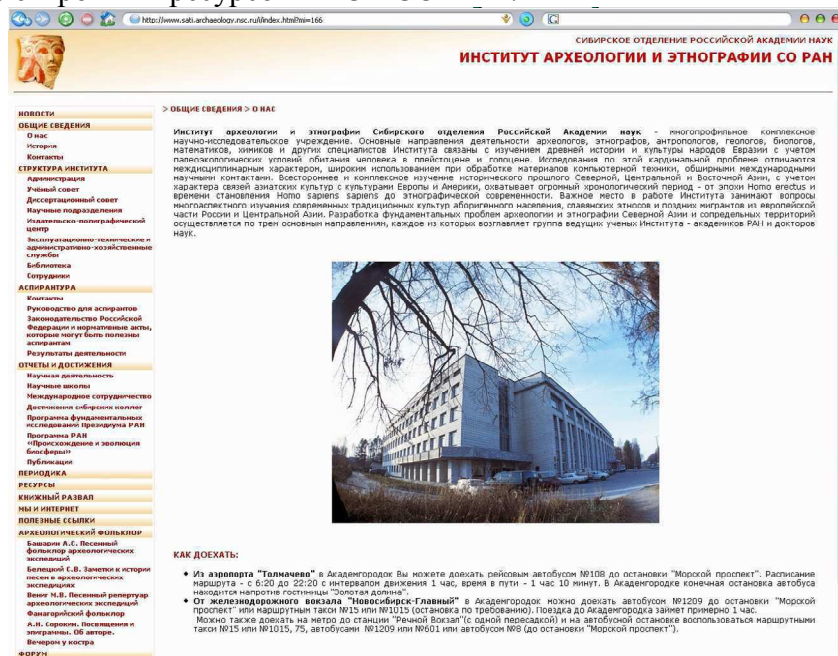


Рис.4. Главная страница сайта ИАЭТ СО РАН (версия 2006 года)

Так в конце 1996 года была создана первая версия сайта ИАЭТ СО РАН, на котором были размещены электронные археологические бюллетени "Компьютеры и археология", формировались узлы ссылок на информационные ресурсы академических центров и 128 университетов мира, археологических сетей ARCHNET и ARCHDATA. Введена первая версия музея «Секреты Сибири» (на английском языке). Был создан опытный вариант виртуального музея "Древнее искусство Приамурья" (на английском и русском языке). В

нём были представлены высококачественные изображения и описания 300 предметов древнего искусства и материальной культуры, найденных на территории Амурского бассейна [Ершов и др., 1998: с. 8-9]. Велась разработка и введение в научный оборот технологии представления научной и культурно-исторической информации в гипертекстовой и мультимедийной форме [Дервянко, Холюшкин, Воронин, Штабной, Воробьёв, 1998, рис.3] (рис. 5).

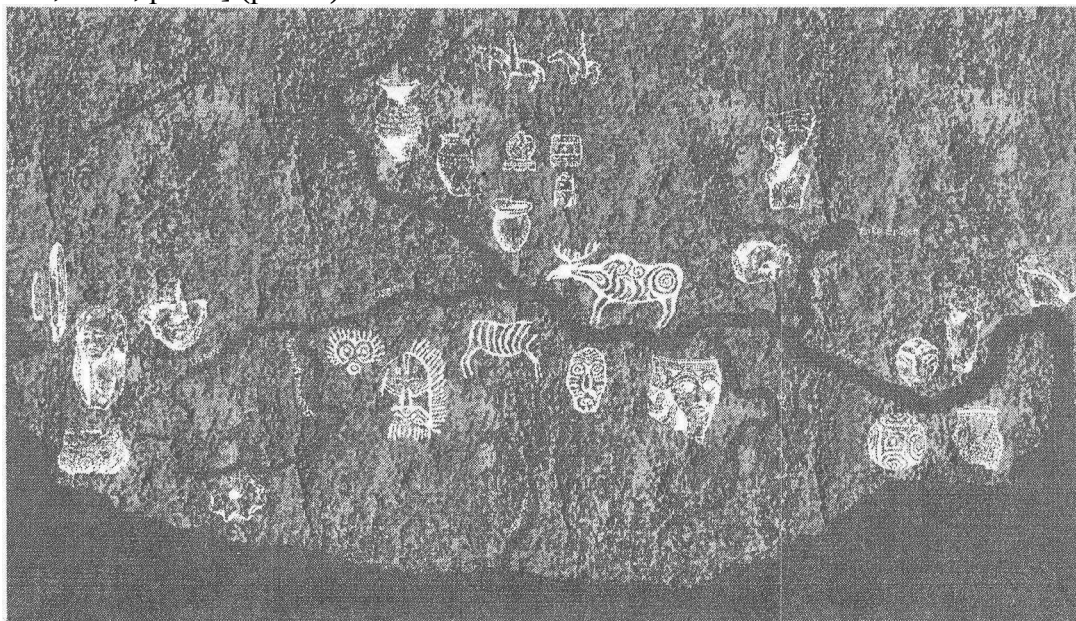


Рис. 5. Навигационная карта музейных экспонатов

В 1995 году в Институте была произведена за счёт грантов группы не только модернизация, но и создана локальная сеть с выделенным сервером, объединившая 9 компьютеров АТ-486. Усилиями Ю.П. Холюшкина и В.Т. Воронина было обеспечено бесплатное для сотрудников Института пользование электронной почтой, первоначально за счёт IREX, а с 1995 года за счёт НГУ [Холюшкин, Воронин, 2005: с.6-7].

Ещё в 1995 году творческий коллектив сектора занял 3 призовое место в тендере по созданию типового варианта локальной сети для гуманитарных институтов и учреждений Академгородка. В качестве награды за это достижение Президиум СО РАН безвозмездно передал на баланс института два мощных по тем временам компьютера АТ 486.

Кроме того сектор при финансовой поддержке РГНФ выполнил комплексное исследование по инициативному проекту "Математический анализ форм и типов неоднородности в археологии палеолита Евразии" (проект № 95-06-17511).

Основные итоги этого периода нашли отражение в публикациях сотрудников группы:

1. *Методология и методика археологических реконструкций*. Новосибирск, 1994.
2. *Методы естественных наук в археологических реконструкциях*. – Новосибирск, 1995.
3. *Математические методы в археологических реконструкциях (методология и методика археологических реконструкций)*. Новосибирск, 1995.

Четвёртый период в истории подразделения начался 4 декабря 1995 года, когда Учёный совет ИАЭТ СО РАН принял решение о преобразовании группы в сектор археологической теории и информатики (САТИ) для разработки новых компьютерных технологий и информационного обеспечения исследований. В 1996 году Ю.П. Холюшкин и В.Т. Воронин выбили в европейском комитете по Интернет (RIPE) солидную (даже избыточную) сетку из 256 IP-адресов. Регистрация локальной сети класса C2 дали Институту возможность прямого доступа к ресурсам Интернет через выделенную линию.

Приобретение ёмкой сетки IP-адресов позволили создать в течение года новую конфигурацию локальной сети в Институте. Для устойчивой работы за счёт грантов сектора было приобретено сетевое программное обеспечение, а также была произведена модернизация компьютеров до класса Pentium и приобретено еще 5 компьютеров этого класса. В 1996 году был выполнен комплекс мероприятий по совершенствованию

управления сетью и обеспечения внешней безопасности (администратор сети – К.В. Штабной).

С 1997 года начались события, которые я назвал борьбой за ресурсы. В результате этой борьбы решением дирекции Института управление сетью было передано в сектор палеолита, однако при этом половина компьютеров сети по-прежнему обслуживалась сотрудниками сектора. Что касается сайта Института, который также был передан в управление сектора палеолита, то инициаторы, по каким-то причинам, отказались от разработок сектора, предпочитая в течение многих лет "радовать" посетителей объявлением "under construction".

Один из тезисов поставленных во главу угла деятельности сектора — сдвиг от воплощения информации к производству знаний, а также создание условий для естественной мотивации использования информационных технологий. Поэтому сектором поддерживаются разнообразные образовательные инициативы и различные образовательные и телекоммуникационные проекты, предполагающие использование распределенных информационных систем.

С этой целью формируется единая система иллюстрированных атласов, энциклопедий, словарей и справочников, включающая в 10575 статей и 1500 иллюстраций. Размещаются в Интернет монографии и статьи сотрудников Института (количество которых приближалось к 500), а также 6 видеофильмов. Кроме того, существовала лента новостей, содержащая около 4000 статей и рассылка по подписке потребителям информации, а также форум. Следующим шагом информатизации образования должно стать создание системы электронных учебных ресурсов и соответствующей технической инфраструктуры.

По теме программы была опубликована монография: Березина И.П., Березин Д.Ю., Холюшкин Ю.П. Немецко-русский и русско-немецкий археологический словарь с общенаучной лексикой. – Новосибирск: РИЦ НГУ, 2003¹ и создан электронный словарь (рис. 6-7).

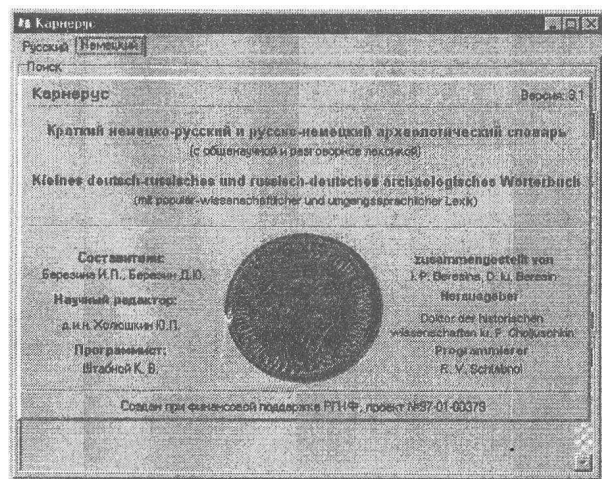


Рис.6. главная страница словаря

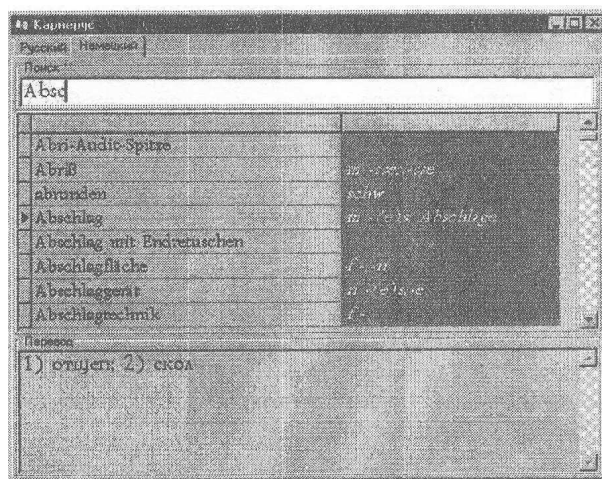


Рис. 7. Рабочее окно словаря.

Созданная система удовлетворяла следующим требованиям [Холюшкин, Березина, Березин, Штабной, 1998: с. 71-74]:

- а) иметь наглядный дружелюбный интерфейс;
- б) обеспечивать возможность редактирования данных, внесённых в словарь;
- в) вносить изменения в содержание словаря (добавление, исправление, удаление терминов);
- г) вести быстрый визуальный поиск слова из словаря.

В период с 1996 по 2007 год в секторе в основном были завершены разработки системной классификации археологии, разработана концепция статистической обработки археологических материалов с целью обнаружения закономерностей, учитывающих наиболее адекватные методы анализа археологических данных, характеризующихся разнотипностью, большим числом пропусков, слабой структурированностью, создано алгоритмическое и программное обеспечение для решения указанных задач. Были разработаны и выложены в Интернет для открытого доступа масштабные информационные ресурсы: информационный центр Института, сайт "Сибирика", сайт

¹ Печатная версия словаря, к сожалению, в русско-немецкой версии не имела артиклей, т.к. Д.Ю. Березин по ошибке вместо отредактированного варианта предоставил в редакцию черновой вариант.

"Гуманитарная паутина", система годовых отчётов, энциклопедические справочники, портал археологических знаний, виртуальный музей "История и культура народов Сибири и Дальнего Востока" и др.

В 1998 году вышел первый выпуск серийного издания "Информационные технологии в гуманитарных исследованиях" (ИТГИ). Это издание, по свидетельству Л.И. Бородкина, ныне чл.корр. РАН и И.М. Гарсковой, ныне доктора исторических наук, сыграло большую роль в процессе адаптации статистических методов и разработки новых алгоритмов и подходов [Гарскова, 2018].

В этом сборнике в период существования сектора археологической теории и информатики в Институте археологии и этнографии СО РАН, одним из постоянных разделов с 1998 г. является раздел "Статистические методы в археологии". В нём публиковались как теоретические и методические материалы, так и результаты прикладных исследований.

Последний, пятый период назван периодом упадка и уничтожения из-за целенаправленного внешнего воздействия. В условиях "реформы науки и образования", когда по её положениям должно было произведено 25% сокращение штатов НИИ, был нанесен удар по двум секторам ИАЭТ СО РАН: сектору истории и археологии стран Зарубежного Востока (старейшему в Институте) и сектору археологической теории и информатики. Часть первого сектора была сокращена, а оставшаяся часть сотрудников переведена в САТИ.

В беседе с В.Е. Ларичевым, по поводу планов дальнейшей совместной работы сектора, я пошутил, напомнив старый анекдот: как господь бог целый год собирал женщин с пониженной социальной ответственностью на океанский лайнер, с единственной целью: утопить всех одновременно. Так и вышло. На следующий год сокращены были все оставшиеся в штате института сотрудники сектора В.Е. Ларичева, а в последующий период сокращены были все программисты сектора археологической теории и информатики. Одного из ведущих программистов, уезжающего в отпуск, заставили подписать заявление на увольнение чуть ли не у трапа самолёта. Таким образом, сектору отрубили руки, оставив на некоторое время только головы. А для того, чтобы дышать было нечем, сократили площади, оставив комнатку для четырёх человек. На мой вопрос: где я смогу разместить д.ф.-м.н. Е.Е. Витяева, В.С. Костина и студентов, последовал ответ М.С. Шунькова: площади института не рассчитаны на совместителей и студентов². Пришлось размещать студентов на площадях ИЭОПП СО РАН и Института математики СО РАН. После сокращения площадей посыпались жалобы директору, что сотрудники сектора не ходят на работу. В 2012 году, накануне полного уничтожения структуры в последнем разговоре с директором я услышал фразу: "у тебя же сотрудники на работу не ходят" (имелись в виду д.ф.-м.н. Е.Е. Витяев и В.С. Костин, д.и.н. В.Е. Ларичев и администратор сети сектора Р.В. Рыбинцев, работавший удалённо). А.П. Деревянко, выслушав мои объяснения, предложил мне выступить на дирекции с докладом о состоянии дел в секторе. Однако мой отчёт не состоялся. Мне результат такого обсуждения был ясен изначально – по окончании срока контракта всех перечисленных совместителей уволили, а сектор прекратил свое существование. Я «поддержал» решение дирекции своим заявлением об увольнении, лишь д.и.н. В.Е. Ларичев³ продолжил работу в отделе палеолита вплоть до своей кончины.

Несмотря на указанные события, издание ИТГИ продолжалось, и это позволило выпустить ещё три монографии (ИТГИ 18, 20, 22) и два выпуска статей ИТГИ (19, 21).

² Однако это не помешало разместить в примыкающей комнате группу д.и.н. Бородавского с большим количеством студентов из НГПУ.

³ Сотрудничество с постоянным автором ИТГИ В.Е. Ларичевым продолжилось после моего увольнения, кроме того мне посчастливилось редактировать его последнюю публикацию, написанную с В.И.Сазоновым [Ларичев, Сазонов, 2016]. Поскольку у меня не было уверенности в том, что книгу издадут, я опубликовал версию своего предисловия [Холушкин, 2014: с. 30-53 с комментариями А.В. Выборнова, 2014: с.53-56 и моим ответом Холушкин, 2014: с. 56-59].

Ниже мы рассмотрим основные результаты коллектива, которые были отражены в ИТГИ на временном отрезке в 20 лет.

1.2. Программа создания статистических методов на страницах издания «Информационные технологии в гуманитарных исследованиях»

В уже упоминавшемся периодическом издании ИТГИ одним из постоянных разделов с 1998 г. являлся раздел "Статистические методы в археологии", в котором публиковались как теоретические и методические материалы, так и результаты прикладных исследований. По данной тематике было опубликовано в ИТГИ 28 публикаций, из которых 10 статей были посвящены методическим вопросам применения методов статистики.

Как было сказано выше, основным центром разработки оригинальных методов математической статистики и анализа данных был сектор археологической теории и информатики ИАЭТ СО РАН [Деревянко, Холюшкин и др., 1995].

Следует подчеркнуть, что попытка установить какую-то общую схему относительно методов и форм статистической обработки и анализа данных, годную во всех случаях, явно несостоятельна. Однако существует определенная "естественная" логика в отношении использования тех или иных методов и эта логика может служить важным подспорьем в проектировании схемы обработки данных в каждом конкретном случае.

Методы обработки и анализа данных можно представить в виде нескольких групп. В каждой отдельной группе находится совокупность однородных или близких по структуре и результатам методов. Группы могут быть упорядочены по степени редукции (свертка, сжатие, обобщение) информации [Деревянко, Фелингер, Холюшкин, 1989: с. 153-154]. В соответствии с таким подходом могут быть рассмотрены следующие группы методов обработки данных:

- вычисление эмпирических распределений по различным классам событий, формулируемых в терминах значений признаков описания объектов статистической совокупности;
- вычисление характеристик распределений (среднее, мода, медиана, дисперсия, энтропия и т.п.);
- анализ зависимости между признаками (корреляция парная, частная, множественная, другие меры и коэффициенты зависимости, регрессионный анализ и т.п.);
- сжатое описание многомерных данных (факторный анализ и многомерное шкалирование);
- выявление структуры многомерных данных (кластерный анализ).

Продвинутый статистический анализ археологических данных предполагает взаимоувязанную обработку данных на всех перечисленных уровнях.

Для данных, сущностную основу которых представляют количественные (числовые) значения, имеется большое разнообразие конкретных алгоритмов для каждого уровня.

Привлечение математической статистики позволяет делать математически точные утверждения при анализе принципиально неточных данных. Это дает нам право рассматривать математическую статистику как универсальный мост через, казалось бы, непреодолимую пропасть, разделяющую эмпирическую и теоретическую науку. Это становится возможным благодаря тому, что математическая статистика использует не формальную, а вероятностную логику. Формальная логика пригодна только для теоретической науки, а вероятностная, оставаясь предельно строгой, как и положено логике, в то же время позволяет работать с недетерминированными, случайными величинами. Вероятностная логика замыкает на себя противоречие между точностью теоретической формы и размытостью эмпирического содержания. Вот какие разделы анализа данных были отражены в выпусках издания ИТГИ:

Предварительный анализ данных

Прежде чем приступать к глубокому статистическому анализу данных, полезно ознакомиться с каждым элементом данных по отдельности. Будем предполагать, что данные представлены таблицей "объект-свойство", то есть каждый объект, будь то отдельная находка или целый слой археологического памятника, характеризуется некоторым набором свойств (признаков), в совокупности составляющих его описание.

Например, свойствами слоя могут быть различные типы артефактов, фаунистические и антропологические остатки, данные ДНК, геологические данные. Для находок, например, каменных артефактов, это может быть материал и количественное и качественное описание артефактов. Каждое свойство обладает своей описательной силой, зависящей от двух моментов: во-первых, насколько разнообразны его значения и, во-вторых, насколько существенные стороны предмета исследования оно отражает. Если для всех объектов рассматриваемое свойство имеет одно и то же значение, то его информативность в точности равна нулю, поскольку анализ данных не позволяет извлекать информацию из констант. Но даже если на всех объектах значения свойства будут различными, это еще не гарантирует его информативности, поскольку не все, что измеряется, имеет какой-либо смысл.

Описательная статистика помогает оценить разброс значений признаков путем построения гистограмм частот, распределений и различных статистик.

В статистике наиболее сильные (т.е. наиболее общие и важные) выводы можно сделать относительно "хороших" переменных. Такими переменными являются количественные, нормально распределенные случайные величины. С помощью описательной статистики исследователь может выяснить, насколько его данные близки к идеалу. Но даже если эти данные далеки от идеала, то в статистике всегда найдутся средства, чтобы сделать обоснованные выводы из их анализа.

Первая задача исследователя, приступающего к статистическому анализу, состоит в определении для каждого признака типа шкалы, в которой он измерен. Для этого достаточно различать три шкалы: номинальную, порядковую и количественную:

- Значение переменной в номинальной шкале – это просто имя. Единственная информация, которую можно извлечь из сравнения двух значений в номинальной шкале – совпадают они или нет.
- Более сильная (информативная) шкала – порядковая. Здесь кроме совпадения, можно определить, какое из значений больше, а какое меньше.
- И, наконец, самая сильная шкала – количественная. Здесь появляется дополнительная информация: насколько велика разность сравниваемых значений. При ее использовании мы от чисто логической информации переходим к информации количественной.

Недостаток информации, содержащейся в значениях, зафиксированной с помощью номинальной и порядковой шкал, приходится компенсировать увеличением числа сравниваемых элементов (увеличением объемов выборки). А при равных объемах выборок статистические выводы для сильной шкалы получаются более определенными.

В задачу предварительного анализа входит проверка корректности данных. Ошибку в данных легче увидеть на графике, чем в таблице. Например, для количественной переменной ошибки (опечатки) часто проявляются в виде выпадающих значений, отстоящих на значительном расстоянии от основной массы значений.

Другой, не менее важной задачей предварительного анализа данных является поиск ответа на вопрос, обладает ли какой-либо (явной или скрытой) структурой анализируемая таблица данных (рис.8). Достаточно простым и эффективным средством является "серый" (или "спектральный") анализ (рис.9). Его суть состоит в том, что анализируемая таблица дополняется графической схемой, которая представляет собой образ таблицы в виде прямоугольника, разделенного на ячейки, подобно клеткам исходной таблицы. При "сером" анализе каждая клетка схемы заполняется (заливается) оттенком серого цвета в зависимости от того, какие значения принимает соответствующий признак для данного объекта. Предварительно промежуток, в который попадают числовые значения всех признаков, разбивается на конечное число равных интервалов. Каждому интервалу сопоставляется определенный оттенок серого цвета по правилу – чем больше значения признаков, которые попадают в данный интервал, тем темнее окрашиваются в серый цвет соответствующие клетки таблицы. Результатом серого анализа является наглядный образ данных, где их структура представлена наиболее отчетливо.

Рис. 8. Орудийные комплексы среднепалеолитических памятников Ближнего и Среднего Востока и Кавказа.

Для выявления скрытой структуры требуется соответствующее преобразование исходной таблицы данных, достигаемое с помощью перестановки строк и (или) столбцов.

Рассмотрим пример данных, взятый из [Деревянко, Холюшкин, Воронин, Костин, 2004: с. 11-15]. На рис. 8 приведена таблица, в которой затененная верхняя строка фиксирует номера оружейных комплексов, а левый затененный столбец – номера археологических памятников, на которых эти оружейные комплексы были найдены. Соответственно на пересечении строк и столбцов указано количество находок. Нулевые ячейки (означающие, что данные пропущены или соответствующие орудия на памятниках не найдены) не заполнены с той целью, чтобы значимые данные были более заметны.

При беглом взгляде на рис. 8 видно, что если эта таблица и имеет какую-либо структуру, то она (эта структура) для наблюдателя представляется скрытой. Поэтому, упорядочивая таблицу перестановкой строк и столбцов, добиваемся того, чтобы эта структура обнаружилась. В наглядной форме эта структура представлена с помощью "серого" анализа (рис. 9).

Для этой цели перед предварительным упорядочиванием сначала были пронормированы значения признаков таблицы так, чтобы каждому исходному значению был поставлено в соответствие значение его ранга в таблице (ранговая статистика). Затем построенная таким образом таблица ранговых статистик была упорядочена перестановкой строк и столбцов. Таким образом, было построено наглядное представление структуры данных, выявленное в результате их "серого анализа" (рис. 9).

Перестановка строк и столбцов при упорядочении таблицы ранговых статистик осуществлялась на использовании следующей идеи. Как правило, хорошо структурированной таблицей является та, в которой не очень часто происходят скачки по величине значений соседних элементов. Поэтому при перестановке строк и столбцов таблицы ранговых статистик эти данные были упорядочены по строкам и соответственно по столбцам таким образом, чтобы суммы расстояний между соседними элементами стали минимальными. Благодаря такой перестановке, строки, соответствующие памятникам, оказались упорядоченными по близости их распределений по артефактам.

Полученная картина показала, что данные таблицы на самом деле обладают некоторой структурой, и, таким образом, имеет смысл с помощью статистических методов анализа исследовать ее более тщательно.

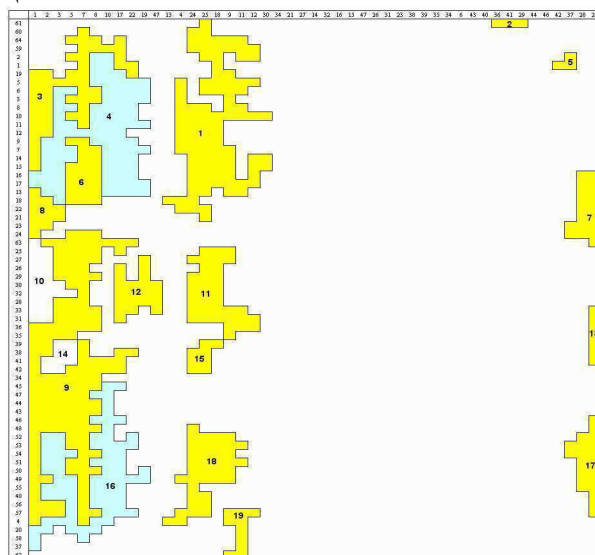


Рис10. Выделение смежных областей.

Для этой цели данные таблицы на рис. 8 (а не их ранговые статистики) были вновь упорядочены по строкам и соответственно по столбцам. Причем, упорядочение данных исходной таблицы выполнялось по такой же схеме, что и упорядочение таблицы ранговых статистик.

На основе этого упорядочения было произведено разбиение матрицы на существенные с точки зрения информативности области.

Было выделено 20 связных областей.

Для визуализации результатов выделения каждая из смежных областей была окрашена единым оттенком серого цвета. В итоге структура данных исходной таблицы предстала еще более отчетливой (Рис. 10, 11).

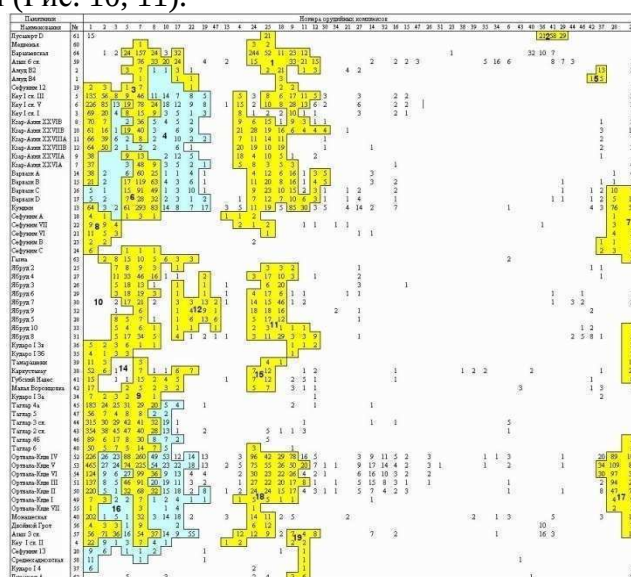


Рис.11. Наполнение смежных областей

Наконец, завершается предварительный анализ данных визуализацией наполненности связных областей числовыми данными (рис.11).

Исследование связей и зависимостей

Следующим шагом анализа в статьях ИТГИ является обнаружение взаимосвязей и зависимостей между признаками. В статистике разработано множество критериев для проверки наличия связей в данных. Но все они построены по одному принципу: в каждом критерии формулируется своя нулевая гипотеза, которая утверждает, что исследуемые признаки являются независимыми случайными величинами, связь между которыми если и проявляется, то исключительно в силу случайного совпадения. Проверка любого критерия начинается с вычисления своей статистики – величины, характеризующей степень отклонения от независимости. Вычисляемая статистика является количественной переменной и подчиняется в условиях выполнения нулевой гипотезы определенному распределению, которое может быть аналитически рассчитано или аппроксимировано программой. Таким образом, значение статистики переводится в так называемую значимость, которая является не чем иным, как вероятностью наблюдения полученного значения этой статистики при выполнении нулевой гипотезы. Если эта вероятность ниже некоторого заранее выбранного порога, например, 5%, то исследователь имеет основания утверждать, что нулевая гипотеза не подтверждается на его данных, из чего с большой вероятностью следует вывод, что между признаками обнаруживается определенная связь.

Поскольку признаки могут быть измерены в любой из трех шкал (номинальной, порядковой и количественной), то для каждого сочетания шкал надо применять свои критерии. Например, если обе переменные измерены в шкале наименований, то можно применять критерий Хи-квадрат, если одна из них – номинальная, а другая – количественная, то можно пользоваться дисперсионным анализом, а если обе количественные, то подойдет корреляция по Пирсону. При сочетании порядковой и количественной переменных приходится огрублять количественную переменную до порядковой и применять методы ранговой корреляции.

Для обычного исследователя, не владеющего в совершенстве методами анализа данных, одна из основных трудностей при работе с пакетами статистических программ заключается в сложности ориентации среди большого числа предлагаемых методов и критериев, которые в основном носят имена их авторов. Чтобы разобраться в том, а для решения каких именно задач применяется тот или иной метод, необходимо самостоятельно изучать специальную литературу.

В данной разработке для преодоления этого препятствия предполагается упорядочить методы по решаемым задачам и условиям (например, сочетаниям шкал признаков). Кроме того, вместо теоретических соображений в пользу того или иного критерия, мы можем привести практические доводы. А точнее – непосредственно на данных пользователя проверять применимость каждого критерия и рекомендовать лишь те, которые покажут свою работоспособность прямо "на глазах изумленной публики". Для проверки работоспособности критерия связи необходимо смоделировать условия нулевой гипотезы об отсутствии связи между переменными. Это можно сделать, разрушив связь между переменными, для чего достаточно перемешать любую из них, то есть переставить ее значения в случайном порядке, не изменяя самих значений. Многократное применение критерия к перемешанным данным должно давать случайные результаты – значимость при этом должна быть равномерно распределена в интервале от нуля до единицы. Если распределение значимости не будет равномерным, значит, критерий на предоставленных данных не работает. Если все критерии откажутся работать, то такие данные подлежат отбраковке как не пригодные к статистическому анализу связей.

Дополнительным сервисом системы может быть проверка чувствительности критериев к связям разного рода – линейным и нелинейным. Для такой проверки необходимо найти способ моделировать связи между переменными путем неслучайного перемешивания.

Методы снижения размерности

К методам снижения размерностей относятся факторный анализ (метод главных компонент) и многомерное шкалирование.

Таблица 1. Факторные нагрузки на типы орудий.

№ типов орудий	Типы орудий	Факторные нагрузки на типы орудий						
		1	2	3	4	5	6	7
7	Продольные скребла	0.880	0.196	0.041	-0.006	0.154	-0.140	0.070
8	Двойные скребла	0.815	-0.209	-0.092	0.055	-0.133	-0.021	0.056
9	Конвергентные скребла	0.671	-0.102	-0.093	-0.045	0.026	0.476	-0.057
25	Зубчатые	-0.050	0.891	0.074	0.027	-0.079	0.003	0.068
24	Выемчатые	-0.096	0.738	-0.146	-0.337	-0.001	-0.083	-0.014
1	Леваллуазские сколы	-0.404	-0.507	-0.459	-0.176	-0.218	0.023	0.145
2	Леваллуазские острия	-0.351	-0.433	-0.229	-0.132	-0.247	-0.132	0.305
18	Резцы	-0.016	0.182	0.838	0.085	-0.116	-0.099	0.036
37	Разные	-0.133	-0.293	0.680	-0.183	0.060	-0.175	-0.289
22	Усеченные отщепы	0.044	-0.040	0.608	-0.024	0.198	-0.008	0.369
19	Проколки	-0.119	0.367	0.432	0.015	-0.287	0.152	0.206
5	Мустьерские острия	0.186	0.004	0.049	0.871	-0.012	0.053	-0.052
3	Леваллуазские ретушированные острия	-0.173	-0.139	-0.088	0.855	-0.067	-0.156	0.049
17	Скребки	-0.027	0.062	0.130	-0.145	0.869	-0.137	-0.023
10	Угловатые скребла	0.112	-0.142	-0.188	0.107	0.673	0.471	-0.037
11	Поперечные скребла	-0.025	0.020	-0.096	-0.092	-0.011	0.885	-0.117
20	Ножи	-0.252	-0.037	-0.141	-0.036	0.185	0.272	-0.654
4	Псевдолеваллуазские острия	-0.158	-0.119	-0.136	-0.259	0.221	-0.178	0.623
28	Сколы, ретушированные со спинки	0.005	-0.142	-0.167	-0.273	0.162	-0.340	-0.613

Эти методы позволяют из многих десятков малоинформативных признаков построить несколько высокоинформативных факторов, содержащих "отжатую" информацию, неравномерно разбросанную по исходным признакам. На примерах можно убедиться, что оба этих метода дают близкие результаты, но метод главных компонент алгоритмически более простой и эффективный, поэтому мы в дальнейшем сосредоточимся на нем.

Факторный анализ, кроме снижения размерности, дает косвенную возможность исследовать связи многих признаков, ибо факторы можно представить в виде линейной комбинации исходных признаков. Те признаки, которые входят с наибольшими

коэффициентами в разложение факторов, образуют группы высокоррелированных признаков (см. табл.1 – пример взят из монографии [Дервянко, Холюшкин, Воронин, Костин, 2004: с. 25]).

Здесь из 47 признаков (типов орудий) предварительно выделены 7 главных компонент, по факторным нагрузкам которых из множества типов орудий были отобраны 19 типов как наиболее значимых. Методом вращения Varimax, при котором максимизируется дисперсия (разброс) факторных нагрузок каждого компонента на исходные переменные, получена матрица нагрузок этих переменных на отобранные главные компоненты (см. табл. 1).

Таким образом, факторный анализ за один проход выделяет все группы связанных переменных. Этот побочный результат данного метода часто оказывается полезен при изучении связей.

Поскольку факторы представляют линейную комбинацию исходных переменных, то полный набор факторов содержит в точности то же количество информации, что и набор исходных признаков.

Сущность метода можно понять, представив себе данные как точки в пространстве с размерностью, равной числу исходных переменных. Точки в этом многомерном пространстве сосредотачиваются в некотором компактном облаке рассеяния. Поиск главных компонент сводится к такому вращению системы координат, при котором вдоль первого фактора наблюдается наибольший разброс точек, вдоль второго – меньше и далее по убыванию.

Но дело в том, что случайные величины всегда содержат шум, вызванный ошибками сбора данных и случайными отклонениями в параметрах объектов. Природа этих ошибок может быть самой разной, но они всегда есть. Значит, и полный набор факторов также содержит шум. Главные компоненты, начиная с фактора номер один, характеризуются наибольшим отношением сигнала к шуму. Чем больше номер фактора, тем меньше полезной информации он содержит. При некотором критическом номере фактора уровень шума становится выше уровня сигнала. Этот и все последующие факторы должны быть отброшены.

Определить количество факторов, которые надо отбросить, достаточно просто. Для этого воспользуемся тем же подходом, что при проверке применимости критериев связи. Действительно, мы можем сформулировать нулевую гипотезу: все исходные переменные независимы друг от друга. Промоделировать выполнение нулевой гипотезы не представляет труда – достаточно перемешать каждую исходную переменную (может быть, кроме одной) и повторить расчет факторов. Все полученные таким образом факторы, начиная с первого, не содержат никакой информации о связях, а только статистический шум, поскольку это гарантировано выполнением условий нулевой гипотезы. Такой статистический эксперимент можно провести многократно. Если теперь сравнить факторы, полученные по исходным данным с теми, что получились в результате экспериментов с перемешиванием, мы сможем определить количество факторов, действительно содержащих больше полезной информации о связях признаков, чем статистического шума.

Сжатие информации с помощью главных компонент является часто подготовительным этапом для структурного анализа, к рассмотрению которого мы и переходим.

Анализ структур

Если анализ связей выявляет признаки, значения которых согласованно изменяются от объекта к объекту, то анализ структур выявляет объекты, на которых согласованы значения определенного набора признаков.

Промежуточное место между анализом связей и анализом структур занимает метод прямого кластерного анализа, разработанный П.С. Ростовцевым [Ростовцев, 1985]. Этот метод позволяет непосредственно на таблице объект-признак выделить, после

переупорядочения строк и столбцов, области неправильной формы, выделяющиеся близкими значениями признаков. Каждая такая область объединяет несколько объектов и несколько признаков в "пятно".

№	Комплексы	IF	IF _{st}	IL	IL _{am}
1	Семиганч	37.10	14.20	20.20	30.70
24	Оби-Рахмат 15-18	-	-	-	35.50
25	Оби-Рахмат 10-14	-	-	-	34.40
9	Чингиз	54.00	29.00	41.10	0 23.00
11	Актогай	48.60	8.60	44.00	6.60
13	Семизбугу В	57.60	15.20	9.80	2.80
12	Семизбугу А	1 50.80	6.00	9.80	2.80
5	Георгиевский Бугор	44.30	8.80	9.50	10.80
18	Кутурбулак	50.70		2.40-	20.00
19	Зирабулак	36.90	27.30	2.00	3.10
10	Кош-Курган	42.30	13.70	3.90	6.10
16	Хантау	3 35.70	3.60	3.80	2.10
17	Бурма	26.80	9.80	0.80	2.10
6	Кара-Бура	23.60	11.50	4.20	6.30
22	Кызыл-Тау пл.2 сдф	22.00	15.00	2 3.30	0.60
20	Кызыл-Тау пл.1 сдф	15.80	11.80	7.10	0.10
21	Кызыл-Тау пл.1 сдф	9.80	7.50	0.00	0.40
15	Семизбугу D	14.30	4.80	6.80	4.10
14	Семизбугу C	21.40	11.90	11.40	7.00
3	Госсор	25.60	4.40	12.20	5 20.60
1	Хонако 3	4 24.40	9.00	14.00	31.00
2	Худжи	27.00	6.00	14.50	6 62.00
8	Огзи-Кичик	35.90	4.60	32.70	44.00
26	Оби-Рахмат 6-9	-	-	-	43.40
27	Оби-Рахмат 2-5	-	-	-	7 43.90
23	Оби-Рахмат 19-21				44.20
4	Джар-Кутан	55.00	41.00	37.80	53.00

Рис.12. Результат работы метода прямого кластерного анализа.

Попытки проинтерпретировать наблюдаемую картину областей могут натолкнуть исследователя на новые, интересные гипотезы. Пример применения прямого кластерного анализа взят из [Деревянко, Холюшкин, Воронин, Костин, 2005: с. 45-46] (см. рис. 12 и табл. 2).

Таблица 2. Дисперсионный анализ областей.

Область	Среднее	Ср.кв.откл.	Объем	Объясняет долю дисперсии (в %)	
				область	Элемент
0	30.920	9.469	10	4.4	0.440
1	51.000	4.546	6	21.7	3.618
2	7.356	4.844	52	32.5	0.624
3	30.657	7.647	7	2.9	0.419
4	26.860	5.459	5	0.9	0.171
5	25.800	7.354	2	0.2	0.121
6	62.000	0.000	1	6.7	6.661
7	43.889	6.889	9	19.3	2.141

Комментарий к таблице 2: Среднеквадратичное отклонение по таблице 16.99.

Объясненная дисперсия 88.5 %.

В частности, рассмотрение структуры областей на рис. 12 приводит к выводу о том, что хорошими признаками для выделения структурной информации являются IF и IL_{am}. Как для всех вышерассмотренных методов, здесь также имеется возможность сформулировать и проверить нулевую гипотезу об отсутствии выделенных областей. В качестве статистики для измерения отклонений от независимости можно выбрать долю объясненной дисперсии (в нашем примере 88.5%), также называемой коэффициентом детерминации R^2 . Такое дополнение превращает метод из эвристического в статистический метод. А это позволяет использовать полученные результаты не только для научного поиска, но и делать вполне обоснованные утверждения.

К методам анализа структур относятся в первую очередь методы автоматической классификации, среди которых наиболее распространен кластерный анализ. Для

проведения кластерного анализа необходимо выбрать несколько признаков – построить так называемое признаковое пространство. Задача состоит в том, чтобы выделить в этом пространстве отдельные сгущения точек – кластеры (рис. 13).

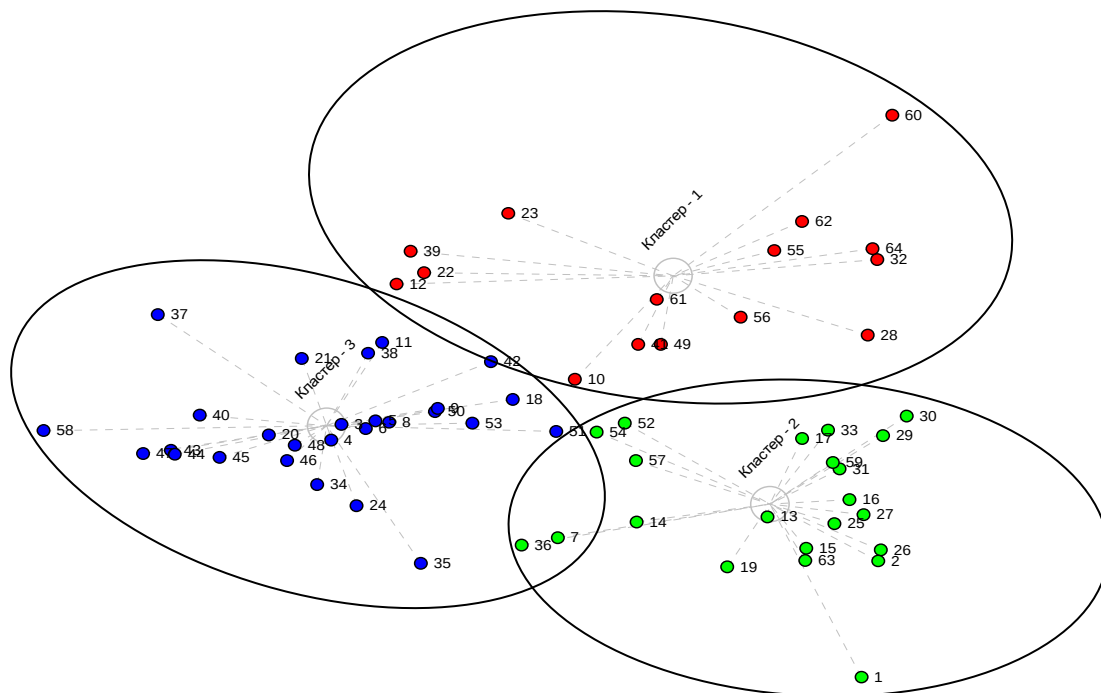


Рис. 13. Результат работы кластерного анализа методом k-средних

Разнообразные типы кластерного анализа активно применялись в археологических исследованиях. В них авторами осознавались недостатки процедур кластерного анализа, главными из которых являются два: отсутствие четких рекомендаций по выбору числа классов и невозможность индивидуального учета отдельных элементов при объединении классов [Деревянко, Холушкин, Воронин, Костин, 2005: с. 5].

В ходе таких исследований было обнаружено, что кластеры, замечательным образом, найденные в первый раз и разумно описанные исследователем, после повторного сбора информации (новых раскопок и нового применения кластерного анализа) могут "рассыпаться" из-за случайности выявленной кластерной структуры (при малых выборках, ненормальных распределениях, плохо обусловленных моделях и т.д.).

С этой целью используется метод повторной выборки с возвращением, известный как метод boot-strap [Efron, 1986]. К сожалению, этот метод мало затронул археологию.

Так Кинти [Kintigh, 1984] использовал выборку методом Монте-Карло, чтобы генерировать псевдодоверительные интервалы для результатов анализа многообразий k-значной кластеризации пространственных данных. Рингроуз [Ringrouse, 1992] использовал boot-strap для оценки подобным способом результатов анализа соответствия. Суть метода состоит в имитации повторного сбора данных, в ходе которой генерируется выборка, совпадающая с исходными данными [Деревянко, Холушкин, Воронин, Костин, 2005: с. 5-6].

Существует множество алгоритмов для выделения сгущений, основанных на различных подходах. Мы выбрали один из простейших и вычислительно наиболее эффективных методов кластерного анализа – метод k-средних. Согласно этому методу, принадлежность объекта к кластеру определяется евклидовым расстоянием между объектом и центром кластера. Объект приписывается к ближайшему кластеру. Процедура начинается с некоторого начального приближения, а затем запускается итерационный процесс, на каждом шаге которого объекты перемещаются между кластерами, что приводит к изменению координат центров кластеров.

Итерации продолжаются до тех пор, пока объекты не перестанут перебегать из одного кластера в другие. При этом достигает своего минимального значения оптимизируемый

функционал – остаточная дисперсия, которая вычисляется как сумма квадратов отклонений координат объектов от центров своих кластеров.

И в этом случае имеется возможность сформулировать и проверить нулевую гипотезу, которая звучит так: в признаковом пространстве точки рассеяны так, что образуют единственный кластер [Жданов, Костин, 2002]. Правда, в этом случае дело обстоит сложнее, чем во всех предыдущих, поскольку разрушить кластерную структуру, оставив в то же время в нетронутым виде связи между переменными, намного сложнее, чем просто разрушить связи. Простое перемешивание признаков здесь не подходит. Приходится вводить дополнительное предположение о том, что точки в пространстве признаков имеют многомерное нормальное распределение, которое не всегда выполняется. Но если мы принимаем такое допущение, то далее обычным путем генерируем данные с разрушенной кластерной структурой, получая экспериментальным путем распределение значений оптимизируемого функционала в условиях нулевой гипотезы об отсутствии многокластерной структуры. По значимости нулевой гипотезы можно обоснованно сказать, наблюдается ли кластерная структура на нашей выборке. Более того, по наименьшей значимости можно даже определить наиболее вероятное количество кластеров.

Но даже возможность получить оптимальную кластерную структуру не позволяет говорить о том, что мы действительно можем извлечь из данных содержащуюся там структурную информацию. Трудность состоит в том, что методы кластерного анализа хорошо работают при небольшой размерности признакового пространства (2-3), а выбор наиболее информативного подпространства признаков превращается в неподъемную переборную задачу. К тому же появляется проблема сравнения результатов классификаций и выбора наилучшей из них.

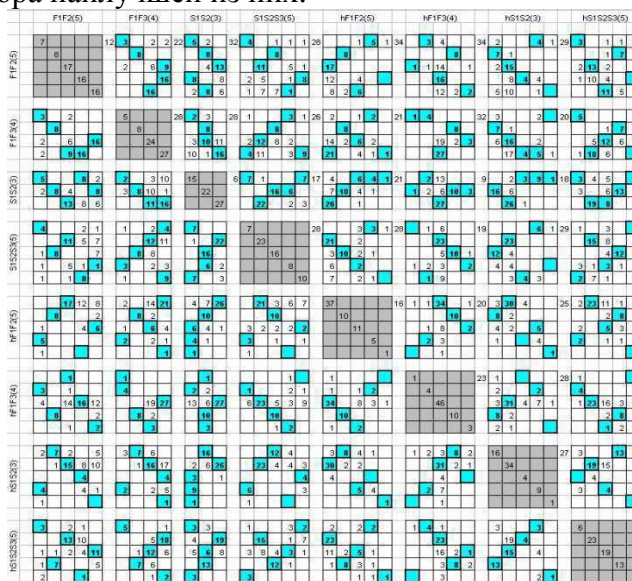


Рис.14. Покластерное совпадение кластеров, построенных по разным классификациям.

Задача сравнения классификаций была поставлена при анализе совпадения классификаций (см. рис. 14), построенных на основе данных по типологии орудий среднепалеолитических индустрий Ближнего и Среднего Востока и Кавказа [Деревянко, Холушкин, Воронин, Костин, 2004: с. 25-66].

Для каждой классификации были выбраны наиболее значимые переменные, позволившие построить наиболее отчетливые группировки:

1. классификация, построенная методами k -средних в пространстве факторов 1, 2 – выделено 5 кластеров;
2. классификация, построенная методами k -средних в пространстве факторов 1, 3 – выделено 4 кластера;
3. классификация, построенная методами k -средних в пространстве шкал 1, 2 [Деревянко, Холушкин, Воронин, Костин, 2004: с. 36] – выделено 3 кластера;

4. классификация, построенная методами *k*-средних в пространстве шкал 1, 2, 3 [Дервянко, Холушкин, Воронин, Костин, 2004: там же] – выделено 5 кластеров;
5. классификация, построенная методами иерархического кластерного анализа в пространстве факторов 1, 2 (выделено 5 кластеров);
6. классификация, построенная методами иерархического кластерного анализа в пространстве факторов 1, 3 (выделено 4 кластера);
7. классификация, построенная методами иерархического кластерного анализа в пространстве шкал 1, 2 (выделено 3 кластера);
8. классификация, построенная методами иерархического кластерного анализа в пространстве шкал 1, 2, 3 (выделено 5 кластеров).

Таблица 3. Совпадение кластеров, построенных по разным классификациям (в %).

Исходные классификации	Сопряженные классификации							
	1	2	3	4	5	6	7	8
1	100.0	81.3	65.6	50.0	56.3	46.9	46.9	54.7
2	81.3	100.0	56.3	56.3	59.4	67.2	50.0	68.8
3	65.6	56.3	100.0	90.6	73.4	67.2	85.9	71.9
4	50.0	56.3	90.6	100.0	56.3	56.3	70.3	54.7
5	56.3	59.4	73.4	56.3	100.0	75.0	68.8	60.9
6	46.9	67.2	67.2	56.3	75.0	100.0	64.1	56.3
7	46.9	50.0	85.9	70.3	68.8	64.1	100.0	57.8
8	54.7	68.8	71.9	54.7	60.9	56.3	57.8	100.0

По этим классификациям построены попарные классификации для перечисленных разбиений. Результаты анализа совпадений кластеров при попарном сравнении состава их элементов приведены в таблицах 3, 4.

Таблица 4. Несовпадение кластеров, построенных по разным классификациям (количество объектов, не попавших в общие кластеры).

Исходные классификации	Сопряженные классификации							
	1	2	3	4	5	6	7	8
1		12	22	32	28	34	34	29
2	12		28	28	26	21	32	20
3	22	28		6	17	21	9	18
4	32	28	6		28	28	19	29
5	28	26	17	28		16	20	25
6	34	21	21	28	16		23	28
7	34	32	9	19	20	23		27
8	29	20	18	29	25	28	27	

В.С. Костиным был предложен вариант решения, где в качестве статистики, измеряющей степень отклонения от независимости классификаций, выбрана максимальная доля совпадающих объектов при оптимальном соответствии кластеров [Костин, 2003]. А задачу выбора наилучшей классификации была трансформирована в задачу объединения результатов большого количества независимо построенных классификаций [Костин, Корнюхин, 2003] и построения на основе этого объединения классификации обобщенной. Наглядное представление обобщенной классификации предлагается на рис. 15. В клетках представленной на этом рисунке таблицы указана степень согласованности включения объектов в одни и те же кластеры. Более темным оттенкам серого соответствует высокая степень согласованности, более светлым – менее высокая.

Оценивая весь оригинальный (не полностью описанный из-за ограничений на объемы в настоящем издании) инструментарий, следует указать наиболее важную особенность методологии и методики, на которой он выстроен: все процедуры и методы завершаются обязательной проверкой статистической значимости полученных результатов.

Следует заметить, что непрерывное пополнение описанного нами выше инструментального арсенала для статистического анализа археологических данных продолжается.

Дальнейшим шагом в этом направлении будет получение не одной, а нескольких обобщенных кластерных структур. Проблема состоит в том, что весь набор признаков, описывающий исследуемые объекты, как правило, отражает их с самых разных сторон,

выделяя разные предметы исследования одного и того же объекта, разные уровни и формы организации и развития мира [Витяев, Костин, 1992].

За каждым предметом исследования стоит свой класс аспектов. Если же мы смешаем все классификации аспектов в одну, то получим либо необоснованно переусложненную картину кластеров (пересечение аспектных или предметных классификаций), либо не получим отчетливой картины вовсе.

	1	2	3	4	5	6	7	8	9	10	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
1	1	2	3	4	5	6	7	8	9	10	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
2	2	1	3	4	5	6	7	8	9	10	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
3	3	2	1	4	5	6	7	8	9	10	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
4	4	3	2	1	5	6	7	8	9	10	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
5	5	4	3	2	1	6	7	8	9	10	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
6	6	5	4	3	2	1	7	8	9	10	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
7	7	6	5	4	3	2	1	8	9	10	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
8	8	7	6	5	4	3	2	1	9	10	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
9	9	8	7	6	5	4	3	2	1	10	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
10	10	9	8	7	6	5	4	3	2	1	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	11	10	9	8	7	6	5	4	3	2	1	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
12	12	11	10	9	8	7	6	5	4	3	2	1	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
13	13	12	11	10	9	8	7	6	5	4	3	2	1	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
14	14	13	12	11	10	9	8	7	6	5	4	3	2	1	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
15	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	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
16	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	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
17	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36																																																																

процессора MS Excel с его развитыми функциональными возможностями и использование языка VBA (Visual Basic for Applications)» [Гарскова, 2018: с. 383].

Завершение создания концепции структурного анализа данных привело к мысли объединить разрозненные программные разработки в едином статистическом пакете и сделать их доступными для более широкого круга заинтересованных пользователей. Для расширения возможностей проектируемого проекта предполагалось использовать отдельные модули пакета PSPP, аналогичного по функциональности SPSS, но распространяемого бесплатно по соглашению "Open source". В нём заложены возможности использования исходного кода программы для включения в некоммерческие программные продукты [<http://savannah.gnu.org/projects/pspp/>] Кроме того, предусматривалась разработка элементов картографического представления полученных результатов [Холюшкин, Воронин, Костин, 2008: с.50].

Поиск возможных аналогов разработки пакета показал, что международный рынок насчитывает несколько сотен коммерческих пакетов в среде операционных систем Windows, DOS, OS/2 и почти столько же коммерческих пакетов, распространяемых бесплатно или за символическую плату. Отметим лишь несколько пакетов, достоинством которых является наличие достаточно объёмной документации на русском языке:

SPSS (Statistical Package for Social Science). Наиболее часто используемый пакет статистической обработки данных отличается гибкостью, мощностью и применим для всех видов расчётов в биомедицине, биологии и социальных науках. Существует полностью русифицированная версия, однако графическое представление результатов в пакте до сих пор оставляет желать лучшего.

STATISTICA (Производитель программы фирма StatSoft Inc. (USA)) [<http://www.statsoft.com/>], которая выпускает статистические приложения, начиная с 1985 года. STATISTICA включает большое количество методов статистического анализа (более 250 встроенных функций). Несложный в освоении этот статистический пакет может быть рекомендован для археологических, социальных и биомедицинских исследований.

Среди частных приложений, используемых в истории и археологии можно отметить программу FUZZYCLASS, разработанную в Лаборатории исторической информатики МГУ под руководством чл.-корр. РАН Л.И. Бородинкина.

FUZZYCLASS – гибкий аппарат, учитывающий неоднородность и неоднозначность исторического источника. Он эффективен в решении не только задач типологии социально-исторических процессов, но и перспективен в археологии, где разделение на типы – процедура интуитивная. Концептуальное значение этой программы заключено в признании размытой природы классов, т.е. постепенного, не скачкообразного перехода реального образа от принадлежности к непринадлежности к этим классам.

Метод нечеткой классификации получил достаточно широкое распространение.

В публикациях АИК можно видеть применение метода в исследованиях по экономической и социальной истории, а также в исторической демографии и в археологических исследованиях [Абдулганеев, Владимиров, 1994: с. 121–128; Они же, 1997; Кокорина, Лихтер, 2008: с. 301–309; Лихтер, 2000: с. 76-82; Беседин, 1996: с.80-81; Он же, 1995: с. 76-77].

При таком обилии компьютерных пакетов и хорошей проработанности тематики оставалась непонятной необходимость создания ещё одного статистического продукта. Тем более что авторские программные продукты, как универсальные, так и специализированные, «не имеют тех преимуществ, которыми обладает коммерческое программное обеспечение: они часто написаны на устаревших языках программирования, имеют не слишком удобный интерфейс, их непросто осваивать, отсутствует обязательное для коммерческих программ сопровождение и обновление. Все это предсказывает таким программам короткую жизнь» [Гарскова, 2018].

В свою очередь, Л.И. Бородинкин, констатировал, что «сложившаяся практика исследований в области исторической информатики базируется на стандартном

программном обеспечении, которое способно решать широкий круг исследовательских задач. Однако стандартные программы не могут давать ответы на нестандартные вопросы, и, кроме того, развитие исторической информатики как самостоятельной научной дисциплины, имеющей не только свой предмет, но и свой специальный методический арсенал, предполагает создание оригинальных алгоритмов, программ и технологий, ориентированных на исторические исследования и образование» [Бородкин, 2004: с.3-6]

По нашему мнению, необходимость создания собственного статистического пакета диктовалась следующими соображениями [Холушкин, Воронин, Костин, 2008: с. 51-52]:

Во-первых, статистические программы относятся к наукоёмкому программному обеспечению, цена их часто недоступна индивидуальному пользователю из-за высокой цены на легальные копии профессиональных статистических пакетов, дороговизны сопровождения по этим пакетам (обучение, консультации, льготная поставка новых версий и т.д.), так стоимость двухлетней лицензии на пользование пакетом SPSS составляла 15000\$ США.

Во-вторых, все известные западные пакеты ориентированы на наличие профессиональных навыков и высокой квалификации, широкого первоначального статистического образования, доступной литературы и консультационных служб. Нам уже приходилось писать о низкой математической культуре археологов Сибири, особенно в области палеолита и насущных мерах её повышения. Так в прекрасной коллективной монографии "Природная среда и человек в палеолите Горного Алтая. – Новосибирск, 2003" (под редакцией А.П. Деревянко) самым слабым местом, как ни странно, был раздел, посвящённый описанию археологической коллекции.

В-третьих, крупные пакеты ориентированы на пользователя, готового к тщательному изучению сложной и объёмистой документации, как на русском (SPSS, STATISTICA), так и на английском языке.

В-четвёртых, абсолютное большинство существующих пакетов являются локальными приложениями, к которым нет доступа из сети Интернет.

Поэтому на начальном этапе были сформулированы основные принципы построения нашей системы и вытекающие из них требования к пакету:

1. Общедоступность пакета – любой пользователь сети Интернет может воспользоваться услугами пакета, при этом не требуется установка пакета на локальную машину, все вычисления производятся на сервере.
2. Высокая скорость обработки информации.
3. Основная категория пользователей – археологи, имеющие собственные данные или гипотезы, которые могут быть проверены на уже опубликованных данных. Для этого сайт должен предоставлять пользователю возможность вводить, сохранять и редактировать собственные данные. Данные должны храниться на сервере как в течение сеанса работы, так и (для зарегистрированного пользователя) между сеансами. Кроме того, должна быть возможность сохранения данных на стороне клиента, то есть экспорт/импорт.
4. Доступ к архиву обобщённых данных (со ссылками на источник). Из общего архива данные могут быть скопированы в личный архив, после чего их можно редактировать и сохранять. Также необходима функция пополнения общего архива, хотя она может быть реализована не полностью автоматически, а с участием человека, администратора данных.
5. Простота в использовании пакета. Для проведения исследования в нашем пакете не требуется знание внутреннего языка, как в других локальных аналогах, взамен этого пользователю предоставляется удобный интерфейс.
6. Простой и эффективный интерпретационный аппарат. Предполагается, что пользователи хорошо ориентируются в собственных данных, но не имеют специальной подготовки в методах статистического анализа. Поэтому:
7. пользователю должна быть предоставлена не только возможность применять методы анализа, но и получать наглядное объяснение, в каких случаях их можно применять и какие выводы можно сделать из полученных результатов, т.е. нужна добротная обучающая подсистема, включая тексты с описанием методов и примеры анализа реальных данных;

8. помимо шаблонных интерпретаций методов у пользователей пакета должна быть реальная возможность самим разобраться в сущности этих методов предоставлением пошаговой демонстрацией производимых преобразований на живых данных, причем каждый шаг должен быть снабжен описанием, достаточным для понимания и самостоятельного воспроизведения;
9. Возможность для научных сотрудников применять методы анализа данных не только в процессе исследования, но и для подготовки научных публикаций. Поэтому:
 - 9.1. вместе с выводом результатов анализа система должна предлагать формулировки выводов, достаточно строгие и точные, чтобы их можно было использовать в публикациях.
 - 9.2. все графические иллюстрации должны быть информативными и выполнены на хорошем дизайнерском уровне в соответствии с требованиями к рисункам в научных публикациях.
10. Наряду с собственными, оригинальными методами наличие достаточно полного набора стандартных статистических методов.
11. Реализация возможности картографического представления результатов статистического анализа.

Для реализации проекта важной составляющей представлялась кадровая составляющая проекта. В составе участников были опытные научные сотрудники и программисты:

Ю.П. Холюшкин – руководитель проекта

В.Т. Воронин – участник проекта

В.С. Костин – участник проекта

А.Н. Нуртдинов – участник проекта

В.В. Воробьев – участник проекта

М.Ю. Ильиных – участник проекта

И.Г. Комиссарова – участник проекта

В.М. Семенов – участник проекта

Р.В. Рыбинцев – участник проекта

А.А. Грищенко – участник проекта

К сожалению, в начальной стадии выполнения проекта (2007 год), в ходе так называемого процесса по упорядочиванию численности институтов РАН были уволены "по собственному желанию" 60% участников проекта:

А.Н. Нуртдинов – специалист по ГИС и мат статистике;

В.В. Воробьев – специалист по базам данных и web-дизайну;

М.Ю. Ильиных – специалист по базам данных;

И.Г. Комиссарова – специалист по мат. статистике;

В.М. Семенов – специалист по трехмерной графике.

А.А. Грищенко – специалист по трехмерной графике.

Поэтому возникли серьезные проблемы в выполнении программы проекта: 4.5.2. "Разработка научных основ распределённой информационно-аналитической системы на основе ГИС и Веб-технологий для междисциплинарных исследований" (научный проект 4.5.2.11 "Создание on-line статистического пакета анализа информации гуманитарных исследований с элементами картографии на сайте "Sibirica" ").

Кроме того, в мае 2008 года случилась ещё одна потеря: неожиданно, из-за переживаний за судьбу сектора, да и за собственную также, скончался один из ведущих участников проекта к.э.н. В.Т. Воронин. Будучи личностью одарённой, он с увлечением разрабатывал проблему времени⁴, а параллельно сочинял музыкальные произведения. Известно, что в своё время он обратил на себя внимание известного композитора Арама Хачатуряна, который посоветовал начинающему композитору серьёзно заняться

⁴ Через год, после кончины В.Т. Воронина мы за свой счёт (Ю.П. Холюшкин и В.С. Костин) издали в 14 выпуске ИТГИ его последнюю работу "Тайны времени" (*Посвящается памяти Платона, Августина Блаженного, Исаака Ньютона и Иммануила Канта*) / Рос. акад. наук. Сиб. отд-ние. Ин-т археологии и этнографии; Ассоциация «История и компьютер». Сиб. отд-ние; Отв. ред. Ю.П.Холюшкин. Новосибирск: ЗАО РИЦ Прайс-курьер, 2009. - 122 с.: ил. - Библиогр.: с.121-122.

музыкальным образованием. Мысль об этом, видно, никогда не покидала его. Незадолго до своей кончины перед очередной аттестацией он мне сказал: вот выгонит меня окончательно на пенсию М.В. Шуньков, тогда всерьёз займусь музыкой [Холюшкин, 2018: с. 14].

Возникла альтернатива: (если учесть, что в 2008 году был намечен 3 этап по увольнениям (25%)), смиренно ждать команды: "с вещами на выход" или спокойно продолжать работу. Мы предпочли второе и привлекли к выполнению проекта дипломников ВКИ НГУ:

Болдырев А.В. – разработчик пользовательского интерфейса [Холюшкин, Воронин, Костин, Болдырев, 2008: с. 54-61];

Ильчук А.Ю. – Разработчик программного интерфейса [Холюшкин, Воронин, Костин, Ильчук, 2008: с. 62-65];

Мок А.В. – Методы анализа связей [Холюшкин, Воронин, Костин, Мок, 2008: с. 66-70];

Данник Е.В. – Методы анализа структур [Холюшкин, Воронин, Костин, Данник, 2008: с. 71-75];

Есина И.В. – Дискриминантный анализ [Холюшкин, Костин, Есина, 2010: с.74-83];

Федотова Д.В. – Регрессионный анализ [Холюшкин, Костин, Федотова, 2010: с.83-89].

При разработке пакета участниками проекта использовались операционная система FreeBSD, Web-сервер Apache, СУБД MySQL, ZEND Studio Client 5.1.0., язык сценариев PHP, язык гипертекстовой разметки HTML, каскадные таблицы стилей CSS, графический редактор Adobe Photoshop 8.0 CS, Internet Explorer 6.0, VRML, ГИС⁵.

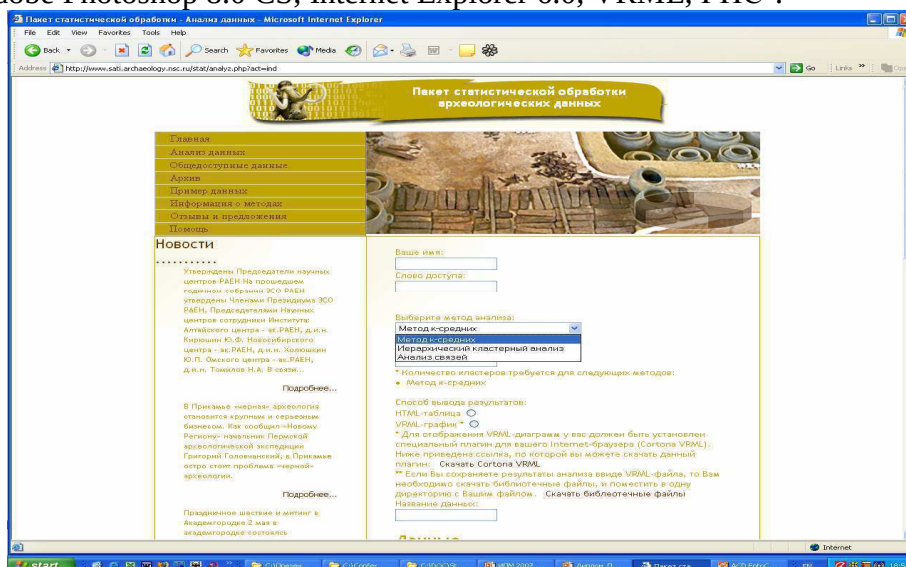


Рис. 16. Главная страница пакета.

Пакет включал полный набор методов статистики, в том числе методы многомерного анализа: кластерный, дискриминантный, регрессионный анализ и др.

В ходе реализации был разработан интерфейс пользователя и администратора системы. Пакет содержит следующие разделы:

Главная страница – содержала описание Пакета и сравнение с аналогами (рис. 16).

Для полноценного пользования пакетом требовалась регистрация пользователя. Для незарегистрированных пользователей долгосрочное хранение данных не поддерживалось. Все данные удалялись сразу после выдачи результата. На основании этого можно сделать вывод, что любые операции с данными и результатами (просмотр, редактирование, опубликование в разделе общедоступных данных, создание личного архива) для незарегистрированных пользователей невозможны.

⁵ Авторы разработки пакета не остановили своего выбора на какой-либо конкретной системе из-за отсутствия средств на приобретение программного продукта, а разработке собственного картографического приложения помешало сокращение единственного специалиста в этой области.

Зарегистрированным пользователям предоставлялись следующие права [Холюшкин, Воронин, Костин, Болдырев, 2008: с. 54-61]:

- анализировать данные (с выбором настроек анализа);
- сохранять проанализированные данные в системных таблицах (с уникальным именем);
- сохранять результаты анализа данных (по выбору пользователя – в виде HTML-таблиц или в виде трёхмерных VRML-диаграмм);
- просматривать и манипулировать собственным архивом;
- отправлять запрос администратору на добавление данных в раздел "Общедоступные данные";
- иметь доступ к полному руководству по использованию статистического пакета и полную информацию о статистических методах, используемых при анализе данных;
- обмениваться сообщениями с другими пользователями пакета посредством Гостевой книги.

Администраторы имели отдельный вход в систему и иной интерфейс для управления этой системой. Администраторы выполняли следующие функции:

- редактирование и удаление некорректных сообщений в Гостевой книге;
- добавлять и редактировать новости в новостном окне;
- отправлять пользователям сообщения по электронной почте;
- проверка на корректность и добавление данных от пользователей в раздел общедоступных.

Входные данные

Во входных данных результаты проведённых исследовательских работ упорядочивались в формате таблиц MS Excel (рис 17). Данные обязательно должны быть в строго определённом виде – табличном представлении.

Рис. 17. Таблица переменных

Описание всех переменных с указанием типов;

Метки: расшифровка меток, используемых в данных.

Кроме того, пользователь выбирает один из двух способов вывода результатов анализа: HTML-таблица или трёхмерная VRML-диаграмма.

Каждая строка представляла объекты (стоянки, или её горизонты). А столбцы (переменные) – значения признаков (характеристик) вводимых объектов. Таким образом, ячейки содержали только значения переменных. В качестве переменных могли быть указаны признаки артефактов, анализируемых типов классификации и др.

После ввода данных пользователь выбирал метод, с помощью которого будет производиться анализ данных. При этом обязательное указание количества кластеров требует только "Метод к-средних". Для остальных методов это необязательно.

Пункт "Сделать эти данные рабочими": позволяет активизировать все данные, находящиеся в полях "Данные", "Описание", "Метки". Все они были занесены в системный буфер, и при последующих обращениях в раздел "Анализ данных" повторно вводить эти же самые данные не требовалось. Активация пункта "Сделать эти данные рабочими» могла потребоваться, например, в тех случаях, если пользователь хочет проанализировать одни и те же данные на разных методах или с различным способом вывода результатов.

Время анализа зависит от объёма данных и от выбранного метода, но обычно не превышает 3-5 секунд.

Анализ данных – раздел, в котором пользователь может проанализировать свои данные (рис.18).

Пер1	Пер2	Метод	Степ. свобод. 1	Степ. свобод. 2	Статистика	Значимость
steq3	steq4	Дисперсионный анализ	14	0	0	0
steq3	steq5	Дисперсионный анализ	12	2	225.291111111111	0.99898176959086
steq3	steq6	Дисперсионный анализ	5	9	7686.71	0.83914758055623
steq3	steq7	Дисперсионный анализ	8	6	2299.6566666667	0.96073167800585
steq3	steq8	Дисперсионный анализ	14	0	0	0
steq3	steq9	Дисперсионный анализ	12	2	195.307777777778	0.99898223804432
steq3	steq10	Дисперсионный анализ	13	1	0	1
steq4	steq5	Дисперсионный анализ	12	2	7126.92	0.99892199465445
steq4	steq6	Дисперсионный анализ	5	9	77765.09	0.81762265788618

Рис. 18. Пример результата расчётов однофакторного дисперсионного анализа.

В период существования пакета для пользователей были доступны:

а) Анализ связей. В пакете были реализованы три метода анализа (дисперсионный, корреляционный и ХИ-квадрат). Целью анализа связей является обнаружение взаимосвязей зависимостей между признаками. В статистике разобрано множество критериев для проверки связей в данных. Но все они построены по одному принципу: в каждом критерии формулируется нулевая гипотеза, которая утверждает, что исследуемые признаки являются независимыми случайными величинами, связь между которыми если и проявляется, то исключительно в силу случайного совпадения. Проверка любого критерия начинается с вычисления своей статистики – величины, характеризующей степень отклонения от независимости. Поскольку археологические признаки могут быть измерены в любой из трёх шкал (номинальной, порядковой и количественной), то для каждого сочетания шкал надо применять свои критерии. Например, если обе переменные измерены в шкале наименований, то можно применять критерий «хи-квадрат», если одна из них – номинальная, а другая – количественная, то можно пользоваться дисперсионным анализом, а если обе количественные, то подойдёт корреляция по Пирсону.

б) Анализ структуры. В пакете были реализованы два метода анализа структуры (метод "к-средних" и агломеративный метод иерархической классификации) (рис19).

Способ вывода результатов анализа. Существует два варианта отображения результатов: HTML-таблица и трёхмерная VRML-диаграмма.

Для отображения VRML-диаграмм необходимо, чтобы на машине пользователя был установлен VRML-клиент (Cortona VRML) (рис.20).

Номер	Имя	Коорд. [1]	Коорд. [2]	Коорд. [3]	Кластер
1	Site	1	1	1	2
2	Layer	1	2	3	2
3	Time	10	11	11	2
4	steq1	47	32	2	2
5	steq2	75	43	10	2
6	steq3	241	170	19	2
7	steq4	528	319	56	2
8	steq5	72	46	9	2
9	steq6	5	6		2
10	steq7	72	44	12	2
11	steq8	514	408	97	2
12	steq9	52	21	5	2
13	steq10	2374	1950	366	3
14	steq11	8234	7091	2315	1
15	steq12	348	256	52	2
16	steq13	106	49	5	2

Номер	Имя	Коорд. [1]	Коорд. [2]	Коорд. [3]	Кластер
1	Кластер 1	8234.0000	7091.0000	2315.0000	1
2	Кластер 2	140.4285	97.0714	19.7857	2
3	Кластер 3	2374.0000	1950.0000	366.0000	3

Количество итераций: 3

Рис.19. Пример вывода табличного результата анализа методом “к-средних”

Технология VRML обеспечивает возможность создания некой комплексной модели, состоящей из трехмерных и двухмерных объектов (рис.21), а также звука и прочей мультимедиа информации. При этом двухмерные объекты могут быть как текстовыми, так и иллюстрациями и фотографиями. При всём этом VRML не предъявляет никаких особых требований к компьютерной платформе или программному обеспечению. Как стандарт, разрабатываемый специально для Интернета, VRML открывает в широкие возможности в сфере представления и передачи информации благодаря малым размерам исходных VRML-документов и высокой информативности трёхмерных образов.

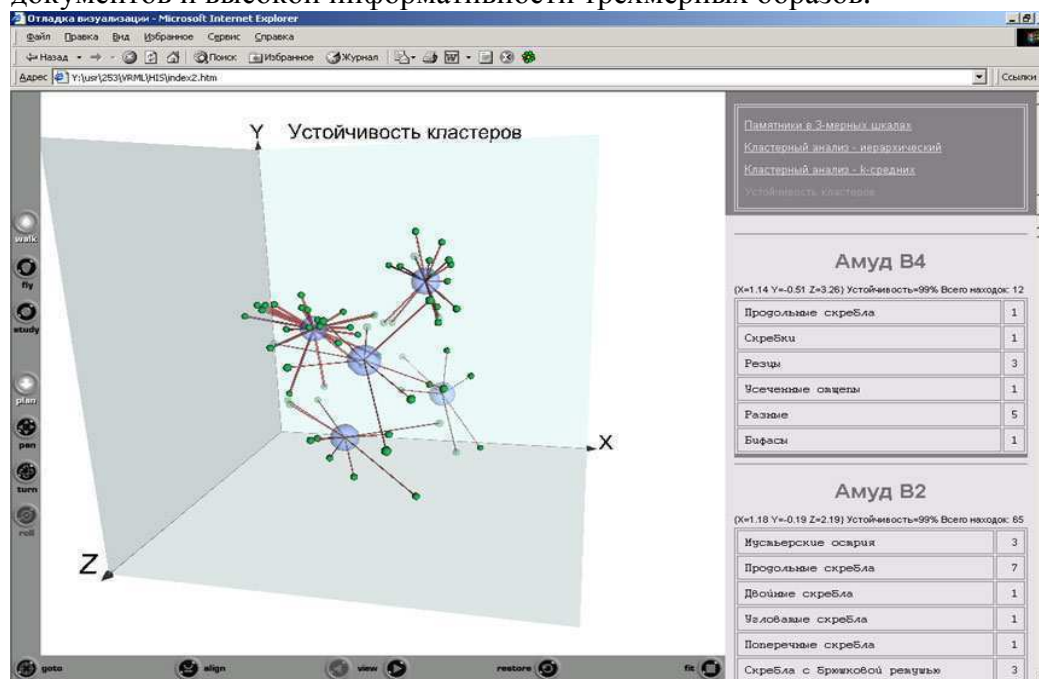


Рис.20. VRML диаграмма кластерного анализа к-средних.

Возможное внедрение интерактивности с аудио объяснением полученных результатов могло бы делать более реалистичными и удобными в понимании и освоении демонстрируемого объекта. Это обстоятельство является важным для неподготовленного пользователя, которыми являются археологи. Именно этот аспект выдвигает VRML в категорию средств предоставления доступа к информации, более полно описывающей реальные объекты. Бесспорным преимуществом технологии VRML является возможность её сетевого использования [Холушкин, Воронин, Костин, 2008: с.53].

Общедоступные данные – содержат данные и результаты анализов, которые пользователи посчитали "интересными".

На странице раздела общедоступных данных (переход в данный раздел осуществляется нажатием соответствующей ссылки в главном меню) представлены данные и результаты анализа, которые пользователи посчитали интересными и полезными для других пользователей системы. Данные в раздел добавляются только после проверки администратором на корректность.

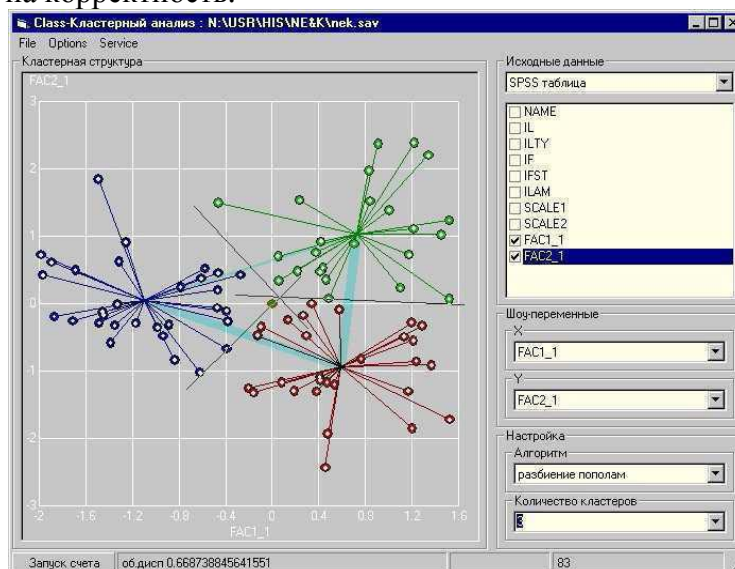


Рис.21. Плоская визуализация кластерного анализа. Анимация устойчивости разбиения множества объектов на кластеры с помощью повторной выборки с возвращением (bootstrap).

Архив – Для зарегистрированных пользователей формируется личный архив, в котором представлены все данные, когда-либо анализировавшие пользователи. Для удобства поиска нужных архивных данных, названия данных состоят из двух частей: собственно название, под которым пользователь анализировал эти данные, и дата анализа. При нажатии на соответствующее название данные будут представлены в полном размере вместе с результатами анализа. Чтобы войти в собственный архив нужно ввести логин (имя, под которым Вы зарегистрировались) и пароль (выбранный при регистрации) и нажать кнопку "Войти". Если логин и пароль правильные, то происходит автоматическое перемещение на главную страницу личного архива. В противном случае система выдаст сообщение о неправильности логина или пароля и нужно повторить процедуру.

При нажатии на кнопку  выбранные данные удалятся из архива.

Кнопка  позволяет отправить администратору запрос на добавление выбранных данных и результатов в раздел общедоступных. После одобрения администратором, эти данные попадут в раздел общедоступных данных.

Пример данных – содержит примеры данных, которые можно просмотреть в полном размере и скачать на свой компьютер.

Несмотря на наличие ряда серьезных разработок в современной археологии, общее состояние дел в теории классификации требует уже иных подходов, основанных на новом направлении, получившем название "интеллектуальный анализ данных". Авторами проекта была поставлена задача разработки "сценариев" решения археологических задач, путем последовательностей применения методов (стратегий) интеллектуального. Ниже приводится описание этих последовательностей.

К сожалению, нами в ходе проведения работ был создан лишь макетный вариант пакета, созданный для проверки схемы функционирования будущего сайта. Все разработки не были до конца встроены в единую технологическую цепочку. В секторе был лишь проблемный программист, но отсутствовали программисты – разработчики, способные погрузиться в сложные задачи.

1.4. Описание последовательностей применения методов (стратегий) интеллектуального анализа данных для решения задач археологии.

После смерти В.Т. Воронина встал вопрос о приёме на его место нового сотрудника. Для этого нужно было найти специалиста, разрабатывающего алгоритмы классификации, основанные на различных априорных предположениях: гипотезе компактности, различных мерах близости, сходством с эталонами, суперцелях, на различных функционалах качества и т.д. Специалистов, связанных с разработкой алгоритмов, основанных на принципах "естественной" классификации можно было найти в ИМ СО РАН. Применяемые там методы удовлетворяли свойству "прозрачности", когда конечный пользователь может детально проанализировать полученные результаты и убедиться в их правильности путем демонстрации на объектах обнаруженных закономерностей и классов.

Разработанные системы "естественной" классификации позволили:

1. анализировать объекты – сравнивать объект по признакам с другими объектами, предсказаниями и идеализированными;
2. анализировать закономерности – определять на каких объектах они получены;
3. анализировать идеализированные классы – определять объекты, относящиеся к этим классам, и определять каким закономерностям они удовлетворяют.

Для решения этих задач возникло два предполагаемых кандидата:

1. В.В. Бериков, д.т.н., который участвовал в интеграционном проекте СО РАН № 89. Успешно сотрудничал с антропологами и сотрудниками кафедры археологии НГУ, возглавляемой Ю.С. Худяковым.
2. Е.Е. Витяев, д. физ-мат.н., с которым В.С. Костин имел совместные публикации. В итоге мы остановились на кандидатуре Е.Е. Витяева (таблица 5).

Таблица 5. Сравнение интересов В.В. Берикова и Е.Е Витяева.

д.т.н. В.В. Бериков	д. физ-мат.н. Е.Е. Витяев
Логические решающие правила	Логические закономерности
Оба метода работают на данных, представленных таблицами объект-свойство.	
Метод Берикова более тщательно подходит к статистическим основаниям отбора правил. В результате полученные правила могут интерпретироваться и по отдельности, хотя вероятность случайного обнаружения ошибочного правила на 100% не исключается.	Метод Витяева использует правила (закономерности) более общего вида, что позволяет ему восстанавливать функциональные зависимости типа физических законов. Кроме обнаружения закономерностей, предсказывающих значения целевого признака, метод включает процедуру естественной классификации, основанной на корректировке реальных объектов предсказанными значениями по очень большому набору закономерностей, которые не анализируются по отдельности, а работают вместе, как преобразователь эмпирических объектов в идеальные образы.

Для реализации этих задач Е.Е. Витяева необходимо использовать Интеллектуальный анализ данных — одно из новых направлений искусственного интеллекта. Этот термин является кратким и весьма неточным переводом с английского языка терминов Data Mining и Knowledge Discovery in Databases (DM&KDD) [Брандт, 2003].

Data Mining — это процесс обнаружения в сырых данных (row data) ранее неизвестных, нетривиальных, практически полезных, доступных интерпретации знаний (закономерностей), необходимых для принятия решений в различных сферах человеческой деятельности.

Предлагаемое применение технологий DM&KDD обусловлено накоплением огромных объемов информации в археологических компьютерных базах данных, которыми стало трудно пользоваться традиционными способами. Последнее обстоятельство связано со стремительным развитием вычислительной техники и программных средств для представления и обработки данных. Большие объемы накопленных данных постоянно приходится модифицировать из-за быстрой смены аппаратного и программного обеспечения БД, при этом неизбежны потери и искажение информации. Одним из средств для преодоления подобных трудностей является создание информационных хранилищ данных, доступ к которым не будет сильно зависеть от изменения данных во времени и от используемого программного обеспечения. Другой подход ориентирован на сжатие больших объемов данных путем нахождения некоторых

общих закономерностей (знаний) в накопленной информации. Оба направления актуальны с практической точки зрения. Второй подход более интересен для специалистов в области ИИ, так как связан с решением проблемы приобретения новых знаний. Следует заметить, что наиболее плодотворным является сочетание обоих направлений.

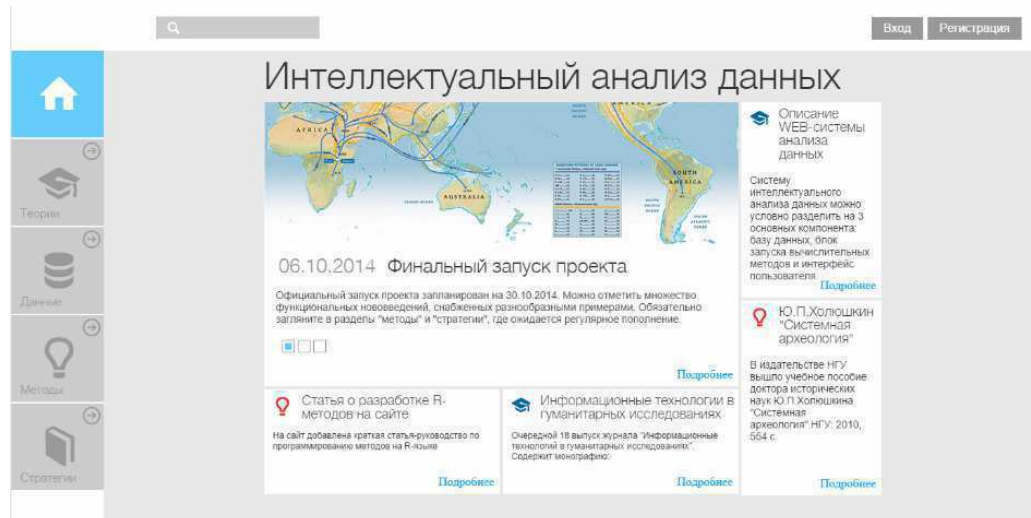


Рис.22. Портал интеллектуального анализа данных

Наличие хранилища данных — необходимое условие для успешного проведения всего процесса KDD. Хранилищем данных называют предметно-ориентированное, интегрированное, привязанное ко времени, неизменяемое собрание данных, используемых для поддержки процесса принятия управленческих решений. Предметная ориентация означает, что данные объединены в категории и хранятся в соответствии с теми областями, которые они описывают, а не в соответствии с приложениями, которые их используют. Такой принцип хранения гарантирует, что полевые и научные отчеты археологов, сгенерированные различными участниками процесса их создания, будут опираться на одну и ту же совокупность данных. Привязанность ко времени означает, что хранилище можно рассматривать как собрание исторических данных, т.е. конкретные значения данных однозначно связаны с определенными моментами времени. Атрибут времени всегда явно присутствует в структурах хранилищ данных. Данные, занесенные в хранилище, уже не изменяются в отличие от оперативных систем, где присутствуют только последние, постоянно изменяемые версии данных. Для хранилищ данных характерны операции добавления, а не модификации данных. Современные средства администрирования хранилищ данных обеспечивают эффективное взаимодействие с программным инструментарием DM и KDD. В общем случае зависимости, выявляемые в базах данных, могут быть представлены правилами, гипотезами, моделями нейронных сетей и т.п. Интеллектуальные средства извлечения информации позволяют почерпнуть из БД более глубокие сведения, чем традиционные системы оперативной обработки транзакций (OLTP — On-Line Transaction Processing) и оперативной аналитической обработки (OLAP). Выведенные из данных закономерности и правила можно применять для описания существующих отношений и закономерностей археологических данных, и прогнозирования их последствий.

Извлечение знаний из БД является одной из разновидностей машинного обучения, специфика которой заключается в том, что реальные БД, как правило, проектируются без учета потребностей извлечения знаний и содержат ошибки.

В технологиях DM&KDD используются различные математические методы и алгоритмы: классификация, кластеризация, регрессия, прогнозирование временных рядов, ассоциация, последовательность [Холушкин, Витяев, Мартынович, 2014: с.11].

Прежде чем мы перейдем к составлению исследовательских стратегий, следует познакомить читателя со спектром предлагаемых системой методов. Программирование

стратегий невозможно себе представить без ознакомления с этим инструментарием, ведь методы служат кирпичиками, или атомарными элементами в схеме стратегии.

Для упрощенной навигации, методы группированы по семантическому признаку и распределены соответствующим образом по директориям в общедоступном системном архиве (таб. 6):

Таблица 6. Структура общедоступного архива методов

Раздел	Путь в файловой системе	Описание
Matrix	/public/matrix	Методы для работы с числовыми матрицами: арифметические, собственные значения, ядра, детерминант
Data	/public/data	Преобразования исходных данных: сортировки, транспонирование, преобразование форматов, исключение нечисловых признаков и прочее
Graph	/public/graph	Операции на графах: составление графа по инцидентности, выделение остова, алгоритмы на графах
Generators	/public/generators	Генераторы случайных последовательностей (различные распределения)
FillGaps	/public/fillGaps	Заполнение пробелов в данных и устранение недочетов
Clustering	/public/clustering	Кластеризации объектов в выборке по какому-либо признаку или их группам
OR	/public/or	Методы исследования операций (OperationResearch)
SubOptimal	/public/suboptimal	Методы исследования операций (OperationResearch)
Statistics	/public/statistics	Статистические методы
Statistics/Basic	/public/statistics/basic	Базовые операции: среднее, дисперсия, ковариация и прочее
Statistics/Criteria	/public/statistics/criteria	Статистические критерии

Следующую таблицу можно использовать как справочник при проектировании собственных стратегий. Однако не следует считать перечисленный ниже список исчерпывающим, поскольку система будет постоянно дополняться новыми методами (с помощью внедрения существующих в R-языке, либо согласно запросу пользовательской аудитории) (таб. 7):

Таблица 7. Библиотека методов системы

Метод (файл)	Вход	Выход	Параметры	Описание
<i>Matrix</i>				
sum.r	A: frame; B: frame;	M: frame	Нет	Выполняет поэлементное сложение 2 таблиц (A+B)
mul.r	A: frame; B: frame;	M: frame	Нет	Перемножает 2 числовых матрицы (AxB)
eigenValues.r	A: frame;	values: double	Нет	Вычисляет собственные значения матрицы A
<i>Data</i>				
sort.r	data: frame;	sorted: frame;	column: char;	Сортирует входную таблицу data по колонке, указанной в параметре "column"
transpose.r	M: frame;	T: frame;	Нет	Делает строки в таблицу T столбцами, а столбцы - строками
numericOnly.r	M: frame;	Numeric: frame;	Нет	Убирает все не-числовые столбцы из выборки M
parse.r	Нет	v: char	s: char;	Разбирает входящую строку s в формате "s1, s2, ..., sN" и создает из неё вектор значений v
concat.r	M: frame; N: frame;	C: frame;	Нет	Склеивает таблицы или векторы M и N в одну таблицу C
<i>Graph</i>				
matrixToGraph.r	M: frame;	From: char; To: char; Weight: double;	Нет	Строит граф G по матрице весов M. Выходная конструкция является перечислением ребер графа
deikstra.r	From: char; To: char; Length: double;	To: char; Length: double; Path: char;	From: char;	Алгоритм Дейкстры для нахождения кратчайших путей на графе. В качестве параметра указывается исходная вершина, в качестве результата - длины и пути, на которых достигаются минимальные длины
vertex_cover.r	From: char; To: char;	Vertex: char	Нет	Быстро находит вершинное покрытие графа. Вершинное покрытие необязательно будет минимальным, т.е. полученное решение - субоптимально
<i>Generators</i>				
normal.r	Нет	Random: double;	u: double; s: double; N: integer;	Генерирует N чисел из нормального распределения с МО и дисперсией s
atomic.r	atoms: double;	Random:	N: integer	Генерирует N чисел из дискретного

	p: double;	double;		распределения с заданными атомами
bernoulli.r	Нет	Random: double;	p: double; N: integer;	Генерирует N чисел из распределения Бернулли с параметром вероятности p
<i>FillGaps</i>				
put.r	data: frame;	filled: frame;	value: char;	Пытается заполнить пустые или NA поля в data с помощью значения value
linearRegression.r	data: frame;	filled: frame;	base: char;	Пытается заполнить пустые или NA поля в data с помощью построения линейной регрессионной модели на базе колонок перечисленных в base
<i>Clustering</i>				
k_means.r	data: frame;	id: integer; cluser: integer; centers: frame;	k: integer;	Кластеризация объектов из data на основе алгоритма k-средних, число кластеров определяется параметром k
ttrees.r	data: frame;	id: integer; cluser: integer;	Нет	Кластеризация на основе критерия доли объясненной дисперсии
<i>OR</i>				
knapsack.r	Item: char; Weight: integer; Price: integer;	Item: char; Price: integer;	W: integer	Решает задачу о ранце: выбирает предметы из Item, суммарный вес Weight которых не превышает параметра W, максимизируя стоимость Price
<i>SubOptimal</i>				
<i>Statistics</i>				
quantile.r				
<i>Statistics/Basic</i>				
cov.r	data: frame;	cov: frame;	Нет	Вычисляет матрицу ковариаций для числовых признаков выборки data
mean.r	sample: double;	mean: double;	Нет	Вычисляет среднее mean по выборке sample
<i>Statistics/Criteria</i>				
fisher.r	sample: frame;	val: double;	Нет	Вычисляет значение критерия Фишера для представленной выборки sample
student.r	sample: frame;	val: double;	Нет	Вычисляет значений критерия Стьюдента для представленной выборки sample
chi.r	sample: frame;	val: double;	Нет	Вычисляет значений критерия Хи-квадрат для представленной выборки sample

Информацию о методах, присутствующих в архиве планируется поддерживать в актуальном состоянии по адресу [<http://sibirica.spsl.nsc.ru>].

1.4.1. Анализ данных черепов неандертальцев и построение классификации

Настало время перейти к реализации настоящих стратегий.

В качестве задачи была выбрана все та же стратегия, описанная в [Холюшкин, Витяев, Костин, 2013]. Она была выбрана потому, что большая часть подготовительной работы уже проделана авторами этой монографии.

Напомним, что исходным материалом служили данные и выводы монографии [Деревянко, Холюшкин, Ростовцев, Воронин, 2001]. Задача звучит примерно следующим образом: по неполным исходным данным, сведенным в таблицу измерений, необходимо получить такое разбиение данных на группы, чтобы оно объясняло и подтверждало выводы [Холюшкин, Витяев, Костин, 2013], а также вычислить расхождение двух этих классификаций (исследовательской и алгоритмической). Таблица исходных данных приведена ниже (см. таблицу 8).

Сначала построим план работы стратегии.

1. Поскольку исходные данные являются неполными, а большинство существующих методов классификации подразумевают работу с цельными объектами, необходимо дополнить данные, заполнив пропуски в исходной таблице;
2. На дополненной таблице мы проведем линейный дискриминантный анализ, чтобы показать возможность разделения данных на группы;
3. С помощью алгоритма кластеризации TTrees получаем разбиение выборки Filled на группы, которые затем объединяются в теоретически предсказанные типы (получаем список групп и список типов);
4. Сравниваем разбиение, полученное на шаге 3 с теоретическим разбиением, в результате чего получаем статистику расхождений по предсказаниям типов содержащихся в таблицах групп и типов;

Предлагаемый план без труда преобразуется в схему стратегии.

Согласно спецификации, стратегия – это набор соединенных между собой блоков-методов, имеющая свой вход и выход, притом каждая связь между блоками-методами описывает, какие выходы исходных методов соответствуют входам целевых методов. Поэтому логично представить схему стратегии в виде набора методов с установленными соответствиями выходов и входов для каждой связи.

Таблица8. Исходные данные для стратегии анализа черепов

	Қазақстан Республикасының Әл-Фараби атындағы Жергілікті Атқарушы Әкімшілігі																																			
№	1	2	3	4	5	6	7	8	9	10	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	
Имя	Свансқомб	Гибралтар	Сакалосторе I	Сакалосторе II	Монте-Чирео II	Фонгешевад	Ла-Шапель	Леферасси I	ЛеМустье I	Сти I	Сти II	ЛяКнауу	Крапина D	Крапина C	Эрнстдорф I	Неацтергаль	Штайнхайм	Ганювние	Петрагона	ДжебелИрхад I	ДжебелИрхад II	Ньярасса	Брокен-Хилл	Салданыя	Схул IV	Схул V	Схул VI	Схул IX	Табун I	Зуттие	ДжебелГарафех VI	Амуд I	Шандар I	Мала	Тешик-Таш	
v	1325	1300	1200	1300	1550	1470	1626	1641	1564	1562	1600	1350	1450	1200	1450	1450	1070	1320	1220	1480	1450	1100	1325	1225	1554	1588	1400	1610	1271	1400	1568	1750	1600	1225	1545	
v1	1	2	2	1	1	1	1	1	1	1	1	2	1	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	
v3	190	182		204	195	208	208	196	201	201	202	198	178		196	199	185	200	209	198	197	191	209	200	206	192		213	183		196	215	207	196	198	
v4	183	173		193		193	196	188	187	185	187				190	186	173		194	190			188	185	193	179		200	173		190	204	197			
v5	186	175		198		197	195	189	201		199				192	198	179		206	190	193		209	200	202	190		212	179		190	208	201		187	
v6	176	173		185		190	194	184	186		185				182	185	173		196			199	192	199	183		198	174		198	202	199				
v7	176	165		184		194	198	184	182	187	184				178	183	158		190			193	192	187	174		193	174		192	202	190				
v8	142	176		142		155	154	156	159	156	144	153	138	169	149	145	147	113	145	149	145	148	142	145	144	148	144	145	141	138	145	154	152		151	
v9	125	117	109	123		117	130	135	128	144					145	147	111	115	128	125	119	109	131	128	126	130	115	130	115		139	133		139		
v10		107		125			131	138	127								113						130					131	115							
v11		93	101		111	108	111	114	114	115	115	113			105		98					96	106		112	112		113	98	116	117	121	116		125	
v12		85	79		89		90	93	91	83		79			96	83	85		83	83	84		80	84	103	95		87	84		101	101	102		93	
v13		59	51		52		51	58	56	54		47			67	54	51		54				62		66	64		53	47			59	68		108	
v14		106	102	114	115		123	120	124	107							99						112		110	98		115	108							
v15		107		106			109	109	102	107		101	108	101	110	99	113	105	102	118	117		91	99	102	106	100	95	98	97	110	115	110	104	104	
v16	117	125	116	127			124	121	121	101		108	101	110	99	113	105	102	118	117			119	119	121	114	120	122	113	125	124	125	126	126	126	
v17		125	130	145			132	143	133	124	135	112					116		138		115	132														
v18		116	116	128				135									107			112																
v19	124	110	117	112	124	126	131	125		121	131	112			105		107					132	131	132	132	122	140	120	120							
v20	40	38	35		43		49	40									39					37	40	40	40				36							
v21	30	28				32	34										578	546				30	34	29												
v22	520	520		590		503	592		589			558					538			597	582	578														
v23	522	503				503	563	534	539	546	524									570	566															
v24	276	282		360		315	322			300	316	302				380	300		308																	
v25	342	338		361		356	373										341		372				373													
v26	124	110	131		131	121	134	120	110	116		116			135	133	120	130	111			138		132	121		130	107	125	133	135	117	134	123		
v27	116		86	117		119	121	122	126	115	107				128	110	108		114	114		117		134	137		120	117		120	134		114			
v28	116	106	107	113		116	116								117	114	128	113				118		122	119		129	108		130	132		134			
v29	67	56		61		74	61	63	59	55	66				62	57	40																			
v30	50	46	52		52		43	51							57	52	44																			
v31		107	106	117		107	117	108	103			106			115	117	100					118		121		118	106		114	96	113	114	120	109	116	110
v32	108		83		109		112	113	109	115	109	103			119	103	97		106			112	112	110	122	123		112	105							
v33	94	82	87		88		91	97							87	57	69		94			92		89				95	90							
v34	61	55	51		53		66	62	57	55	53	58			58	54	65																			
v35	48	45	49		50		42	51							52	43						75.1	74.4	69.4	72.0	71.8	75.0		68.1	77.0		74.0	71.6	73.4		76.3
v36		76.8	78.0		75.0	79.0	76.0	79.0	76.5	71.6	76.5	67.7	85.4	83.7	74.0	73.9		43.0	72.5	71.3	73.2	75.1	74.4	69.4	72.0	71.8	75.0		68.1	77.0		74.0	71.6	73.4		76.3
v37		65.0	65.0		60.0	62.0	60.0	62.0	65.0	65.0	65.0	65.0					60.0	57.5	61.2	63.1	59.9	57.1	62.7	62.7	62.7	62.7	62.7		61.0	62.0		64.4	64.4		70.2	
v38	88.0	89.0	55.5		54.4	55.4	53.4	54.8	58.2	57.2	57.5	55.4			59.0		53.0		84.1	79.3	85.9	86.2	79.7	76.8	90.3	86.5	87.5		89.7	81.6		90.3	87.5		93.1	
v39		80.1	76.8		79.4	76.0	83.3	85.4	85.3								84.1	79.3	85.9	86.2	79.7	76.8	90.3	86.5	87.5		89.7	81.6		90.3	87.5		93.1			
v40		63.7	71.1		71.6	70.1	71.2	72.7	76.0	79.9	75.2	81.9			70.5		74.2		40.3	43.7	43.5	69.0	73.1		75.7	77.8		77.9	69.5	84.1	80.1	78.6	76.3		82.8	
v41		45.7	45.1		45.0		45.7	47.7	48.2	41.3		39.7			50.0	41.9	47.5		40.3	43.7	43.5	69.0	73.1		75.7	77.8		77.9	69.5	84.1	80.1	78.6	76.3		82.8	
v42		33.5	29.5		28.1		26.8	29.0	30.4	29.0	25.4				36.8	29.2	29.5		27.6			64.1	67.6	70.2	71.6	69.4		66.2	69.5	70.3	75.9	74.7	72.4		68.9	
v43		73.3	71.1		68.4		69.9	69.0	71.3	70.1	70.6	73.2	65.1	66.4	77.9	71.4	71.4		74.5			64.1	67.6	70.2	71.6	69.4		66.2	69.5	70.3	75.9	74.7	72.4		68.9	
v44		43.3	47.9		43.1		48.4	46.9	46.2			44.4			83.4	86.1	83.4		87.6	87.7		82.4	87.6	87.7		82.4	87.6		82.4	87.6		82.4	87.6		82.4	
v45		97.3	86.3		85.5		83.8	86.2		83.5	82.4	90.2			107.6		95.3					88.9	74.8		80.3	82.0		80.0	80.3		80.0	80.3		80.0	80.3	
v46	94.4	113.6	99.2		102.4		94.7	96.8		115.2		110.2			110.2		110.2					90.8	91.7	93.4		93.4		100.0	101.7		101.6	100.0	100.0		100.0	
v47		55.6	53.8		56.5		58.4	55.8							51.6		54.3		56.2			56.0		53.1	50.9		56.2	55.0		53.3	55.8	55.5		53.5		
v48		30.1	30.2		31.9		34.6	32.2							29.0		29.6					30.0		28.4	26.0		30.3	32.4								
v49		56.3	58.2		57.4		51.4	56.3	55.1	51.2		52.0			60.7	58.8	54.0		50.7		59.9	57.9		53.3	55.2		52.5	57.5		58.2	55.8	57.2	59.2			
v50		45.8		53.4		53.9	54.3	55.6	57.2	54.5	50.5				58.7	52.0	52.4		50.7		57.4	53.6		59.2	64.1											

По схеме стратегии почти автоматически строится стратегия в визуальном редакторе. Сценарий подробно описан в разделе "Пример разработки стратегии" [ИТГИ, 20]. Суть идеи в том, что следует добавить одноименные блоки на канву стратегии, расположить их аналогично схеме и достроить граф стратегии до графа, изображенного на схеме.

Для примера можно рассмотреть схему стратегии из раздела "Пример разработки стратегии". На блок-схеме помимо графа стратегии представлены соответствия выхода-входа (рис.23):

Уточним план, предложенный выше, но уже в терминах схемы стратегии (рис.24):

1. На вход стратегии (блок "Начало TTreeCluster.rs") подается переменная с именем data и типом frame (таблица);
2. Вход стратегии передает data методу "Заполнение пропусков";
3. "Заполнение пропусков" дополняет данные (здесь использован метод /public/FillGaps/LinearRegression.r, следуя выкладкам из [Холюшкин, Витяев, Костин, 2013]), генерируя выходную таблицу data типа frame;
4. "Заполнение пропусков" передает полную таблицу data методу "Линейный дискриминантный анализ";

5. "Линейный дискриминантный анализ" вычисляет факторные нагрузки для выборки data и передает результат – factors типа frame–на выход;
6. Те же самые данные "Заполнение пропусков" передает в "Кластеризация, затем сборка";
7. Наконец, "Заполнение пропусков" обеспечивает теоретической типизацией заключительный метод "Проверить кластеризацию";
8. "Кластеризация, затем сборка" выполняет разбиение на группы, согласно алгоритму TTree, получая разбиение исходной выборки на группы groups, а затем пытается объединить их в типы types. Обе переменные groups и types являются списками. Метод расположен в архиве/public/clustering/ttrees.r;
9. "Кластеризация, затем сборка" передает эмпирически полученную типизацию types контрольному методу "Проверить кластеризацию" под именем classes;
10. "Проверить кластеризацию" сравнивает 2 типизации извлеченные из data и types и выдает таблицу-результат сравнения на выход (блок "Конец TTreeCluster.rs").

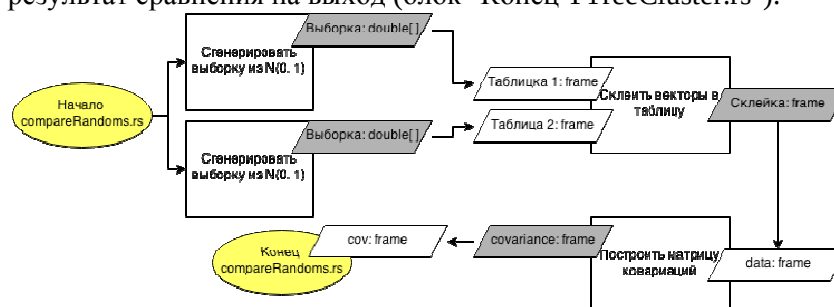


Рис.23. Схема тестовой стратегии.

Разберем немного подробнее методы, используемые в рамках данной стратегии.

Первым на очереди метод является метод "Заполнение пропусков". Основой для метода является идея о том, чтобы представить целевой вектор значений v и представить как линейную комбинацию базисных векторов x и y :

$$v = \alpha x + \beta y + \gamma \quad (1)$$

Рассматриваемый метод строит линейные модели вида (1), аппроксимирующие существующие (уже заполненные) значения вектора v . Для этого:

1. Строится источник линейной модели, задаваемый пользователем как база линейной регрессии в виде списка колонок исходной таблицы данных:
`regressionNames = c(base1, base2);`
`regressionBasis = input[regressionNames];`
2. Для каждой колонки с отсутствующими значениями вычисляется линейная регрессионная модель вида (1) с помощью R-функции `lm`:
`for (name in names) {`
`v <- input[[name]];`
`model <- lm(x ~ ., data.frame(x = v, regressionBasis));`
3. Выполняется предсказание согласно полученной модели `model` и строится вектор предсказанных значений. За это отвечает функция R-языка `predict.lm`:
`predictedValues <- predict.lm(model, regressionBasis);`
4. Отсутствующие значения вектора v дополняются предсказанными значениями:
`missed = is.na(v);`
`v[missed] <- ifelse(`
`typeof(v) == "integer",`
`as(predictedValues[missed] + 0.5, typeof(v)),`
`as(predictedValues[missed], typeof(v));`

Линейный дискриминантный анализ построен на принципе разделения объектов с помощью введений нескольких (k) интегральных линейных показателей в k -мерном пространстве состояний R^k :

1. Из данных выбрасываются все нечисловые значения, а остальные приводятся к `double`:
`toAnalyze <- data[, sapply(data, is.numeric)]`
`toAnalyze <- as.data.frame(sapply(toAnalyze, as.double));`
2. Вычисляем факторные нагрузки для линейной дискриминации объектов в факторном пространстве, в этом нам поможет функция языка `lda`:
`fa <- lda(Type ~ ., toAnalyze);`

3. Строим векторы значений согласно построенным моделям факторных нагрузок:
`factorVals<- predict(fa)$x;`

Особое внимание следует уделить методу кластеризации. Авторский метод TTrees был предложен П.С.Ростовцевым [Ростовцев и др., 1994]. Суть идеи состоит в постоянном разбиении множества объектов на группы таким образом, чтобы это разбиение отличалось наибольшей объяснительной критериальной способностью.

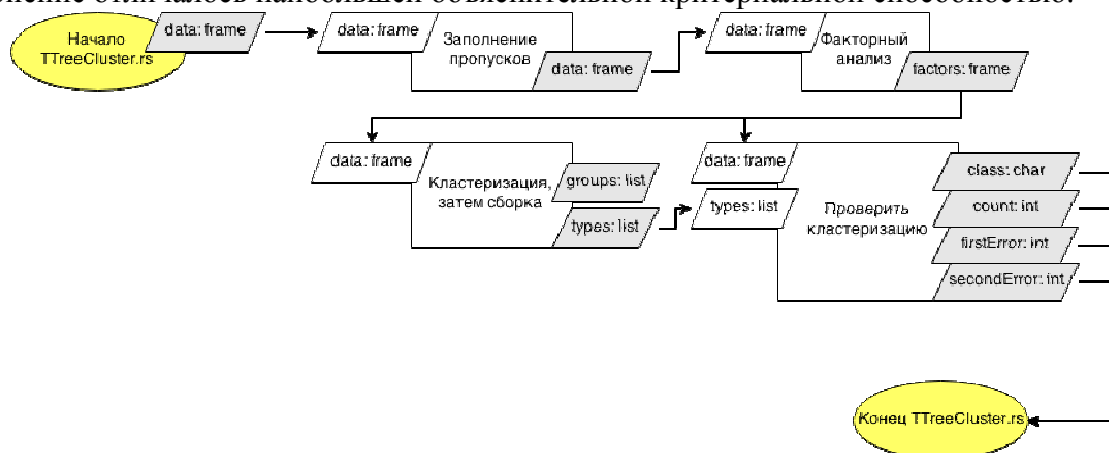


Рис.24. Схема стратегии TTreeCluster.rs.

Разбиение является хорошим, если оно имеет максимальную объясненную, или межгрупповую дисперсию (и, соответственно, минимальную необъясненную, или внутригрупповую дисперсию). Отметим, что разбиения продолжаютсЯ до тех пор, пока размер группы не окажется пренебрежительно мал, либо группа станет однотипной с точки зрения теоретической таксономии.

Авторы попытались воспроизвести работу оригинального метода кластеризации при программировании метода-аналога на R языке:

1. Если размер текущей группы ≤ 5 (в этом случае при разбиении получилась бы подгруппа неудовлетворительно малого размера $=1$) или группа полностью гомогенна, то она оставляется без изменений и возвращается "как есть":
`if (nrow(group) <= 5 | length(unique(group[[Y]])) == 1)`
`return(list(list(path=path, data=group)))`;
2. Иначе ищется оптимальное с точки зрения критерия дробление группы и для каждой подгруппы повторно запускается процедура дробления:
`criteria<- sapply(X, varMax, group = group);`
`criteria<- criteria[,which.max(apply(criteria, ...))];`
`dataPieces<- c(split(group[criteria$permutation,], ...));`
`result<- list();`
`for (dataIndex in 1:length(dataPieces))`
`result<- c(result, breakGroups(...,dataPieces[[dataIndex]]));`
3. Функция varMax из листинга п. 2 выполняет поиск оптимального разбиения для каждого признака присутствующего в переданной выборке. Результат возвращается в виде списка описаний, содержащих:
 - 3.1. Сортировку объектов в виде отображения-подстановки (permutation);
 - 3.2. Значение критерия объясненной дисперсии (fValue);
 - 3.3. Размеры получающихся при дроблении групп (lengths);
 - 3.4. Границы значений признака, разбивающие весь интервал значений признака на группы (borders);`permutation<- order(groups[[attribute]]);`
`dataVector<- groups[permutation, attribute];`
`borderCombos<- subset(expand.grid(`
`border2=3:(length(permutation)-3),`
`border3=5:(length(permutation)-1)),`
`border3 - border2 > 1);`
`criteria<- apply(borderCombos, 1, checkGrouping,`
`target = groups[permutation, Y]);`
`opt<- criteria[[which.max(sapply(criteria, ...))];`
`opt$borders<- cbind(`
`start=dataVector[c(1, cumsum(head(opt$lengths, -1)) + 1)],`
`end=dataVector[cumsum(opt$lengths)]);`

В листинге borderCombos – это все возможные разбиения, по которым будет искаться оптимум, а apply несколькими строками ниже получает значение критерия для всех таких разбиением поочередным применением функции checkGrouping к таблице разбиений;

4. Функция checkGrouping вычисляет межгрупповую дисперсию для представленного разбиения на прямую:

```
fVal<- 0; start_next<- 1; avg<- mean(target);
lengths = c(
  grouping[['border2']] - 1,
  grouping[['border3']] - grouping[['border2']],
  length(target) - grouping[['border3']] + 1);
for (group in 1:length(lengths)) {
  start<- start_next;
  start_next<- start + lengths[group];
  part<- target[start:(start_next - 1)];
  fVal<- fVal + lengths[group] * (avg - mean(part)) ^ 2;
}
```

Метод "Кластеризация, затем сборка" немногим отличается от описанного выше метода. Помимо операций разбиения исходной выборки на группы, присутствует часть кода, отвечающая за сборку воедино тех групп, которые относятся к одному типу. В итоге получается цельный список групп, образующих эмпирический тип.

```
for (groupinggroups) {
  groupData<- group$data;
  counts<- as.data.frame(table(groupData[['Type']]));
  classIdx<- counts[which.max(counts[, 'Freq']), 'Var1'];
  classIdx<- as.character(levels(classIdx)[classIdx]);
  ids<- as.integer(rownames(groupData));
  foundClasses[classIdx] <- c(foundClasses[classIdx], ids);
}
```

"Проверка кластеризации" заключается в сравнении двух таксономий и выдает числовые показатели, описывающие расхождения в количествах объектов, попадающих под тот или иной тип.

Запрограммировав все необходимые методы, можно приступить к преобразованию схемы стратегии с Рис. 23 в исследовательскую стратегию в рамках системы анализа данных. Для этого следует обратиться к руководству раздела "Техника разработки стратегии" (рис.24).

Подведем некоторый итог. Что у нас на входе? Согласно схеме Рис. 23, при запуске стратегия получает набор данных data: frame с пропущенными значениями. А на выходе? После исполнения, стратегия должна вывести нам на экран сводную таблицу со сравнением двух классификаций: теоретической и эмпирической. В таблице должны присутствовать целочисленные колонки class, count, firstError, secondError.

Что ж, можно и проверить. Для этого выберем файл с исходными данными. Это таблица 6, несколькими страницами выше. Чтобы стратегия могла нормально функционировать, необходимо загрузить файл с этой таблицей на сервер, воспользовавшись архивом в разделе "Данные" (рис.25):

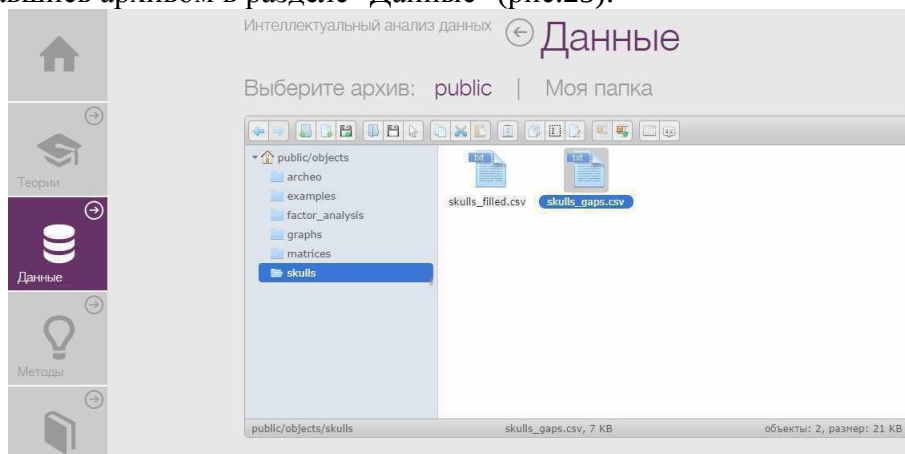


Рис. 25. Загрузка данных для обработки стратегией.

Запускаем стратегию на свежих выгруженных данных.

Результат работы стратегии представлен на изображении ниже. В целом, это именно то, чего мы ожидали, исходя из схемы стратегии (рис.26):

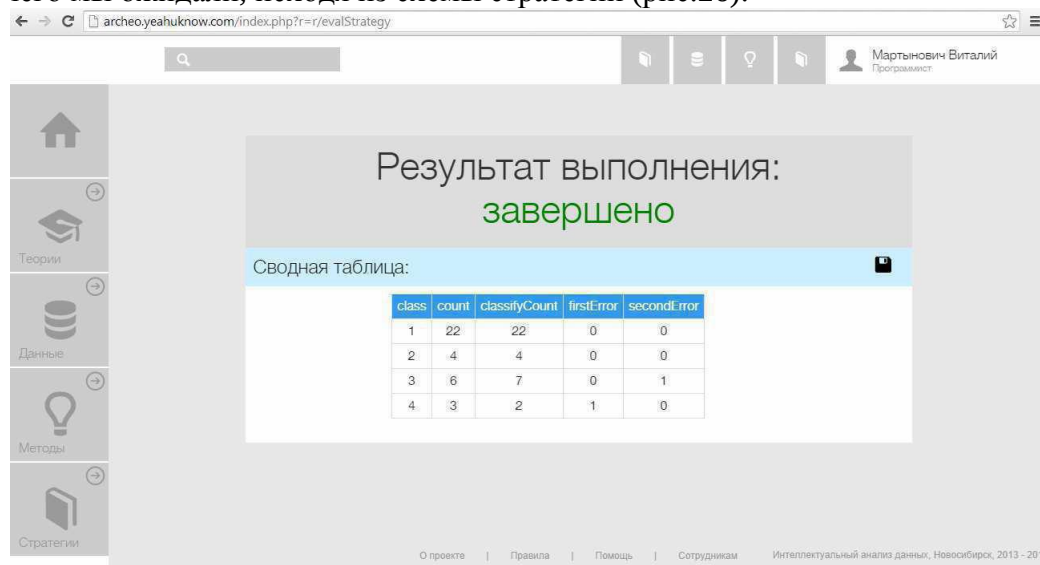


Рис.26. Результаты анализа данных по черепам.

Из протокола можно увидеть, что почти все объекты были распределены правильно. Ошибочно определен только один объект, попавший вместо теоретически предсказанного типа 4 в тип 3. Все остальные объекты заняли своё место в эмпирической таксономии, совпадающей с таксономией теоретической.

Данные по классификации черепов (указаны идентификаторы элементов исходной таблицы):

```

Тип `1` 4 22 28 29 16 7 10 5 14 12 18 2 6 13 21 8 11 15 3 9 17 20
Тип `2` 23 19 34 24
Тип `3` 1 27 30 35 25 31 26
Тип `4` 33 32

```

Таким образом, можно сказать, что теоретическая таксономия является абсолютно корректной и обоснованной. Заинтересовавшийся читатель может получить обоснование полученной иерархии групп, используя алгоритм TTrees из одноименного метода системы. TTrees дает разбиение исходной выборки в виде ветвящегося дерева, где каждая группа – это лист дерева, а метки на пути дерева позволяют восстановить «объяснение» произведенного разбиения в виде накладываемого на тот или иной признак ограничения. Ниже можно увидеть пример такого разбиения, произведенного методом /public/Clustering/TTrees.r (рис.27):

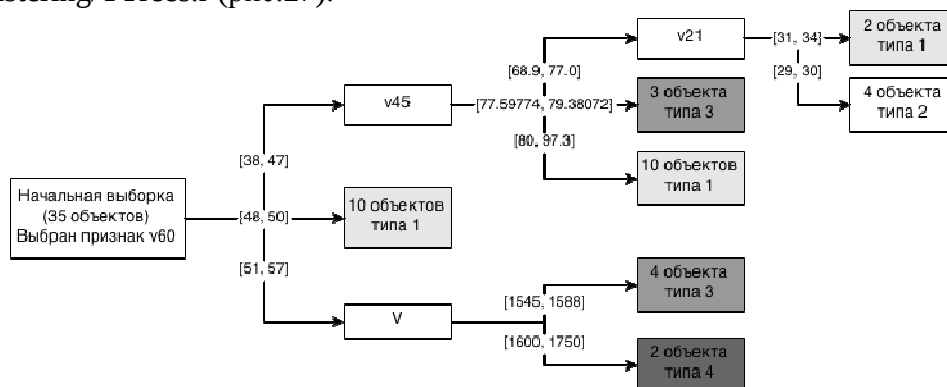


Рис. 27. Разбиение на группы для исходных данных.

В листьях дерева содержится описание полученных групп. Метки на не листовых узлах в совокупности с метками на ветвях дерева дают описание группы в терминах исходных признаков. Цветом подсвечены группы, которые впоследствии будут собраны в единый тип.

1.4.2. Программирование стратегии для исследования данных по памятникам палеолита

Для понимания и установления природы и характера различий палеолитических комплексов большое значение имеет исследование структурных характеристик археологических данных. Основой для исследований служит типологическая таблица, подготовленная авторами оригинальной работы [Деревянко и др., 2005] на основе данных, собранных по материалам публикаций (см. таблицы 9 и 10).

Таблица 9. Список типов орудий из комплексов Ближнего и Среднего Востока и Кавказа.

1	леваллуазские сколы	17	скребки	33	рабо
2	леваллуазские острия	18	резцы	34	орудия с черешком
3	леваллуазские ретушированные острия	19	проколки	35	чопперы
4	псевдолеваллуазские острия	20	ножи	36	чоппинги
5	мустьерские острия	21	ракле	37	разные
6	Лимасы	22	усеченные отщепы	38	бифасы листовидные
7	продольные скребла	23	транше	39	угловатые
8	двойные скребла	24	выемчатые	40	долотовидные
9	конвергентные скребла	25	зубчатые	41	клововидные
10	угловатые скребла	26	резцовые острия	42	бифасы
11	поперечные скребла	27	сколы с брюшковой ретушью	43	микроорудия
12	скребла с брюшковой ретушью	28	сколы ретушированные со спинки	44	диски
13	скребла с крутой ретушью	29	остроконечники тейяские	45	кливер
14	скребла с утонченной спинкой	30	треугольные орудия с выемкой	46	пластины с притупленным краем
15	двусторонние скребла	31	псевдорезцы	47	Nachrlbragim.
16	скребла с противолежащей ретушью	32	сколы с выемчатым концом		

Совокупность типы орудий – локации будут объединены в классы, объясняющие их схожесть. На основе полученной классификации могут без труда быть получены выводы о культурных и семантических взаимосвязях рассматриваемых объектов.

Таблица10. Список орудийных комплексов Ближнего и Среднего Востока и Кавказа

1	Амуд В4	14	ВарвазиА	27	Ябруд 4	40	Монашеская	53	Ортвала-Клде V
2	Амуд В2	15	Варвази В	28	Ябруд 5	41	Губский Навес	54	Ортвала-Клде VI
3	Кеу I сл. I	16	Варвази С	29	Ябруд 6	42	Малая Воронцовская	55	Ортвала-Клде VII
4	Кеу I сл. II	17	Варвази D	30	Ябруд 7	43	Таглар 2 сл.	56	Двойной Грот
5	Кеу I сл III	18	Сефуним А	31	Ябруд 8	44	Таглар 3 сл.	57	Азых 3 сл.
6	Кеу I сл V	19	Сефуним 12	32	Ябруд 9	45	Таглар 4а	58	Среднехаджохская
7	Кзар-Акил XXVIA	20	Сефуним 13	33	Ябруд 10	46	Таглар 4б	59	Азых 6 сл.
8	Кзар-Акил XXVIB	21	Сефуним VI	34	Кударо I За	47	Таглар 5	60	Медвежье
9	Кзар-Акил XXVIA	22	Сефуним VII	35	Кударо I 3б	48	Таглар 6	61	Лусакерт D
10	Кзар-Акил XXVIB	23	Сефуним В	36	Кударо I 3в	49	Ортвала-Клде I	62	Лусакерт А
11	Кзар-Акил XXVIII	24	Сефуним С	37	Кударо I 4	50	Ортвала-Клде II	63	Газма
12	Кзар-Акил XXVIII	25	Ябруд 2	38	Каркустакау	51	Ортвала-Клде III	64	Баракаевская
13	Кунджи	26	Ябруд 3	39	Тамарашени	52	Ортвала-Клде IV		

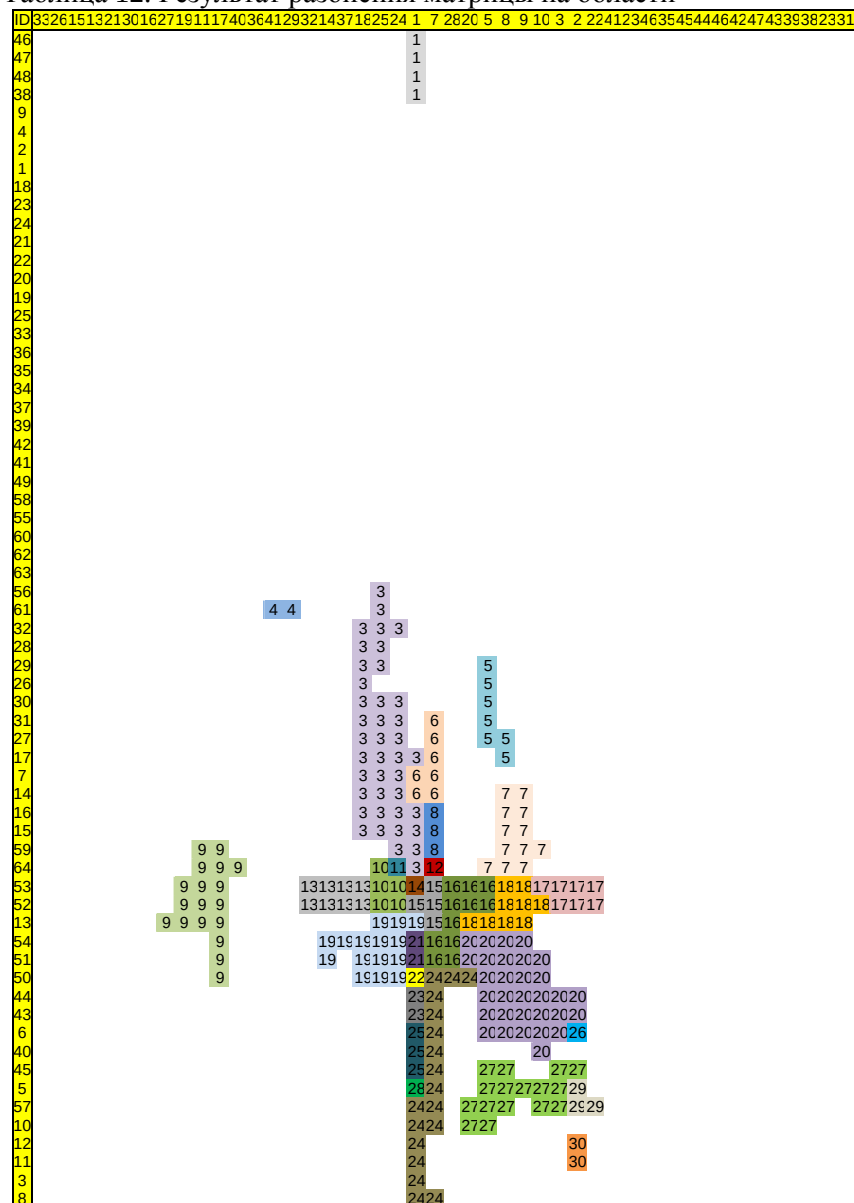
Теперь можно сформулировать исследовательскую задачу, а вместе с ней и задачу разрабатываемой стратегии. Имея набор типов орудий (таблица 9) и данные по количеству типов орудий в перечисленных локациях (таблица 10) получить классы орудий, связывающие родственные типы и культурно близкие локации. Класс должен представлять собой набор пар (тип орудия, локация), а также правила, описывающие внутреннюю структуру такого класса.

Типы орудий были перенумерованы и представлены в таблице 9, мустьерские комплексы (локации) сведены в таблицу 10. В данных, используемых для исследовательской стратегии, содержатся уже лишь закодированные объекты, идентификаторы в первом столбце этой таблицы (таблица 11а) соответствуют локациям из таблицы 10, идентификаторы в первой строке – это типы орудийных комплексов из таблицы 9.

План исследований в рамках стратегии может быть следующим:

1. Как видно из таблицы 11а, многие ячейки оставлены пустыми. В данном конкретном случае это означает, что они просто отсутствуют. Для работы алгоритма необходимо заполнить эти ячейки нулями;

Таблица 12. Результат разбиения матрицы на области



5. "Перестановка строк" принимает таблицу (frame) data и действует следующим образом. Во-первых, по исходной таблице строится матрица расстояний между строками (расстояние между двумя строками вычисляется по стандартной евклидовой метрике). Далее исходная задача сводится к задаче коммивояжера, если положить строки – вершинами графа, а веса ребер – расстояниями между строками. В граф вводится "нулевая" вершина. После этого запускается алгоритм решения задачи коммивояжера, который дает субоптимальное решение path. Перестановка строк таблицы data согласно порядку строк в path есть конечный результат работы метода ordered.
6. "Перестановка строк" передает таблицу ordered методу "Транспонировать таблицу". Это сделано для того, чтобы вновь воспользоваться перестановкой элементов таблицы, но уже для столбцов а не для строк;
7. "Транспонировать таблицу" превращает столбцы матрицы в строки и инициирует исполнение второй части сортировки, выполняя метод "Перестановка строк (2)";
8. "Перестановка строк (2)" отдает ordered методу "Транспонировать таблицу", которая возвращает таблицу в исходное состояние;
9. Все готово к выполнению метода "Разбиение на области". Разбиение на области получает упорядоченную таблицу (frame) areas и производит на ней кластеризацию (см. описание ниже).
10. Итог выполнения этого метода и всей стратегии – таблица classes, в которой каждой ячейке указан номер класса, к которому она принадлежит.

Таблица 13. Вклад областей в объяснение полученного разбиения

Область	Объясняет, %	Область	Объясняет, %
15	18.05	20	0.98
23	16.18	1	0.97
14	15.76	6	0.71
25	8.84	29	0.70
16	6.23	7	0.50
11	4.26	26	0.48
22	3.45	19	0.42
24	3.28	9	0.25
2	2.63	27	0.25
21	2.36	30	0.24
18	2.30	4	0.22
8	1.83	3	0.18
12	1.73	17	0.16
10	1.31	5	0.15
28	1.26	13	0.14

Основная сложность при разработке такого рода стратегий состоит в использовании нестандартных методов. В данном случае таковым является метод "Разбиение на области" (запрограммированный метод выложен в общий доступ по адресу /public/Data/AreaCluster.r). Для нестандартных методов необходима процедура программирования этих методов на R-языке с учетом спецификации системы. Все остальные части стратегии выбираются из общедоступного архива и разработка стратегии проводится в духе системы интеллектуального анализа данных на основе визуального редактора.

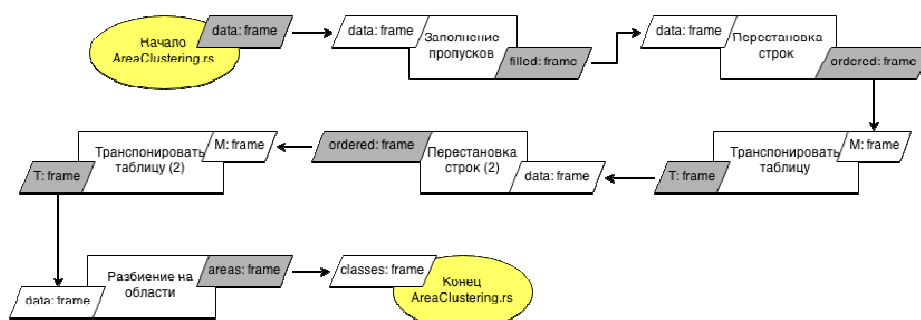


Рис. 28. Схема стратегии AreaClustering.rs.

Метод "Разбиение на области" начинает свою работу с процедуры инициализации.

```

v <- list(); e <- list(); nu <- mean(input);
addEdge <- function (v1, v2) {
  e[[v1]] <- c(e$v1, v2);
  e[[v2]] <- c(e$v2, v1); }
idF <- function(i, j) { return(paste('v_', i, '.', j, sep='')); }
  
```

Изначально вся таблица представляется в виде графа, где вершины – это ячейки таблицы, а ребра соединяют две вершины тогда и только тогда, когда они являются в таблице соседними:

```

for (i in 1:nrow(input)) { for (j in 1:ncol(input)) { #ячейки
  v[[idF(i,j)]] <- data.frame(id=idF(i,j), val=input[i, j]);
  if (j<ncol(input)) { # есть сосед справа
    v1<-idF(i,j); v2<- idF(i,j+1); #имена вершин на ребре
    e[[v1]] <- c(e[[v1]], v2); e[[v2]] <- c(e[[v2]], v1); }
  if (i<nrow(input)) { # есть сосед снизу
    v1<-idF(i,j); v2<- idF(i+1,j); #имена вершин на ребре
    e[[v1]] <- c(e[[v1]], v2); e[[v2]] <- c(e[[v2]], v1); }
  }}
  
```

Затем перебираются все доступные ребра и выбирается такое, для которого изменение суммарной объясненной дисперсии по итоговому разбиению – максимально:

```

q <- function(val) { return((mean(val) - nu)^2 * length(val)); }
edgeDelta <- function (v1, v2) {
  return(q(c(v1$val, v2$val)) - q(v1$val) - q(v2$val)); }
maxCrit <- function () {
  emax <- list(h='', m='', crit=-Inf);
  for (from in names(v)) { for (idx in 1:length(e[[from]])) {
    to <- e[[from]][idx]; #конец ребра
  
```



```

d<- edgeDelta(v[[from]], v[[to]]); #оценка критерия
  if (d> emax$crit) { emax <- list(h=from, m=to, crit=d); }
if (d == 0) return(emax);# дисперсия не может возрасти
}}
return(emax);
}
Наконец, выполняется слияние вершин графа по оптимальному ребру:
v[[host]] <- rbind(v[[host]], v[[merged]]); # объединить области
e[[host]] <- setdiff(c(e[[host]],e[[merged]]), c(host,merged));
v[[merged]] <- NULL; e[[merged]] <- NULL; #убрать вершину
if (length(v) <= 15) break;# повтор пока более 15 областей
Оставшаяся часть стратегии отвечает за отрисовку значений в итоговую матрицу:
for (vert in names(v)) {
  cells <- v[[vert]];
  for (i in 1:nrow(cells)) {
    cell <- cells[i,];
    input[cell$x, cell$y] <- vert;
  }
}

```

С готовой схемой и набором запрограммированных методов достаточно просто реализовать стратегию. Визуальный редактор служит прекрасным инструментом для конструирования пользовательских стратегий:

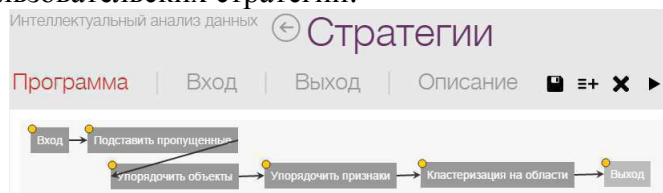


Рис 29. Реализация стратегии в визуальном редакторе стратегий

Результаты, полученные нами для приведенной стратегии представим в виде двух таблиц. Первая, таблица Ordered, получена на стадии перестановки строк и столбцов исходных данных, с целью минимизации суммарных расстояний между строками и столбцами таблицы. Вторая, Classified – результат процедуры классификации, каждая ячейка этой таблицы соответствует ячейки таблицы Ordered. Внутри ячеек указан класс, к которому эта клетка принадлежит.

В таблице ниже представлены полученные в результате кластеризации области (30 штук). Каждая область окрашена в свой цвет, который соответствует цвету в таблице 10. Область 2 оставлена пустой для простоты восприятия (белое пространство).

Интересен также анализ вкладов различных выделенных областей в общее распределение объектов. Для этого проанализируем таблицу долей объясненной дисперсии (табл. 13), которая приходится на ту или иную область. Таким образом, мы сможем показать, какие области объясняют полученное распределение хорошо, а какие – нет.

Следует отметить, что области выделялись таким образом, чтобы максимизировать суммарную объясненную дисперсию. Каждое объединение областей в алгоритме метода AreaCluster.r сопровождается потерей объясненной дисперсии, и эта потеря должна быть минимальной. Соответственно области – это наиболее экономные объединения объектов вида (Тип орудий, Локация) с точки зрения потерь объясненной дисперсии. Итоговая объясненная дисперсия составляет 95.8%, а её потери – 4.2%.

1.4.3. Анализ научных течений в новой археологии

Следуя за идеями, изложенными в работе [Холушкин, Витяев, Костин, 2013], мы воспроизведем построение классификации различных археологических школ и течений. Суть идеи состоит в выделении трех достаточно ясно выраженных направлений на основе данных о взаимном цитировании авторов.

За основу была взята таблица 14 с количеством постраничных ссылок одного автора (указанного в строке таблицы) на другого (колонка таблицы). План разработки следующий:

1. Стратегия bib.rs получает на вход таблицу (frame) перекрестных ссылок;

2. На полной таблице мы производим дискриминантный анализ, получая дополненную таблицу с возможностью выделения кластеров;
3. Строим точечный график, которые показал бы распределение объектов (авторов) в пространстве выделенных факторных нагрузок;
4. Анализируем график визуальнo, убеждаясь в качестве дискриминантного анализа;
5. Проводим повторный анализ с помощью нейронных сетей;
6. Сравниваем качество предсказанных нейронными сетями классов и теоретических;

Таблица 14. Исходные данные для стратегии Bib.rs

№	Автор	Класс	Бинфорд	Хилл	Фриц	Плог	Уотсон	ЛеБланк	Редмен	Уоллон	Стрювер	Лион	Лонгакр	Мартин	Айзек	Доран	Кларк Д.	Фленнери	Ренфру	Дитц	Рауз
1	Бинфорд	1	82	38	12	6	5	1	2	1	3	5	2	6	9		34	3	24		2
2	Хилл	1	3	39	4	16	1		2		1	1		1			2		2		
3	Фриц	1		12	5		3		1					1		1	1	1			
4	Плог	1		4	1	1	3		1					1		1	5	1			
5	Уотсон	1	3	2	2		21	2	4		5		2				4	5			3
6	ЛеБланк	1		2	2		8		3		4		1				4	4	3		2
7	Редмен	1		2	2		1	1	4		4		1				4	4	2		3
8	Уоллон	1	2	8	1	1				1		1					1	1			
9	Стрювер	1	2	1	1				1		2				3		1		1		
10	Лион	1			6	2	1		1									1	2		
11	Лонгакр	1	4	12	4	12								1			1	1	1		
12	Мартин	1	1	1	5	15	3							6			4		2		
13	Айзек	2	1				2								8		4				
14	Доран	2														5	17		1		
15	Кларк Д.	2	8	7	1	5	1				1		2		3	3	138	5	6		1
16	Фленнери	3	6	2	2	1	1		2	1		3					3	17	27		4
17	Ренфру	3							1						2		8		8		
18	Дитц	1	13	8		1						2			1		9	2	1		
19	Рауз	1	12	57	1	2							3			1	5	1			

Схема для Bib.rs достаточно очевидна (рис.30):

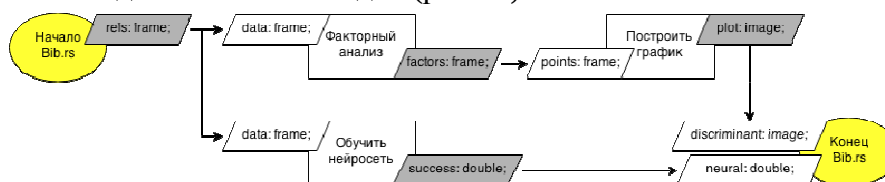


Рис. 30. Схема стратегии Bib.rs

Для дискриминантного анализа был применен уже готовый метод из системы public/Clustering/factor.r. Его результат (factors: frame) был передан следующему методу для отрисовки графика:

```

data<- input$data;
now<-format(Sys.time(), "%b%d%H%M%S");
plotFile<- paste("plot", "-", now, ".png", sep="");
png(filename=plotFile);
plot(data);
dev.off();
result<- list(plot=plotFile);
  
```

Метод plot.r рисует график по переданному ему набору точек и отдает на выход имя графического файла, содержащего этот график. На этапе дискриминантного анализа была получена следующая картина (рис.31).

Диаграмма распределения классов на основе дискриминантного анализа показывает полное соответствие предсказанных классов с классами теоретическими. В данном случае дискриминантный анализ оказывается в состоянии построить адекватную классификацию и является абсолютно состоятельным.

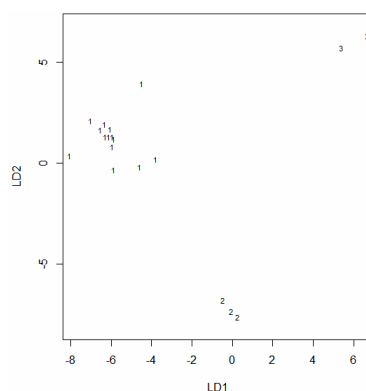


Рис. 31. Диаграмма распределения классов после дискриминантного анализа

Сравнимся теперь с нелинейными по своей природе моделями, построенными на основе нейронных сетей:

Как видно из таблицы 15, нейронные сети достаточно плохо справляются с задачей предсказания на таком маленьком числе объектов. Путем оптимизации параметров архитектуры нейронной сети максимальное чего удастся достичь – это точность порядка 84%.

Таблица 15. Результат метода neural_nets.r

№	ФИО	Класс	Предсказан
1	Бинфорд	1	0.999997
2	Хилл	1	1.000007
3	Фриц	1	0.999999
4	Плог	1	1.000001
5	Уотсон	1	1.000002
6	ЛеБланк	1	1.142861
7	Редмен	1	0.898387
8	Уоллон	1	0.999999
9	Стрювер	1	1.000078
10	Лион	1	0.999998
11	Лонгакр	1	1.000003
12	Мартин	1	1.000015
13	Айзек	2	2.999974
14	Доран	2	2.333327
15	Кларк Д.	2	2.906050
16	Фленнери	3	0.999999
17	Ренфру	3	2.997197
18	Дитц	1	1.011620
19	Рауз	1	1.000000

Сам метод выглядит следующим образом:

```
library("neuralnet");
options(stringsAsFactors=FALSE);
tr.in.init<- input;
tr.out.init<- input[target];
tr.n<- setdiff(colnames(tr.in), c(target));
f <- as.formula(paste(target, '~', paste(tr.n, collapse = '+')));
result<-data.frame(id=character(0),exp=integer(0),n=integer(0));

for (control in 1:nrow(input)) {
  tr.in <- tr.in.init[-control,];
  tr.out<- tr.out.init[-control,];
  neuro<- neuralnet(f, data=tr.in, hidden=3, threshold=0.0001)
  testdata<- tr.in.init[control,-c('Class')];
  tr.predict<- compute(neuro, testdata) #Run
  net.res <- list(
    id=input[control,'Name'],
    exp=input[control,'Class'],
    n=tr.predict$net.result);
  result<- rbind(result, net.res);
}
```

Алгоритм достаточно прост: из выборки выкидывается один объект, а для остальных строится модель на основе нейронной сети. Затем, для исключенного объекта

выполняется предсказание на основе полученной модели. Иначе говоря, выполняется скользящий контроль. Результаты всех шагов сводятся в единую таблицу, которая дана выше как Таблица 15.

При рассмотрении выше описанных работ было бы несправедливым не упомянуть о двух публикациях сотрудников НГУ А.Ю. Борисенко и Ю.С. Худякова, написанных в соавторстве с математиками [Борисенко и др., 2002: с.21-36; Борисенко и др., 2008: с. 90-98].

Первая публикация посвящена поиску закономерностей в истории природных аномалий и катастроф на территории Северного Китая и Кореи в древности и средневековье. В ней благодаря «проведению математического анализа сведений о катастрофах, содержащихся в корейских и китайских летописях, удалось выяснить внутренние причинно-следственные связи между различными природными явлениями. Эти связи позволяют построить математические модели, с помощью которых можно будет выявить долговременные тенденции в периодичности аномальных природных явлений и прогнозировании масштабных последствий для населения и природной среды Земли» [Борисенко и др., 2002: с.36].

Вторая статья посвящена проблемам классификации панцирных пластин номадов Центральной Азии. Соавторы пытались осуществить свои разработки путём представления результатов в виде логических высказываний, отражающих причинно-следственные связи между характеристиками на языке, близком к естественному языку логических суждений.

Отличительной особенностью этих методов является возможность построения логико-вероятностной модели изучаемых объектов, представленной в виде списка логических закономерностей ("знаний"). Примененные методы позволили, в частности, проводить анализ форм защитного покрытия [Борисенко и др., 2008: с.90-98].

2.0. Проблемы автоматизации информационных потоков в археологии

2.1. Введение

С позиций системологии любая научная дисциплина в процессе развития проходит путь от простейших описаний предмета эмпирического исследования, через все большую детализацию, уточнение, специализацию собственного профессионального языка, до того состояния, когда становится возможным описывать изучаемые явления и процессы предельно конкретно и точно, используя языки все более абстрактных разделов математики.

Каждый шаг вперед на этом пути есть результат приобретения новой информации об изучаемом объекте: сами по себе математические методы не дают новой информации; они лишь позволяют преобразовать имеющуюся в постановке задачи информацию в такой вид, который более обозрим, удобен, полезен, пригоден для наших целей [Тарасенко, 2000: с.84].

Исходным понятием в этом ряду являются *технологии создания археологической информации*. Они охватывают все этапы и стадии формирования первичных данных, возникающих в процессе полевых, камеральных и кабинетных археологических исследований, включая работу с литературными источниками. Наиболее важными критериями целесообразности применения подобных технологий является полнота, достоверность и адекватность формируемых в исследовательском процессе данных. Все эти технологии в археологии необходимо разумно основывать на принципах и подходах, выделенных Л.С. Клейном [Клейн, 2001]. Среди этих принципов можно отметить следующие:

1. Принцип всесторонности, основанный на максимально возможном охвате информации.
2. Принцип комплексности, выраженный в преодолении тяги к уникальным вещам и поднятии интереса к массовому рядовому материалу.
3. Принцип универсальной подготовленности, согласно которому каждый археолог должен иметь представление о тех науках, специалистам которых он передает для анализа свой материал.
4. Стратиграфический принцип, под которым понимается возможность выявления относительной хронологии вскрываемых археологических материалов на основе их последовательности во времени.
5. Структурный принцип и принцип сопряженности. Первый из этих принципов подразумевает локальное взаиморасположение находок в археологическом комплексе, а второй – требует не разрознивать вещи из замкнутого комплекса ни в основной документации, ни в хранении, ни в публикации.
6. Полнота исследования – этот подход диктует требование тотального вскрытия памятника и сплошных раскопок местонахождений. Согласно этому подходу на стоянке, городище стало желательным раскопать всю их площадь. Из могильника взять не выборку погребений, а все – иначе не ясна структура могильника, не видно распределение разных видов погребений по его площади [Клейн, 2001: с. 76].
7. Принципы бережности и сохранения информации – на нем основана аккумуляция научных знаний и возможность обмена ими. При таком подходе главным требованием является фиксация информации записями, рисунками, чертежами и т.д. Из этого вытекает протокольный подход, который был впервые сформулирован Тэйлором в виде "хорошей" аксиомы для археологов: «важно не то, *что* ты нашел, а *как* ты нашел это», и излишне напоминать, что «*как* ты нашел это, можно установить только из твоих записей, но не из самих находок» [Клейн, 2001: с. 82].
8. Принцип проверяемости. Это общеметодический принцип всякого исследования. Археолог не видел ни событий, ни общей картины исследуемого периода, а видел лишь источник, т.е. их отражение в вещи, отделенных от них бездной времени. И видел мельком, только один раз. Его наблюдение – раскопки неповторимо, т.к. памятник в результате раскопок исчез.

Поэтому археологический отчет – это отражение отражения [Клейн, 2001].

9. Методичный подход предполагает изучение всех возможных разновидностей археологических артефактов и памятников без каких-либо произвольных пропусков.

Хотя традиционные формы и методы археологических исследований практически реализуют использование этих принципов и подходов, однако важное значение должно придаваться форме, в которой фиксируется археологическое знание. Здесь в первую очередь ставится задача использовать современные информационные технологии.

В любой научной организации рано или поздно появляется проблема "наведения порядка" в массиве документов, потому что в настоящее время ни одна из стадий кругооборота документов не обходится без бумажной волокиты. Для обеспечения целостности документов и упорядоченного ведения документации необходима система электронного документооборота, которая обеспечивает быстрый поиск необходимой информации и снижает вероятность случайного уничтожения документов, а также позволяет снизить расходы на канцелярские принадлежности и оргтехнику.

Ключевой характеристикой археологических учреждений, обосновывающей их принципиальное отличие от каких-либо промышленных или финансовых корпораций, является то, что обрабатываемая этими учреждениями информация формируется в результате тесного переплетения двух разнородных информационных потоков, которые условно можно охарактеризовать как "управленческий" и "научный".

2.2. Система электронного управленческого документооборота (СЭУД)

СЭУД представляет собой совокупность технических и аппаратно-программных средств и технологий для создания и управления хранилищами документов.

Управленческие данные характеризуются высокой степенью структуризации. Они являются центральным звеном при организации административно-управленческой деятельности научных организаций. Эта система должна обеспечить среду, являющуюся основой для систем коллективной административной работы, учитывать специфику управляющих процессов организаций и интегрировать потоки информации в единую систему обработки описательной и числовой управленческой информации. Платформа системы электронного документооборота должна включать в себя необходимые инструментальные средства для обеспечения устойчивого функционирования СЭУД.

Автоматизация информационных потоков управленческой информации предполагает:

- а) внедрение доменной архитектуры локальной вычислительной сети археологического учреждения (интранет), средств централизованного администрирования и межсетевого взаимодействия;
- б) оптимизацию состава и расположения серверов с целью уменьшения нагрузки на оборудование и каналы связи;
- в) систему комплексных закупок необходимых технических и программных средств, организации серверного узла;
- г) решения по защите локальной вычислительной сети и данных от несанкционированного доступа из Интернет, определение политики информационной безопасности и их реализация;
- д) выбор системы хранения данных. В зависимости от объема информации и количества пользователей необходимо соответственно наличие на сервере одной из существующих СУБД: MySQL, MS ACSESS, MS SQL Server, ORACLE, IBM DB2 Database или Lotus Notes/Domino и др. Также возможно применение готовых специализированных программных комплексов типа 1С-управление. Выбор должен быть основан на представленном разработчиком технико-экономическом обосновании;
- е) внедрение систем коллективной работы;
- ж) интеграцию планово-финансовых, бухгалтерских, статистических и других отчетных данных;
- з) к обязательным возможностям стоит относить: автоматическую регистрацию документов, ведение их регистрационных карточек, ведение номенклатуры дел, связывание документов, прикрепление файлов, работу со справочниками и словарями, разграничение прав доступа, наличие поиска по реквизитам и полному тексту;

- и) корпоративную электронную почту, и службу новостей;
- к) службы электронных конференций и интерактивного обмена должны обеспечить потребности общения сотрудников археологических учреждений и филиалов в режиме реального времени;
- л) предоставление средств электронной публикации информации о деятельности организации во внешний мир с использованием ресурсов и служб Web-сайта организации;
- м) обеспечение контролируемого и стандартизированного доступа к результатам административной и экспертно-аналитической деятельности институтов и их филиалов извне.
- н) IP-телефонию.

2.3. Система электронного научного документооборота (СЭНД)

В археологических организациях существенную часть корпоративных данных составляет научная информация в бумажной и/или электронной формах, поступающая наряду с "чисто" управленческими данными. Эта информация образует научный поток, в силу задач организации тесно интегрированный с управленческими данными. Научные данные являются, как правило, слабоструктурированными или неструктурированными.

Важнейшей процедурой в такого рода технологиях являются модели данных, регулирующих не только форму представления фиксируемых (вводимых) данных (тексты, рисунки, снимки, чертежи, таблицы и т.д.), но и те материальные (бумага, киноплёнка, аудио и видеокассеты, компьютерные средства), а также логические (макеты данных) носители, на которых эти данные предусматривается размещать для их последующего использования.

Для этой цели служат разнообразные системы управления базами данных и знаний (СУБД), в частности, ориентированных на гипертекстовую и мультимедийную форму представления информации.

Но для археологических учреждений существует ряд специфических проблем. Дело в том, что из представленных СУБД самой мощной и эффективной является ORACLE, которая требует наличия высокооплачиваемых опытных системных администраторов. Поиск вакансий в Internet показывает, что опытные программисты готовы работать как минимум за 45-55 тысяч рублей в месяц. Но всякая ли бюджетная организация готова платить такие суммы?

Что касается более дешевого программного обеспечения для создания такого продукта, сейчас существует много бесплатно распространяемых средств разработки. Примером могут служить программные сборки Denwer и TopServer, в которые включены виртуальный сервер (Apache), MySQL (в TopServer еще и phpMyAdmin), поддержка языков PHP, JAVA, HTML, CSS, PERL, MYSQL и многое другое.

Это потребует меньших затрат на содержание системы, однако не освобождает организацию от приема на работу приходящих или штатных программистов¹.

Для небогатых организаций археологического профиля наиболее доступным является внедрение СЭНД собственными силами, включая и сопровождение её с первых дней применения. В чем здесь преимущества?

1. в СЭНД будут только те функции, которые действительно необходимы. Система не будет требовать больших ресурсов, чем это необходимо. Каждая археологическая организация сможет сама для себя определить необходимый набор функций.
2. стоимость созданной своими силами СЭНД будет гораздо ниже готовой. Стоимость созданного продукта будет включать стоимость скачивания свободно

¹ В САТИ ИАЭТ СО РАН существовала наиболее дешевая в РФ система создания и обслуживания программных продуктов, основанная на использовании труда студентов ВКИ НГУ и ФИТ НГУ с приемом их на работу (с зарплатой от 1000 до 4000 руб.). При такой низкой зарплате естественной была высокая текучесть кадров, однако она компенсировалась непрерывностью процесса подготовки новых студентов, обеспечивающих преемственность в процессе разработки. Так, на вопрос одного из американских коллег на международной конференции: во сколько обошлось создание виртуального музея? – последовал ответ: 8080! Чего? Долларов? Евро? Что Вы – рублей, ответил автор доклада, чем поверг аудиторию в удивленно – шоковое состояние.

распространяемых программ, зарплату разработчика и системного администратора с хорошими навыками работы на компьютере.

3. программный продукт будет настраиваем настолько, насколько это требуется. В тезаурусах будут лишь изначально используемые слова. Внешний вид форм и карточек будет изначально создан с учетом пожеланий будущих пользователей СЭНД.
4. наличие тех или иных дополнительных возможностей будет зависеть только от времени, затраченного на создание или изменение СЭНД.
5. безопасность СЭНД будет всегда под контролем. Кроме того, к администратору будут при возникновении малейших неполадок или проблем обращаться сразу. Заодно станет ясно, надо ли изменять программный код, переобучать пользователей или закрывать слабое место в системе.
6. во время внедрения не потребуются дополнительных специалистов и капиталовложений.

К этим шести большим достоинствам собственной разработки СЭНД можно сразу добавить недостаток: качество такого продукта будет зависеть исключительно от качества знаний, опыта и добросовестности принятого на работу специалиста, а также от правильного руководства процессом разработки. Но та экономия, которая получится в результате применения собственной СЭНД, может перекрыть все издержки на создание.

2.3.1. Проблемы внедрения СЭНД.

Сейчас настало время решить вопрос, могут ли создаваемые базы данных стать эффективным средством доступа к археологическим материалам. Ответ, по мнению Ж.-К. Гардена, однозначен: развитие "сетей данных" в разных науках рано или поздно затронет и археологию. Такие создаваемые общими усилиями "сети данных" должны включать в себя не только исходные материалы раскопок, но и результаты исследовательских работ [Гарден, 1983: с. 225]. Всё это позволит осуществлять информационный поиск не традиционным путем изучения отдельных трудов и периодики, хранящихся в библиотеках², а прямым запросом к локальным или сетевым базам данных. Главное новшество, по мнению Ж.-К. Гардена, здесь не в применении электронной техники, а в возможности прямого доступа к полному тексту первоисточников, собранных либо в одном фонде, либо в распределенном виде, но не рассеянном по разным печатным изданиям, зачастую в урезанном тезисном виде. На это понадобится несколько десятилетий, чтобы отдельные исследователи и исследовательские учреждения приспособились к новой практике, которую большинство археологов пока еще не поддерживают [Гарден, 1983: с. 225-226]. При переходе к новым способам хранения археологической информации в сетях данных должны быть преодолены и другие отрицательные факторы: логические ("зоопарк" форматов описания), экономические (затраты времени и другие препятствия доступа к информации) и психологические ("жадность" в обобществлении результатов собственных исследований, нехарактерная для настоящих ученых) [Гарден, 1983: с. 225].

В САТИ ИАЭТ СО РАН велись работы по унификации терминологии путём создания ряда информационных ресурсов: археологическая Web-энциклопедия; Web-энциклопедия "Археология и этнография Приобья"; материалы к словарю для доисториков (М. Брезийон); плейстоценовые гоминиды; древности Алтая; каменные изваяния Алтая; искусство чжурчжэней; бронза Сибири; серебро Сибири; энциклопедия мифов обских угров; ритуальные колчаны обских угров; атрибутика медвежьего праздника у обских угров; культурное наследие народа Саха; мифология Коми; Международный справочник исследователей Арктики по социальным наукам; археологи и этнографы России; мировая

² В настоящее время в России созданы распределенные корпоративные библиотечные системы, которые могут вести поиск по запросам заинтересованных пользователей.

антропология, этнография и археология в лицах и др. Все перечисленные ресурсы были размещены на сайте "Sibirica" (рис. 32) и отражены в содержании выпусков ИТГИ.

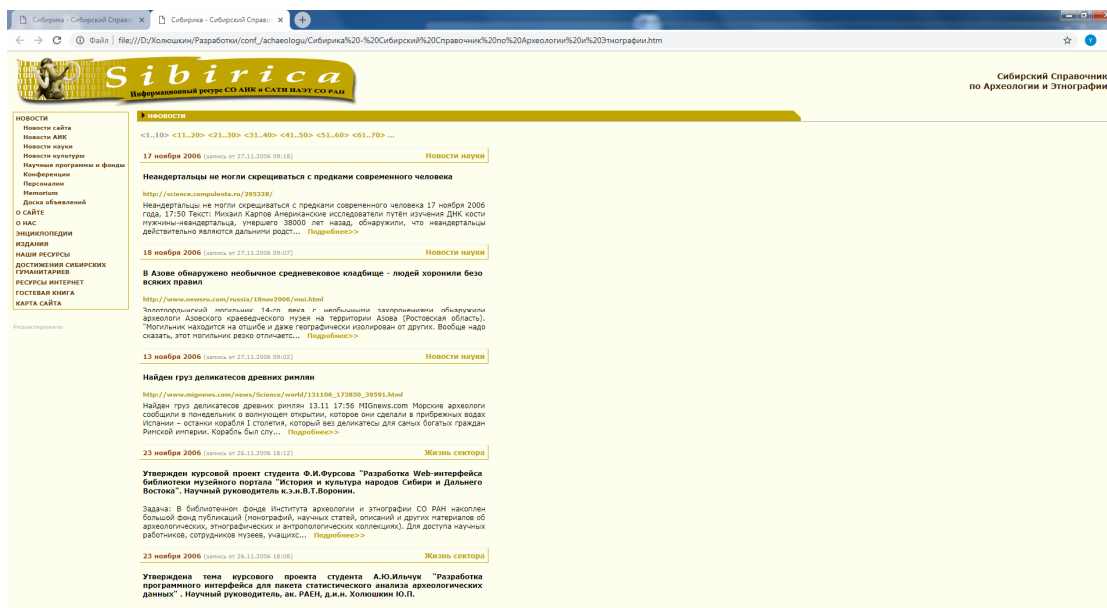


Рис.32. Сайт "Sibirica"

2.3.2. Создана База данных по духовной культуре угорских народов Западной Сибири (рис. 33). Целью данного проекта являлась разработка базы данных по духовной культуре обских угров, позволяющей структурировано хранить археолого-этнографические данные, манипулировать и осуществлять быстрый доступ к ним, с удобным и интуитивно понятным пользователю интерфейсом.

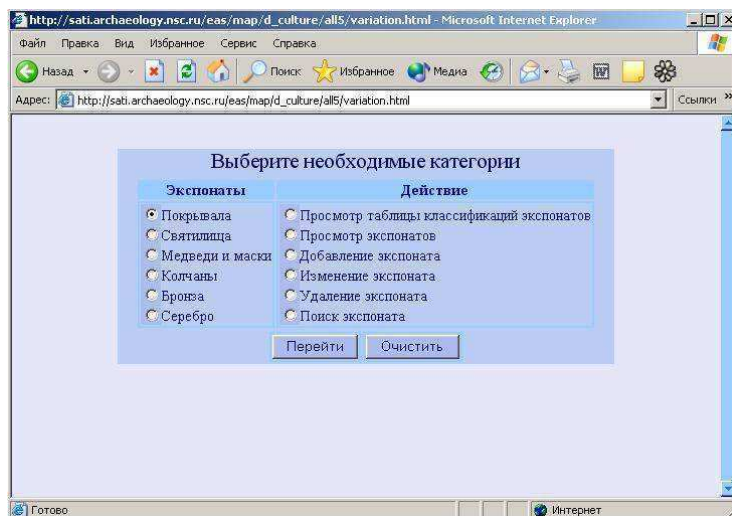


Рис. 33. Главная страница информационной системы по духовной культуре обских угров. База данных содержит информацию о жертвенных покрывалах, культовых местах, атрибутах медвежьего праздника, колчанах, а также предметах из бронзы и серебра.

При проектировании разработаны алгоритмы манипуляции данными, а также система разграничения доступа для различных групп пользователей. Спроектирован интерфейс, эффективный и интуитивно понятный пользователю с любой квалификацией. Программно реализованы:

- база данных с этнографическими данными;
- система поиска, запросов на модификацию записей об этнографических экспонатах;
- база данных пользователей системы;
- контроль уровня доступа пользователей и запросы на модификацию записей о пользователях;
- удобный и наглядный интерфейс.

На основе применения XML (расширяемый язык разметки) построена гибкая и легко переносимая структура данных. Эту структуру с легкостью можно использовать в различных приложениях (от офисных программ до крупных систем управления базами данных), а также объединять с другими этнографическими системами и базами данных.

Для базы данных создан комплекс программ по манипулированию, обработке и визуализации этнографических данных и данных о пользователях на языке РНР.

По результатам работ опубликована коллективная монография "Разработка новых методов информационных технологий представления и обработки археологических и этнографических данных. Материалы научного отчета по интеграционной программе СО РАН за 2003-2004 гг. (проект № 149) ", 8.5 п.л.

2.4. Полевые отчеты в ИАЭТ СО РАН

Создание отчета (перевод полученных при раскопках материалов в электронную форму, максимально отражающую по полноте всю собранную документацию) – дело очень трудоемкое и занимает достаточно продолжительный период времени. Кроме того, часто не все материалы полевых исследований отражаются в отчетах. Прежде всего, это касается различных наблюдений или идей автора (ов) раскопок, которые отражены только в полевых дневниках. Еще 10 лет назад на интернет-форумах по археологической тематике активно обсуждался вопрос – нужен ли компьютер в поле, а уже при современном развитии информационных технологий и методики обработки материала выезд на полевые исследования без компьютеров кажется немыслимым.

Однако эффективная автоматизация становится возможной только после разработки информационной модели данных для упорядоченного хранения всех описаний и иллюстративных материалов, необходимых для генерации отчета. Пока не зафиксирован определенный формат хранения данных, задачи автоматизации рассыпаются, как дом, построенный на песке.

Информационная модель отчета. В тексте отчета должно быть представлено достаточно информации для того, чтобы специалист имел возможность составить объективную картину проведенных полевых работ и мог проверить справедливость выводов авторов раскопок. Для этого в нем должно найти отражение не только описание проведенных работ и извлеченного археологического материала, но и авторское видение обстоятельств исследования, составляющих ближайшее окружение эмпирического объекта, а также имевшихся в распоряжении исследователей материальных, финансовых и человеческих ресурсов [Костин, Постнов и др. 2010: с. 60-65].

а) Разработка системы генерации полевых отчетов в ИАЭТ СО РАН

Первая попытка создания электронных полевых отчетов предпринималась И. Черниковым под руководством А.В. Постнова. В 2010 году в САТИ стала разрабатываться система генерации полевых отчетов, в рамках дипломных работ студентов ВКИ НГУ (Рис.34).

В процессе выполнения проекта была предложена настраиваемая (гибкая) структура хранения описательной информации, включающей тексты, фотографии, карты и планы, локальные и GPS координаты элементов раскопок и находок. Такая архитектура программной системы позволяет отделить задачу создания содержательной структуры представления данных от задач разработки интерфейса и базы данных. В результате выделилась роль так называемого "архитектора" системы, который фактически самостоятельно разрабатывает систему, не обращая при этом к помощи программистов.

В результате разработки Web-приложения были решены следующие задачи:

- Разработка базы данных;
- Разработка пользовательского интерфейса;
- Разработка интерфейса для ввода географических координат;
- Разработка модуля генерации отчета.

Первая версия системы поддерживает генерацию полевых отчетов в виде HTML-документа.

Генерация отчёта предусмотрена в двух видах:

- По шаблону, разработанному "архитектором";
- Полная (без шаблона) – для обзора всех описаний, относящихся к отчету.

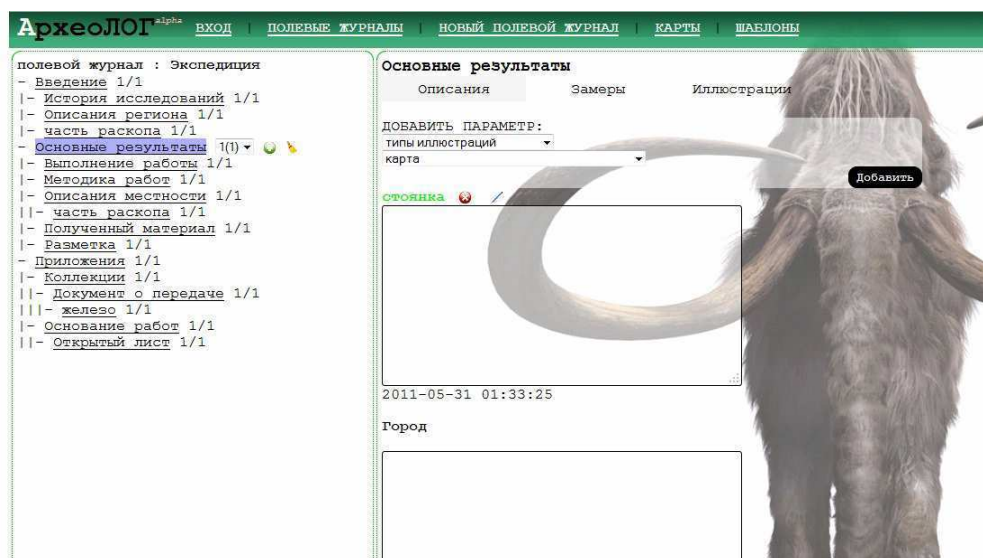


Рис.34. Страница формы полевого отчета (по Костин, Постнов, ... 2010, с.60-65)

В дальнейшем предполагается дополнить систему возможностями вывода в форматах документов MS Word, PDF и LaTeX. Однако, отсутствие финансирования и ясного представления о структуре отчёта не позволили достичь положительного результата.

2.5. Научный отчет

Одним из направлений является разработка ресурсов, освобождающих научных сотрудников от рутинной работы. Создаваемая система в первую очередь была предназначена для ученых секретарей научных подразделений ИАЭТ СО РАН. Известно, что ежегодные научные отчеты отнимают много времени. Большая часть времени уходит на рутинную работу по сбору и перепроверке статистической информации (например, список сотрудников, их публикаций, экспедиционных поездок, участия в конференциях и т.д.).

Информация должна храниться из года в год и постоянно пополняться. При этом желательно организовать так, чтобы каждый сотрудник имел возможность добавить или изменить ту часть, которая касается лично его. Конечно, эта работа требует некоторой систематизации.

Можно попытаться хранить всю информацию в текстовых файлах (например, в формате Word или Excel, или же в каком-либо ином). Однако этот способ хранения данных может порождать характерные для подобных случаев проблемы. Например:

- проблему синхронизации при попытках внести изменения – один редактирует, остальные вынуждены ждать;
- необходимость соответствующего программного обеспечения для чтения и редактирования общих файлов;
- вольность использования форматирования разными людьми в силу разных представлений о конечном документе.

Стандартным решением таких проблем является отделение информации от формата ее представления, для чего идеально подходит инструментальный, включающий современные базы данных и способы доступа к ним по сети интернет. Современные инструменты манипулирования данными в сетевых базах данных позволяют их пользователям:

- выполнять одновременное и независимое редактирование отдельных записей;
- формировать итоговый отчет в одном из принятых форматов.

Кроме того, для доступа к информации, хранимой в базах данных, целесообразно использовать веб-интерфейс, который обладает рядом бесспорных преимуществ:

- обеспечивает доступность с любого компьютера, подключенного в сеть;

- не требует установки дополнительного программного обеспечения.

Таким образом, наряду с решением перечисленных выше проблем, ориентация на использование баз данных в подготовке и ведении научных отчетов дает возможность научным сотрудникам, их руководителям и ученым секретарям получать доступ к их ресурсам практически из любого места.

Исходя из сложившихся традиций оформления и порядка ведения отчетов, в секторе археологической теории и информатики были вычленены основные составляющие статистической информации для отчета, которые мы будем в дальнейшем называть объектами:

- гранты;
- конференции (рис. 35);
- экспедиции;
- публикации сотрудников сектора;
- международное сотрудничество;
- сотрудники сектора и др.

В силу некоторых ограничений реляционной базы данных в ней нельзя хранить объект произвольной структуры. Это означает, что каждое поле хранимого объекта должно быть простого типа (например, текст или число). Поэтому для хранения подобных объектов могут потребоваться дополнительные структуры, которые мы назовем подобъектами.

Рис. 35. Пример заполнения формы по отчету об участии в научной конференции.

В подобъектах предполагается хранить те части объекта, которые не укладываются в простой тип. Например, список участников, соавторов, список работ и т.п. Подобъекты по своей структуре аналогичны объектам. Причем, если их структура не укладывается в простые типы, то они могут также содержать подобъекты более низкого ранга. Впрочем, в данном случае все подобъекты оказались простого типа.

На основе выявленной структуры и ориентировочного содержания отчетов для каждого объекта и подобъекта были созданы таблицы в базе данных и сформированы скрипты для отображения списка уже существующих в базе записей, их редактирования и создания новых.

Полученная система позволяет вводить, редактировать и просматривать всю статистическую информацию. После добавления скрипта, формирующего отчет на основе введенной информации, получаем готовую систему для хранения, редактирования и формирования отчета. [Холюшкин, Воронин, Воробьев, 2004: с. 72-74; <http://www.sati.archaeology.nsc.ru/index.html?mi=otchet-y-za-period>] (Рис.35).

Воспользовались системой немногие, а после ликвидации сектора система и вовсе прекратила свое существование.

2.6. Электронная библиотека.

Сектором было в свое время осуществлено освоение и эксплуатация прикладных программ для наполнения пилотного варианта электронного каталога в базе данных научной библиотеки ИАЭТ СО РАН (археология и этнография) [Холюшкин и др., 2002: с. 15-20; 2003: с. 81-85;].

Начало работ по разработке проекта совпало с разделом библиотеки между ИАЭТ СО РАН и остальными Институтами бывшего ИИФФ СО РАН, что практически вывело из числа участников сотрудников библиотеки. Для участия в разработке, наполнении и размещения в Интернет электронных каталогов научных изданий нам было предложено директором ГПНТБ СО РАН Б.С. Елеповым, разработать "пилотный" вариант каталога электронной библиотеки ИАЭТ для включения в состав сводного каталога корпоративной распределённой библиотечной системы Новосибирского региона. Для этой цели САТИ поучило финансовую поддержку ИОО Фонд содействия (Фонд Сороса).

На момент создания библиотечного сайта общий объем фонда библиотеки ИАЭТ СО РАН составлял около 40 тыс. единиц хранения. В него входили книги, брошюры, продолжающиеся и периодические издания по тематике исследований Института и смежным научным дисциплинам, а также электронные ресурсы на сменных носителях и полнотекстовые документы из открытых источников, размещенные во внешней сети.

При создании пилотного варианта в качестве программного обеспечения СУБД базы данных библиографических описаний было использовано свободно распространяемое программное обеспечение СУБД WinIsis. Для его освоения и последующего использования потребовалась локализация, так как кириллизированных версий этого программного продукта не существовало. На основе разработанной участниками проекта кириллизованной версии программного обеспечения разработан единый формат данных электронного каталога и отработаны конвертеры (Рис. 36, 37, 38).

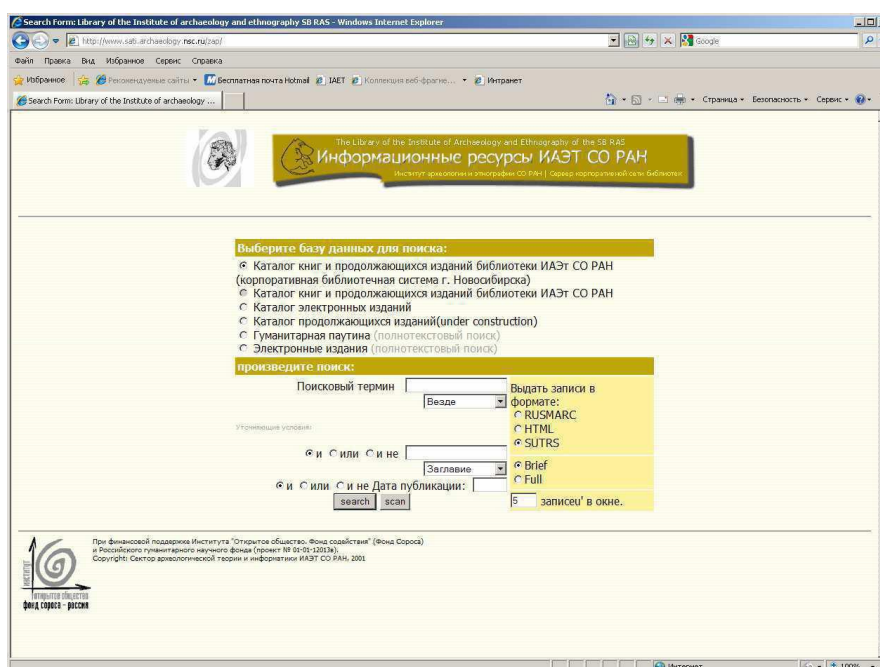


Рис.36. Главная страница электронной библиотеки

С описанными выше технологиями создания и накопления археологических данных тесно связана технология предоставления (доступа к) информации археологических исследований, с которыми она согласуется по моделям и формам, в которых она фиксируется в археологических базах данных. В технологии доступа включаются в первую очередь основные процедуры и операции по обеспечению удобного пользования собранными данными, обеспечивающие действенную защиту данных от несанкционированных действий пользователей, имеющих доступ к хранимой информации, или от сбоев оборудования.

Очевидно также, что доступ к данным более свободен тогда, когда информационный поиск может происходить по любому термину, или комбинации с другими терминами [Гарден, 1983: с. 224]. Развитие сетей данных в разных науках рано или поздно затронет и археологию. Суть сетей данных давно известна: это связанные между собой исследовательские центры, которые общими усилиями создают базы данных по данной отрасли знаний. Они могут включать в себя не только исходные материалы, но и результаты исследовательских работ. Все это позволяет осуществлять информационный поиск прямым запросом в фонды базы данных, либо непосредственно, либо по каналам связи [Гарден, 1983: с. 225]. Главное новшество здесь состоит в возможности доступа к первоисточникам, собранным в одном фонде. Выводы Ж.-К. Гардена основаны на прошлом опыте археологии и не могли в то время учесть новшества и концепции новой научной методологии в применении распределенных систем, разработанных в конце XX – начале XXI вв.

Такие новшества были внедрены в первую очередь в библиотечном деле и музееведении, где были разработаны стандартные обменные форматы языков представления RUSMARC, USMARC (в библиотеках), CIMI (в музеях) и стандартные протоколы взаимодействия (Z39.50, HTTP). Эти технологии нашли применение в областных библиотеках и библиотеках вузов (например, НГУ, БГПУ, Красноярский ГПУ). Нам известен лишь единичный пример попытки внедрения подобных технологий в археолого-этнографических НИИ. Такой проект реализовывался в ИАЭТ СО РАН (Рис.37). В соответствии с целями и задачами проекта создания библиотечной системы по гуманитарным наукам в Сибири ресурсы электронного каталога через Web-шлюз и Z39.50-шлюз включены в сводный каталог региональной распределенной корпоративной библиотечной системы Новосибирской области (URL: <http://z3950.uiggm.nsc.ru/zgwc/>, см. рис.38).

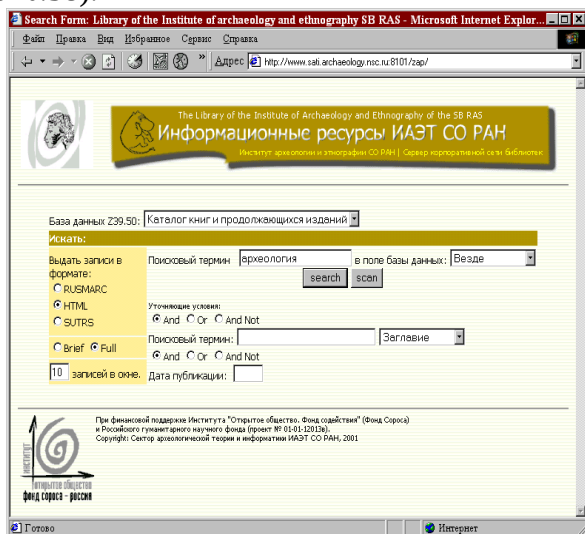


Рис.37. Web-страница WWW–Z39.50 шлюза к базам данных Z-сервера сектора археологической теории и информатики ИАЭТ СО РАН.

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. Доработан конвертер электронного каталога из формата базы данных под управлением СУБД WinIis 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.

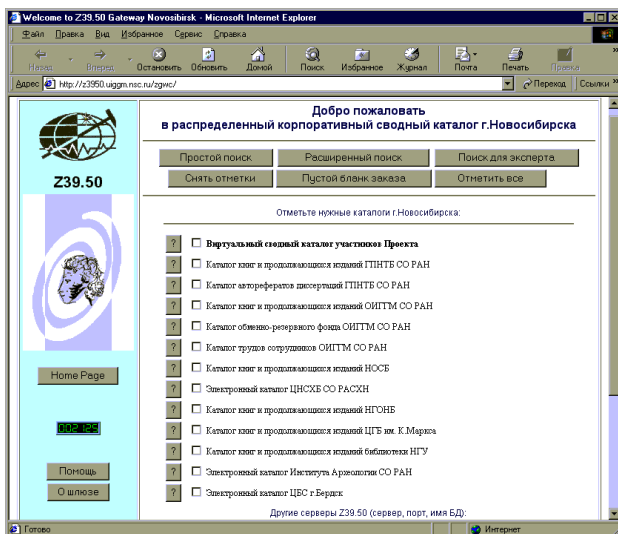


Рис.38. Web-страница сводного каталога Новосибирской региональной распределенной корпоративной библиотечной системы.

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. По окончании работ Г.Ю. Жилицкой произведено заполнение БД 8000 единицами каталога. Осуществлено подключение электронного каталога научной библиотеки ИАЭТ СО РАН к корпоративной сети г. Новосибирска.

Однако куратор библиотеки М.В. Шуньков заявил, что ИАЭТ не нужны самоделки и ИАЭТ будет приобретать у ГПНТБ ИРБИС, к сожалению, вплоть до ликвидации САТИ покупка не состоялась, не воспользовался ИАЭТ СО РАН и содержанием БД (8000 ед.).

2.7. Сайт для поддержки, организации и проведения научных конференций.

Этот сайт, разработанный в САТИ был ещё одним "мертворожденным ребёнком". Ю.П. Холюшкин неоднократно обращался к учёному секретарю ИАЭТ Сагайдачному, а затем к Коровушкину о возможности подключения ИАЭТ СО РАН к системе конференций ИВТ СО РАН и к системе видео конференций, однако положительного отклика не получил. На свой страх и риск Холюшкин принял решение о разработке сайта конференций к I археологическому съезду. Задача состояла в разработке Web-интерфейса для просмотра и изменения информации о проводимых институтом конференциях.

Сама проблема заключается в быстром, удобном и надёжном хранении и использовании информации о данных мероприятиях.

Организации решают эту проблему путем разработки web-приложений (сайтов), в которых они могут держать всю эту информацию, а также часть ее будет доступна любому заинтересованному в этом пользователю сети Интернет.

Данная проблема встала и перед институтом Археологии и Этнографии СО РАН, постоянно проводящим такие мероприятия. Появилась необходимость в хранении архива прошедших, анонса предстоящих и расписания текущих научных конференций. Также потребность в записи участников на определенную конференцию с предоставлением ими своих работ (докладов) и многие другие проблемы.

С целью решить данную проблему, выпускниками ВКИ НГУ А.Е. Мархелем и В.В. Зинченко был разработан пользовательский интерфейс к базе данных, в которой предполагалось размещать всю вышеперечисленную информацию. Интерфейс представлял собой web-сайт. При его помощи потенциальные участники конференции могли, внося свои личные данные, послать заявку на регистрацию администратору сайта, который в свою очередь, просмотрев доклад данного пользователя, будет регистрировать его на определенную конференцию, отправив ответ. Для администрирования также был разработан удобный интерфейс, благодаря которому пользователь обладающий этими правами сможет также вносить информацию о новых предстоящих конференциях и о последних новостях археологии.

Для разработки интерфейса был выбран продукт Denwer, включающий в себя сервер Apache 1.3.33, программу PHP версии 4.4.2, СУБД MySQL 4 и многие другие компоненты. В качестве языка программирования был выбран язык PHP, а соответственно для этого необходим интерпретатор этого языка, в нашем случае это PHP 4.4.2.

Разработанное web-приложение относительно простое в использовании, как и большинство других сайтов.

На главной странице (Рис.39), был размещён информационный раздел, в котором выводятся последние новости. За новостями следует информация о конференциях:

- *Прошедшие конференции*
- *Текущие конференции*
- *Будущие конференции*

Кликнув на ссылку с названием конференции, вы попадете в раздел просмотра более подробной информации о ней.

В самом низу выводится имя последнего зарегистрировавшегося участника.

С левой стороны сайта, под логотипом, располагалось вертикальное меню, состоящее из следующих пунктов:

- *Новости института*
- *Анонс конференций*
- *Текущие конференции*
- *Архив конференций*
- *Регистрация участников*
- *Рассылки*
- *Ин-т Археологии и Этнографии*

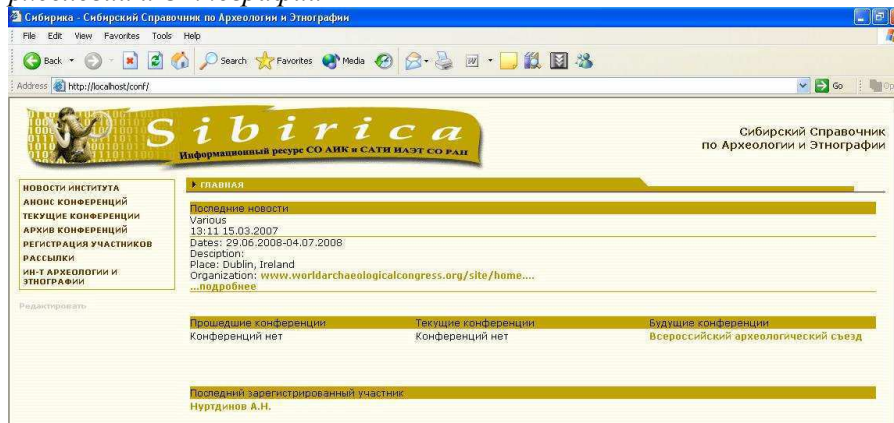


Рис.39. Главная страница конференций

Рассмотрим каждый пункт в отдельности.

Новости института.

Кликнув по данной ссылке, откроется раздел с архивом всех новостей института. Под текстом каждой новости, расположена ссылка «...подробнее», с помощью которой вы можете просмотреть выбранную новость подробнее.

Анонс конференций.

В разделе анонса конференций выводится информация о предстоящих конференциях.

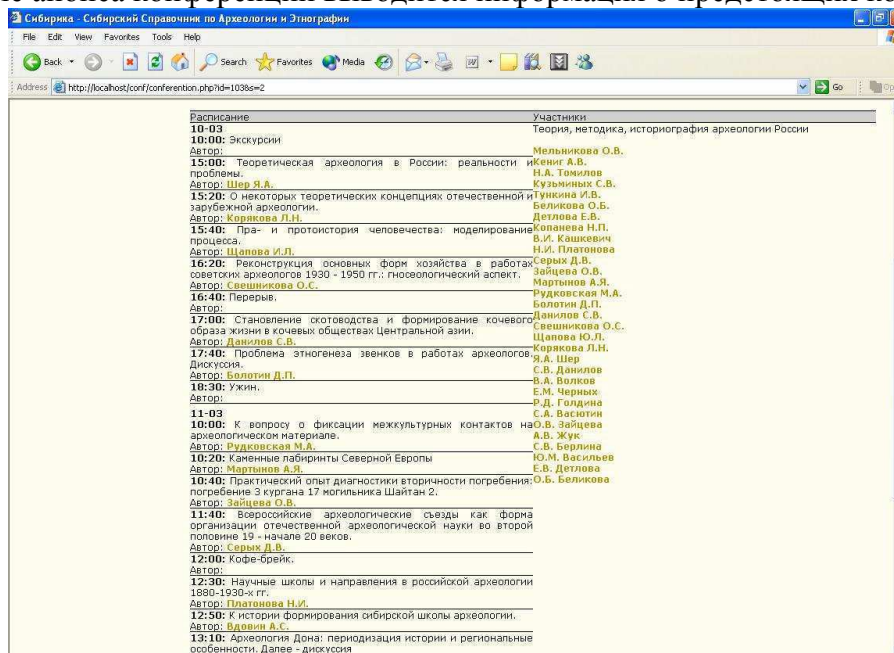


Рис.40. Расписание работы секции "Теория, методика, историография археологии России".

Текущие конференции.

В данном разделе содержится информация о конференциях, проходящих в данный момент.

Архив конференций.

В архиве конференций хранится список уже прошедших конференций.

В каждом из последних трех разделов хранится информация о конференциях. Кликнув по ссылке «...подробнее», расположенной возле каждой конференции, вы сможете увидеть подробную информацию о ней, а именно:

- *Дата проведения*
- *Место проведения*
- *Организатор*
- *Описания*
- *Секции*
- *Участники*
- *Расписание*

Расписание составляется свое для каждой секции отдельно. Название каждой секции является ссылкой, перейдя по которой, слева появится расписание секций и список ее участников справа.

Перед текстом события в расписании, находится ссылка с фамилиями авторов, перейдя по которой, можно просмотреть доклад, с которым она выступают.

Кликнув по конкретному участнику, из списка справа, вы сможете просмотреть более подробную информацию о нем:

- *Дата рождения*
- *Организация*
- *Научное звание*
- *Ученая степень*
- *Город*
- *Конференция*
- *Секция*
- *Доклад*

Регистрация участников.

Если у пользователя возникла необходимость зарегистрироваться на заинтересовавшую его конференцию, то необходимо кликнуть на данную ссылку, после чего перед ним появится форма с полями для ввода.

Поля, которые помечены звездочкой (символом «*») обязательны для заполнения. Т. е. зарегистрироваться будет невозможным, если не заполнены эти поля.

В число таких полей входят:

- *"Имя"* – Необходимо ввести свое имя, это минимум личной информации, необходимой для успешной регистрации.
- *"E-mail"* – Адрес электронной почты необходим только для того, что бы администратор, после рассмотрения вашей заявки на участие смог вас уведомить вас об успешной регистрации, либо наоборот, указать вам причину отказа.
- *"Конференция"* - В данном выпадающем меню должна быть выбрана конкретная конференция, участником которой хотел бы стать пользователь. Ведь он не может послать заявку на участие "ни в какой" конференции.
- *"Доклад"* - Нажав кнопку «Обзор...» необходимо найти файл, формата Microsoft Word (*.doc, *.rtf), содержащий вашу работу (доклад) с которой вы хотели бы выступить на конференции. Эта работа будет просмотрена администраторов сайта.

Также существуют не обязательные для заполнения поля, такие как *"Дата рождения"*, *"Научное звание"*, *"Ученая степень"*, *"Город"*. Несмотря на то, что поля не обязательные, их заполнение желательно.

2.8. Виртуальный музей

Электронные коллекции были сосредоточены в Виртуальном музее истории и культуры народов Сибири и Дальнего Востока, создаваемом как система хранения, доступа и навигации по ресурсам музейных коллекций [Холюшкин и др. 2008: с. 29-36].

Наполнение и доступ к информации электронного каталога музея осуществлялся под управлением MySQL в операционной системе Linux. Обслуживание пользовательских запросов и генерация HTML-страниц производится с помощью программ на языке PHP, JavaScript, XML.

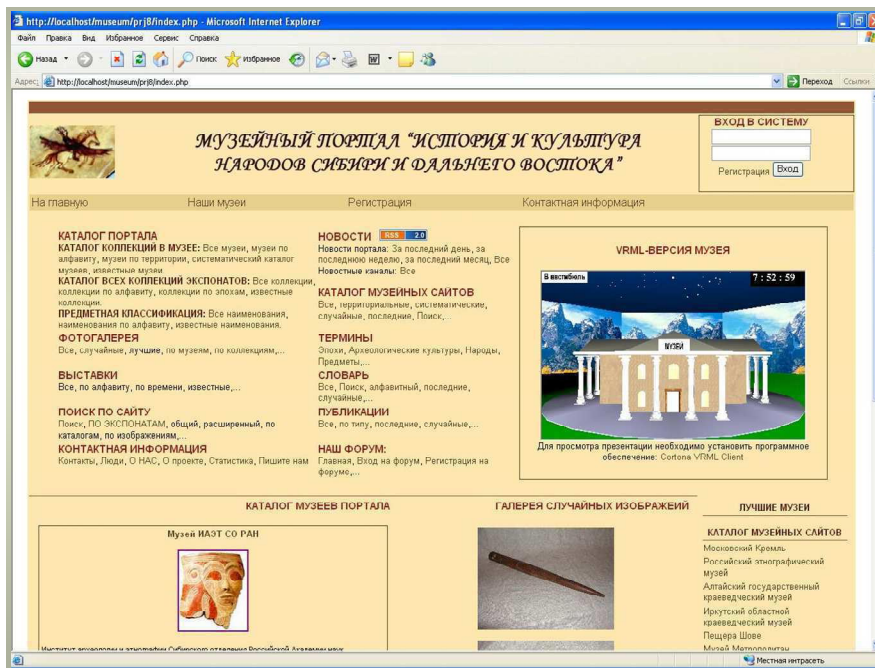


Рис.41. Главная страница музея.

Как информационный ресурс портал знаний по археологии и этнографии обеспечивает следующие возможности (Рис.41):

1. хранение и получение исчерпывающей информации о музейных экспонатах;
2. соответствие стандарту международной организации по стандартизации ISO № (Z39.50) по хранению информации о культурном наследии – № (СИМ);
3. создание и хранение различного рода информации о музеях, виртуальных и реальных;
4. сохранение изображений предметов, оцифрованных фотографий и набросков, так что каждый экспонат может хранить несколько изображений, и представления этих изображений в разных уровнях качества;
5. просмотр их в удобных последовательностях;
6. создание различных классификаций и типологических таблиц в виде директорий и удобную навигацию для них;
7. быстрый поиск по самым "популярным" полям описания;
8. расширенный поиск по любым полям формы описания;
9. создание и просмотр публикаций и поддержку просмотра других публикаций, соответствующих музейной тематике и размещенных за пределами музея, и информацию о них;
10. поддержку просмотра полезных ссылок, которые соответствуют музейной тематике, и информацию о них,
11. поддержку просмотра последних новостей по музееведению и тп.;
12. более полный доступ и расширенную информацию (например, по экспонатам) пользователю посредством назначения ему дополнительных привилегий;
13. полный инструментарий по манипулированию различными категориями данных назначенному пользователю с предоставлением ему дополнительных привилегий;
14. удобный и простой интерфейс;
15. связь с VRML-версией музея и картографической системой.

Одним из приоритетных направлений развития музейного Web-сайта Института археологии и этнографии СО РАН и популяризации его музейных коллекций является активное использование технологий моделирования виртуальной реальности для поддержки полной иллюзии реального свободного доступа в залы и хранилища музея для

их посетителей и самостоятельного изучения находок и их описаний применительно к их вкусам и интересам.

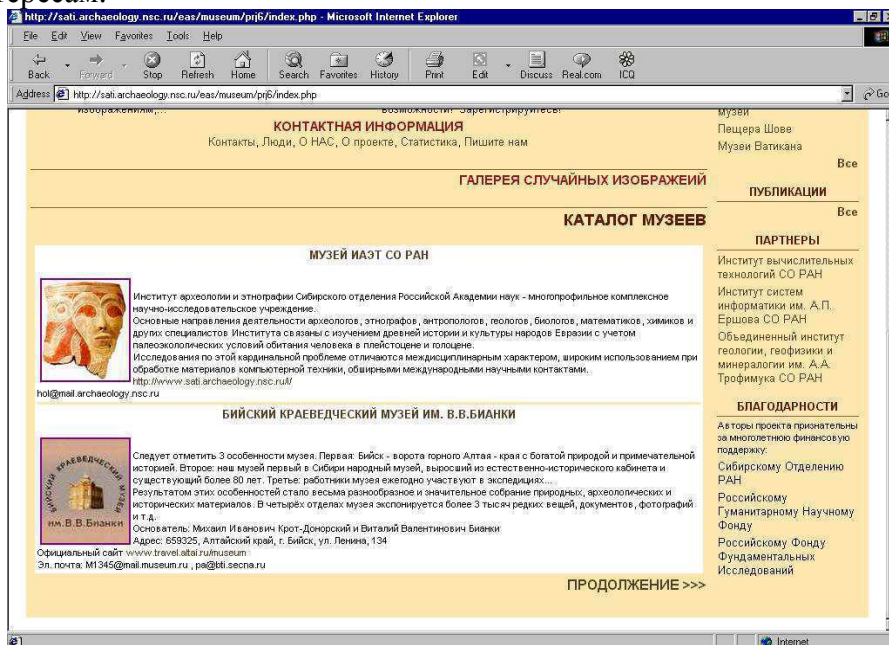


Рис.42. Каталог музеев на главной странице.

Компьютерные трёхмерные реконструкции, созданные по технологии виртуальной реальности и снабженные гипертекстовыми и мультимедийными структурами, могут помочь формированию существенно более целостного, наглядного и детального представления, дадут возможность погрузить музейные объекты в историко-культурный контекст, предоставят возможность самостоятельного обучения и исследования. Публикация таких моделей в сети Internet и/или на компьютере (т.н. музейном киоске) должна дать новую дополнительную возможность огромной аудитории значительно расширить свои представления о культурной и исторической ценности музейной экспозиции, а также о существующих в музее проектах реставрации.

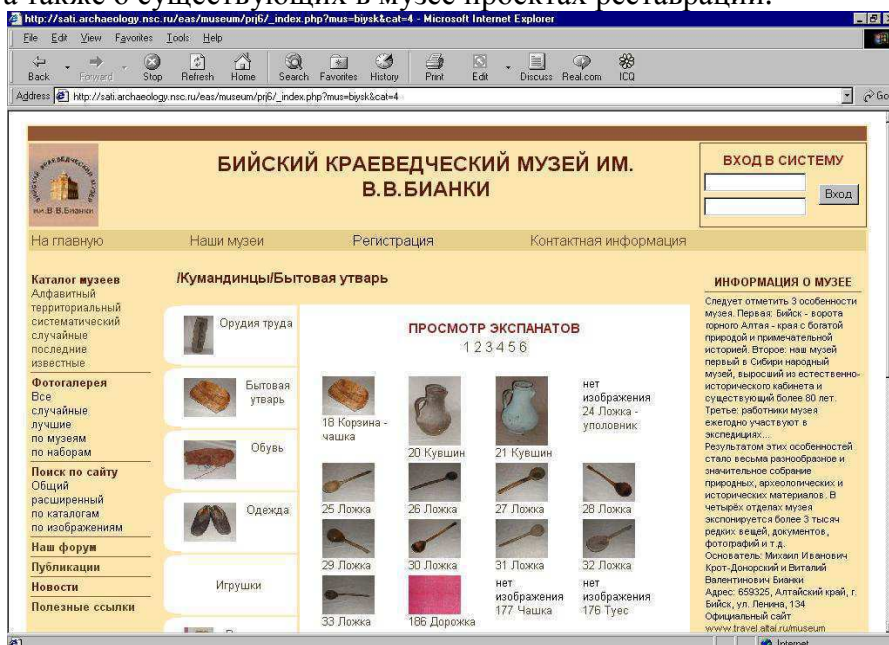


Рис. 43. «Кумандинская коллекция». Музейные предметы «Бытовая утварь».

В ходе реализации проекта осуществлена комплексная программа исследований по созданию и развитию проблемно-ориентированной среды по гуманитарным наукам и разработке на этой основе гуманитарных информационных ресурсов. Исходя из этого принято и реализовано решение о создании информационного ресурса, включающего специализированный узел сервера САТИ с информацией о музейных коллекциях.

В секторе археологической теории и информатики ИАЭТ СО РАН был выполнен комплекс работ по созданию виртуального музея по истории и культуре народов Сибири. Эти электронные материалы составляют основное содержание наполнения виртуального музея ИАЭТ СО РАН и служат предметом каталогизации ее ресурсов в форме электронного каталога.

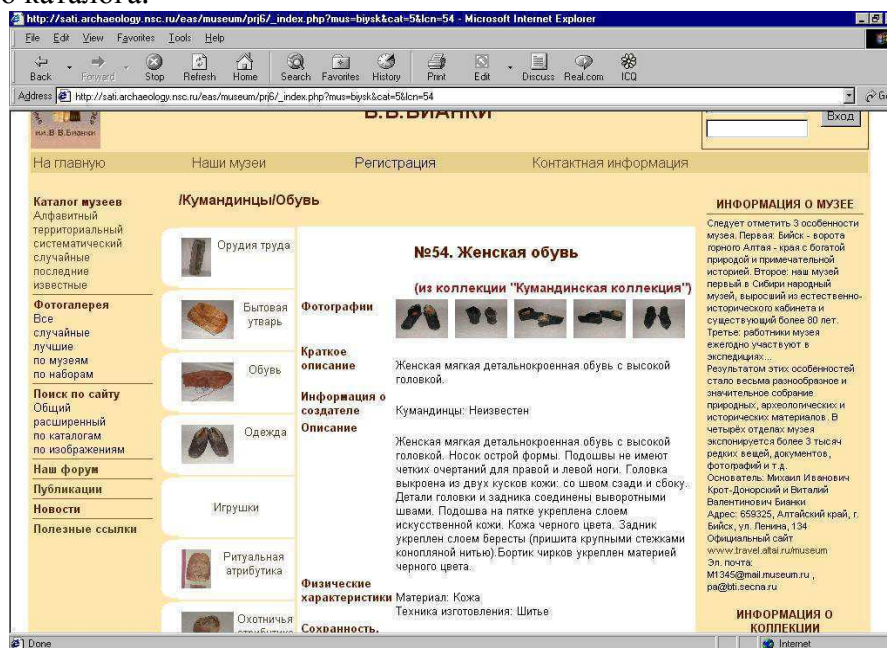


Рис. 44. «Кумандинская коллекция». Экспонаты «Обувь».

Источниками информации являются фонды научной библиотеки ИАЭТ СО РАН, а также три сайта: Виртуальный музей «Древняя история, культура и искусство Северной Азии» (http://sati.archaeology.nsc.ru/gen-i/virtual_e.htm), «Сибирика» (<http://www.sati.archaeology.nsc.ru/sibirica/news/index.html>), «Геоинформационные системы» (<http://sati.archaeology.nsc.ru/eas/map/>).

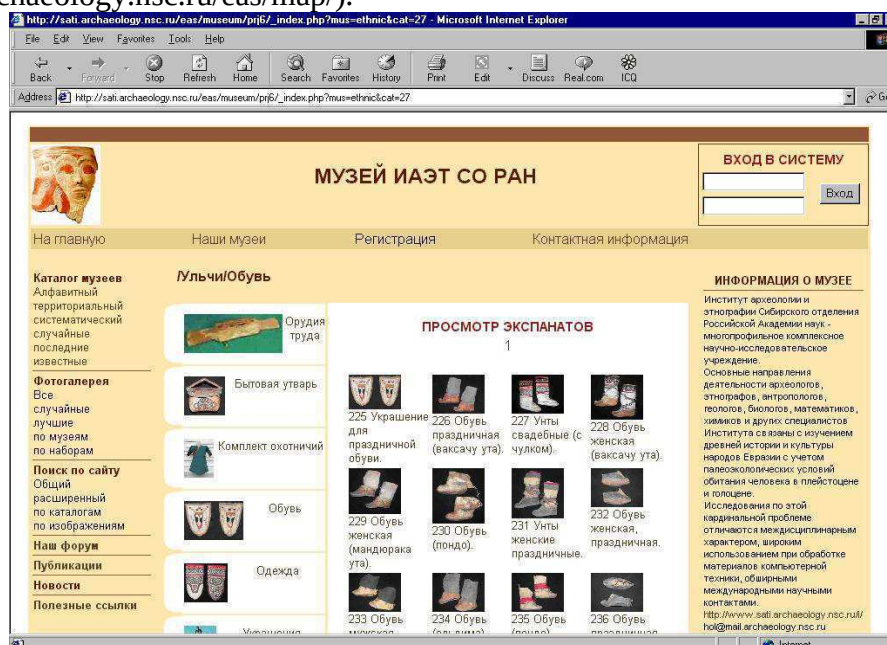


Рис.45. Коллекция «Ульчи». Музейные предметы «Обувь».

Доступ к ресурсам виртуального музея этнографии Сибири и Дальнего Востока имеет два режима:

- свободный по сети Интернет – для пользователей (научных сотрудников, преподавателей, студентов и учащихся),
- авторизованный – для привилегированных пользователей (администраторов).

Создание виртуального музея с обеспечением доступа к его данным в сети Internet предназначено не только для предоставления доступа к имеющимся музейным фондам, но и для накопления, систематизации, сохранения новых поступлений в эти фонды в будущем.

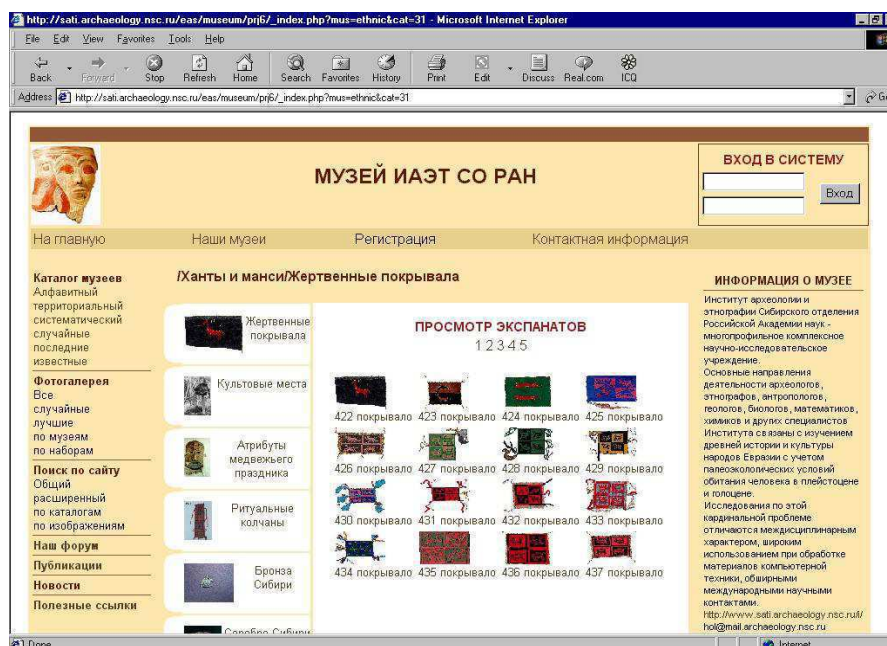


Рис. 46. Коллекция «Ханты и манси». Экспонаты «Жертвенные покрывала».

Виртуальный музей разрабатывался как система хранения и навигации по ресурсам музейных коллекций. Кроме того, в музее имела комплексная справочная система, предоставляющая информацию о текущем состоянии научно-исследовательской и краеведческой, а также экспозиционно-выставочной деятельности.



Рис. 47. Коллекция «Ханты и манси». Экспонаты «Жертвенные покрывала».

Основой информационного наполнения музея является коллекции изображений предметов (оцифрованных фотографий, рисунков, карт, схем и других форм графического представления материалов в электронной форме) и их описаний. Кроме того, наполнение включает другую информацию, например, публикации статей, монографий, новости музееведения.

На начальных этапах на создаваемом музейном портале представлены два музея (Рис.42):

1. Музей ИАЭТ СО РАН,

2. Бийский краеведческий музей им. В.В. Бианки.

Была возможность просмотра коллекций в различных последовательностях.



Рис.48. Коллекция «Ханты и манси». Экспонаты «Атрибуты медвежьего праздника».

Духовная и материальная культура кумандинцев. В виртуальном Бийском краеведческом музее имени В.В. Бианки представлена коллекция "Духовная и материальная культура кумандинцев" (коллекция предоставлена для размещения научным сотрудником Бийского краеведческого музея Соловьевой Светланой Анатольевной). Последняя насчитывает 214 экспонатов (Рис.43-44). Коллекция включает следующие разделы: обувь (рис.51), бытовую утварь, одежду, игрушки, ритуальную и охотничью атрибутику, украшения.

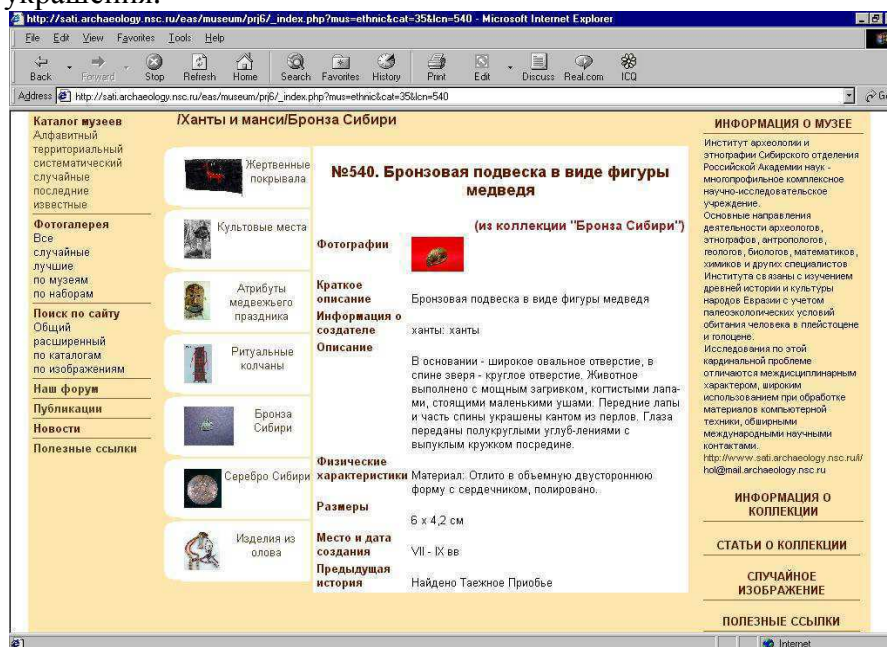


Рис. 49. Коллекция «Ханты и манси». Экспонаты «Бронза Сибири».

В виртуальном музее Института археологии и этнографии СО РАН представлены 2 коллекции:

1. Духовная культура обских угров (ханты и манси) включает в себя следующие коллекции (Рис.46-50):
 - а) Жертвенные покрывала - 76 экспонатов;
 - б) Культовые места – 9 экспонатов;
 - в) Атрибуты медвежьего праздника – 22 экспоната;

- г) Ритуальные колчаны – 6 экспонатов;
 - д) Изделия из бронзы – 33 экспоната (Рис.49);
 - е) Изделия из серебра – 11 экспонатов (Рис.50);
 - ё) Изделия из олова – 36 экспонатов.
2. Духовная и материальная культура ульчей насчитывает 197 экспонатов (Рис.45).

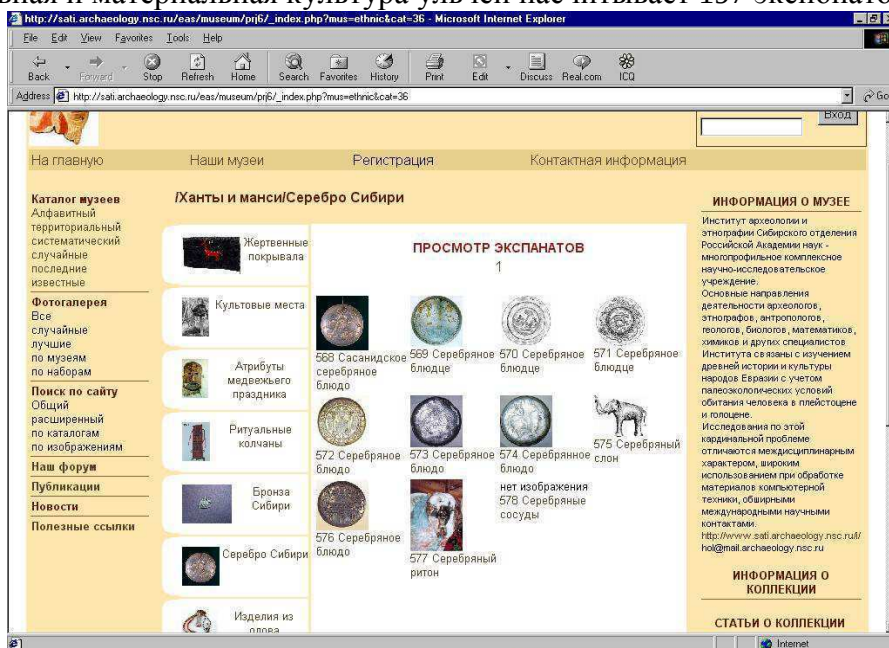


Рис. 50. Коллекция «Ханты и манси». Экспонаты «Серебро Сибири».

Всего в музее представлено 1899 оцифрованных изображений экспонатов (Рис.51).

Созданный Виртуальный музей истории и культуры народов Сибири и Дальнего Востока позволял:

- 1) Работать безопасно и конфиденциально в сети Internet с любого компьютера
- 2) Хранить и получать исчерпывающуюся информацию о музейных экспонатах
- 3) Создавать экспонаты в специальной базе данных

Соответствует стандарту международной организации по стандартизации ISO № (Z39.50) по хранению информации о культурном наследии – (СИМІ).

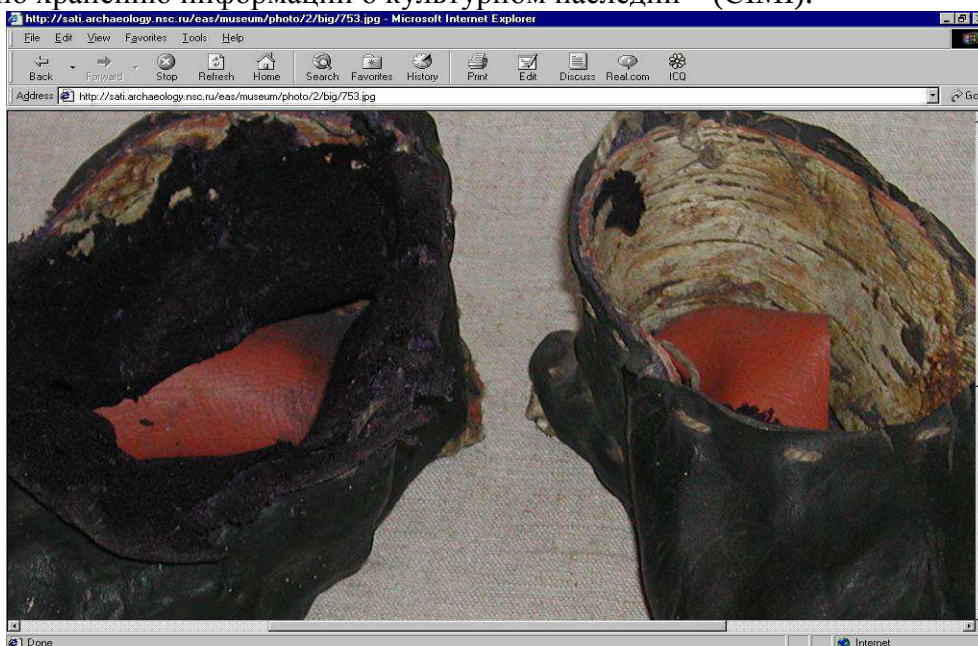


Рис. 51. Одна из фотографий в окне браузера. «Кумандинская коллекция». Экспонаты «Обувь». Музейный предмет №54. «Женская обувь».

Информация, представленная в музее, могла быть полезна для исследователей различных специальностей (историков, археологов, этнографов, краеведов), а также

преподавателей, искусствоведов и т.д. Кроме того, материалы могут привлекаться для образовательных целей в контексте ряда гуманитарных дисциплин, например, в учебном процессе на гуманитарных факультетах Российских университетов и в средней школе при изучении археологии, краеведения, истории, этнографии, искусствоведения. Потенциальные пользователи такой информации школьники, студенты и преподаватели ряда специальностей.

2.9. Виртуальный VRML-музей Института археологии и этнографии СО РАН.

Одним из приоритетных направлений развития музейного Web-сайта Института археологии и этнографии СО РАН и популяризации его музейных коллекций было активное использование технологий моделирования виртуальной реальности для поддержки полной иллюзии реального свободного доступа в залы и хранилища музея для их посетителей и самостоятельного изучения находок и их описаний применительно к их вкусам и интересам.

Виртуальной реальностью мы называем имитацию на персональном компьютере реального мира с помощью трёхмерных и интерактивных моделей.

Компьютерные трёхмерные реконструкции, созданные по технологиям виртуальной реальности и снабженные гипертекстовыми и мультимедийными структурами, могут помочь формированию значительно более целостного, наглядного и детального представления, дадут возможность погрузить музейные объекты в историко-культурный контекст, предоставят возможность самостоятельного обучения и исследования. Публикация таких моделей в сети Internet и/или на компьютере (т.н. музейном киоске) должна дать новую дополнительную возможность огромной аудитории значительно расширить свои представления о культурной и исторической ценности музейной экспозиции, а также о существующих в музее проектах реставрации.

Наиболее распространенным является применение этой технологии в рамках технологий 3D в интерактивном панорамном видео и мультимедиа CD-ROM.

Интерактивное панорамное видео, в настоящее время реализовано на сайтах Дарвиновского музея, Государственной Третьяковской галереи, Государственного музея изобразительных искусств им. А.С. Пушкина, создает образ мира на основе реальных изображений. Программные технологии типа Quick Time Virtual Reality или Surround Video преобразуют панорамное изображение так, что плоское отображаемое пространство воспринимается как трёхмерный мир.

В мультимедийных CD-ROM наличие виртуальной реальности стало уже чуть ли не стандартом. Уровень современных CD-ROM-технологий позволяет организовать 360-градусный обзор в любой плоскости без ограничений. Причем инструментальные пакеты, применяемые при создании таких дисков, обеспечивают наиболее точное воссоздание объектов в 3D и высокое качество изображений. Однако, по имеющимся сведениям, затраты на разработку мультимедийного компакт-диска составляют от 10 до 30 тысяч долларов США. Создание же собственными силами 3D-моделей авторскими средствами разработки CD-ROM на персональных компьютерах нереально, т.к. требует также немалых затрат на приобретение готовых инструментальных пакетов. К тому же модель по технологии CD-ROM нельзя опубликовать в Internet.

Нами, сотрудниками сектора археологической теории и информатики ИАЭТ СО РАН, было принято решение обратиться к третьему типу 3D-технологий – **VRML**, который является одним из основных и наиболее развитых инструментов разработчика трёхмерных миров в Internet. **VRML** расшифровывается как язык моделирования виртуальной реальности (Virtual Reality Modeling Language).

Этот платформенно независимый язык разметки напоминает хорошо известный разработчикам Web-страниц язык HTML, но отличается возможностями изображения трёхмерных объектов, источников освещения, управляемого пользователем и автоматического перемещения «видеокамеры».

Документ VRML применительно к музейному контексту представляет собой просто текстовый файл, который содержит описания залов, свойств их поверхностей (цвет, текстура материала, освещение и т. п.) и изображений музейных экспонатов. VRML-документ запрашивается с Web-сервера и поступает пользователю в виде исходного текста, точно также, как и уже давно ставший привычным HTML-документ. Просматривающий VRML-документ и преобразующий его текст в трёхмерную графику браузер должен иметь т. н. VRML-plugin, создающий на экране монитора иллюзию пространственного движения наблюдателя и окружающих его объектов.

Создание трехмерных интерактивных миров к тому времени, когда его признали актуальным для музея, было для сотрудников сектора совершенно неосвоенной областью. Наибольшие сомнения вызывал вопрос качества получаемых по VRML-технологии изображений, их восприятия посетителями музея в виртуальной реальности. Поэтому было первоначально решено разработать экспериментальную модель музейных объектов на основе уже имеющихся на сайте Института экспонатов по материальной и духовной культуре обских угров. Модели в VRML принято называть сценами.

Выбор темы для экспериментальной VRML-сцены должен был позволить задействовать и оценить такие привлекательные в музейной практике возможности виртуальной реальности, как возможность путешествия по реально несуществующим, залам музея.

В ходе эксперимента был создан ряд комнат (исследователи угорских народов, изделия из металлов, атрибуты медвежьего праздника, жертвенные покрывала и др.). Это обеспечило возможность на материале этих комнат проводить на компьютере операции, обычно выполняемые художником экспозиции по её архитектурно-художественному проектированию. Созданная модель позволила разработать пространственное и цветовое решение экспозиции.

Первоначально на участках VRML-сцены были размещены ссылки на HTML-документы (гипертекст), а в будущем предполагается разместить и ссылки на аудио- и видео-документы (гипермедиа). Таким образом, размещение в VRML-сцене ссылок на тексты, фотоизображения, видео, музыку, речь и другие звуки предоставит возможность формировать компьютерные экспозиции, в которых обеспечивается:

1. многогранное, многосредовое представление музейных предметов;
2. существенно большая информативность модели для пользователя;
3. возможность подключения к информации, находящейся на других сайтах, без прерывания обзора экспозиции в VRML-сцене.

Гипермедиа, по утверждению специалистов, погружает пользователя в сферу усиливающих ощущений (когда одни ощущения порождают другие), обостряя восприятие виртуального мира и оказывая таким образом эмоциональное воздействие, сравнимое с тем, которое можно получить в компьютерных играх.

В стандарте VRML 2.0 предусмотрена возможность создания элементарных анимаций объектов. Для более сложных анимаций применимы скрипты на Java или JavaScript, а также и инструментарий Parallel Graphics, поддерживающий эти возможности VRML.

Анимации, реализованные в такой модели, предназначены для нетрадиционного общения посетителя с виртуальными музейными предметами.

При просмотре сцены, представляющей экспозицию выставки, в стандартном Internet-браузере можно, например, зажечь очаг в чуме, или взять для более внимательного осмотра находящихся там предметов быта. Эффект может еще усиливаться при просмотре видеоролика о хозяйстве, быте и обрядовой сцене.

Это способствует созданию у посетителя виртуальной экспозиции эмоционального эффекта присутствия в чуме в качестве его хозяина, на принципиально новом уровне погружает его в культурно-исторический контекст, даёт наглядное представление о функциональных возможностях экспонируемого образца.

Кроме того, лёгкость перемещения в пространстве внутри и вне здания музея в горизонтальной и вертикальной плоскостях, возможность пребывания в любом помещении (в том числе несуществующем в реальности), получения любого ракурса или перспективы из любой площадки, возможность "птичьего полета", вращения горизонта создают особое настроение причастности, погруженности в историко-культурный контекст, стимулируют желание всесторонне исследовать объект, создают особый эмоциональный настрой и наделяют опытом, который практически невозможно получить другим способом.

Завершение разработки музейного сайта, обладающего подобными возможностями, несомненно, повысит свою привлекательность и для широкой Internet-аудитории, и для музейных специалистов.

Можно сказать, что с возможностью погружения музейных объектов в VRML-миры представление определённого класса музейных объектов приобретает новое измерение. Использовать ли это новое измерение – вопрос для музейных специалистов. Можно предложить следующие способы его применения.

Это измерение можно ввести в музейные базы данных, дополнив ссылки на плоские изображения музейных предметов ссылками на файлы с их VRML-представлениями, если они имеются. А пользовательские интерфейсные оболочки таких баз данных можно дополнить подключением к средствам просмотра 3D.

Имея доступ к VRML-моделям объектов через сеть Internet или музейную информационную систему, музейный специалист получит возможность ознакомиться с интересующим его предметом в деталях (механикой, динамикой, функциональностью, внутренним устройством, архитектурными особенностями неподвижного памятника) без выезда в командировку или посещения места хранения. Это сократит усилия и сроки написания научных работ, предварительного знакомства при отборе и группировке вещей на выставки, проведения сравнительного анализа с предметами в других музеях.

В музеях, университетских курсах по музееведению, археологии, этнографии и школьных кружках по изучению истории использование VRML-моделей музейных объектов облегчит изучение тех предметов, непосредственная манипуляция которыми в музее не разрешается, или предметов с недоступными или утраченными элементами.

Наличие готовых VRML-моделей музейных объектов (в музейной информационной системе, например) даст возможность сравнительно быстрого построения сцен выставок и интерьеров.

Разумеется, VRML имеет недостатки и ограничения, из которых особо следует отметить следующие:

качество прорисовываемых из документа VRML изображений заметно ниже качества CD-ROM-изображений.

при скорости модема меньше 28.8 Кбод и объёме оперативной памяти меньше 16 MB VRML-программы, получаемые из Internet, работают медленно, неустойчиво и с ошибками.

Однако в условиях увеличения пропускной способности отечественных сетей и смены поколений компьютерного парка эти ограничения уже сейчас практически перестают быть реальным сдерживающим фактором применения VRML.

В перспективе музей мог бы представлять археологические и этнографические коллекции по всем регионам Сибири и Дальнего Востока.

В настоящее время для обозрения этнографических коллекций представлено два региона: Западная Сибирь и Дальний Восток. В указанной экспозиции пока представлен лишь раздел этнографии.

Осмотр экспозиции начинается с зала "Ведущие этнографы Западной Сибири". В зале посетители могут ознакомиться не только с помещенными фотоматериалами, но и с краткими биографическими сведениями, списками опубликованных работ, а в будущем предполагается сделать гиперссылки на работы, опубликованные в Интернет (Рис.52).

Из зала Персоналий посетитель мог по выбору посетить ряд экспозиций³:

- Изделия из металла
- Атрибуты медвежьего праздника и ритуальные колчаны
- Жертвенные покрывала
- Искусство хантов Нижней Оби

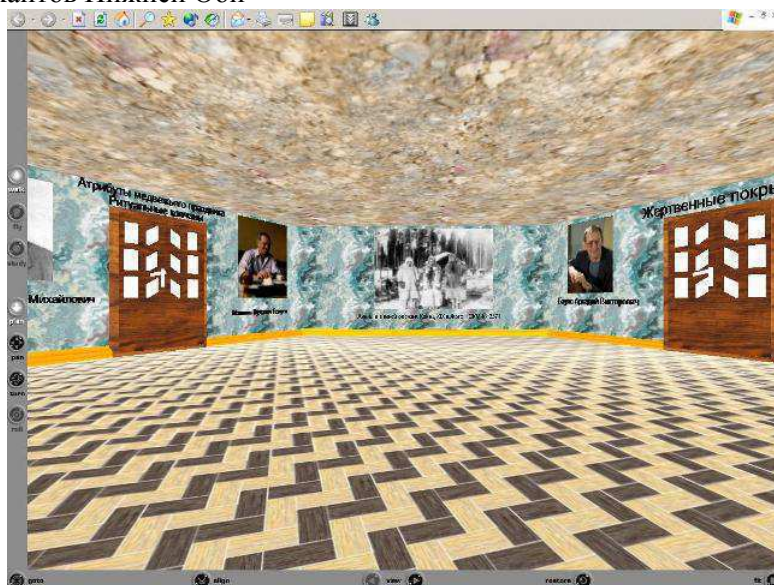


Рис.52. Комната “Видные этнографы Западной Сибири”

В экспозиции “Изделия из металла” помещены для просмотра изделия из серебра, бронзы, а также олова и свинца. Материалы для экспозиции представлены д.и.н. А.В. Бауло (Рис.53).

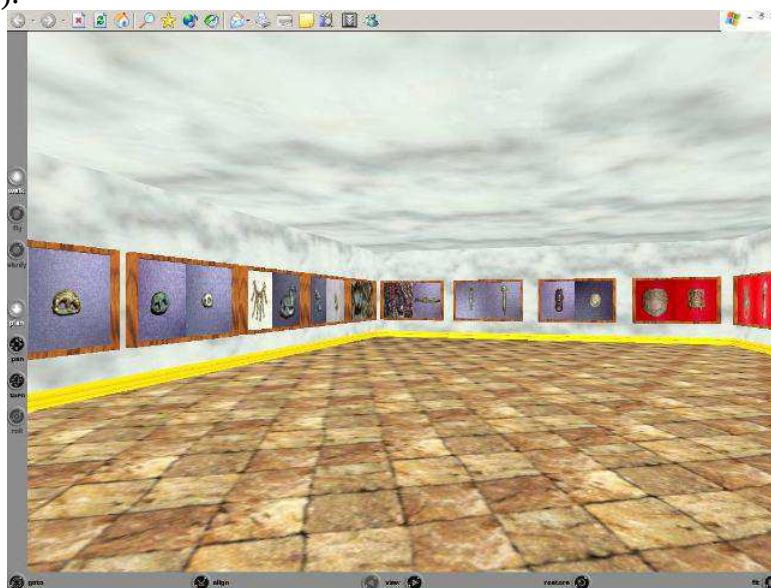


Рис.53. Комната “Изделия из металла”

В комнате “Атрибуты медвежьего праздника и ритуальные колчаны” собраны экспонаты из “Атрибуты медвежьего праздника” и “Ритуальные колчаны.” Из предметов атрибутики медвежьего праздника у сынских хантов в экспозиции представлены маски,

³ Бауло А.В. Жертвенные покрывала обских угров из Ханты-Мансийского музея // Народы Сибири: история и культура.-Новосибирск,1997.-С.90-101

Бауло А.В. Ритуальные колчаны обских угров // Гуманитарные нау-ки в Сибири. - 1999 б. - № 3 - С. 36 - 39.

Бауло А.В., Маршак Б.И. Серебряный ритон из хантыйского святилища // Археология, этнография и антропология Евразии. - 2001. - № 3. - С. 164 - 172.

Гемуев И.Н. Еще одно серебряное блюдо из Северного Приобья // Изв. СО АН СССР Сер. История, филология и философия. Новосибирск, 1988. - № 3, вып. 1

Гемуев И.Н. Мировоззрение манси: Дом и Космос. - Новосибирск, 1990 а

В музее представлены материалы сайта "Sibirica":

<http://www.sati.archaeology.nsc.ru/Museums/DecorateArt/>

выполненные из дерева, Их практически нет в музеях России; крупнейшая коллекция хранится в музее пос. Овгорт Шурышкарского района Ямало- Ненецкого АО (рис.54).

В экспозиции виртуального музея использованы также материалы полевых исследований А.В. Бауло у сынских хантов в 2000 – 2002 гг. В экспозиции представлены также шкуры медведей из домашних святилищ хантов и колчаны обнаруженные в поселке Тугияны (р. Обь) Белоярского района Тюменской области.



Рис.54. Комната “Атрибуты медвежьего праздника и ритуальные колчаны”

Галерея “Жертвенные покрывала обских угров” создана на основе материалов полевых исследований д.и.н. И.Н. Гемуева и к.и.н. А.В. Бауло. Она позволяет впервые познакомиться с неопубликованными материалами этих авторов и комплексом вопросов, связанных с феноменом жертвенных покрывал в религиозно-обрядовой практике северных групп манси и хантов. Вводит в курс проблем генезиса и эволюции жертвенных покрытий, семантики имеющихся на них изображений, использования данного типа культовой атрибутики во время религиозных церемоний (рис55).

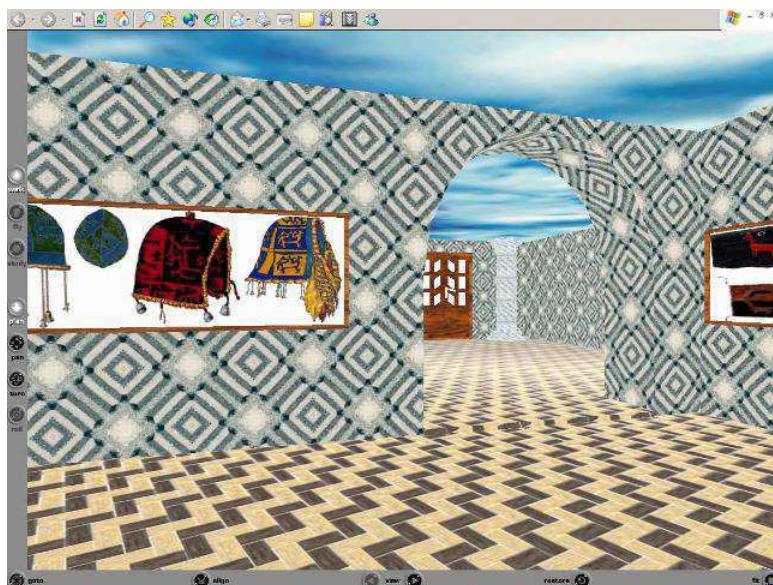


Рис.55. Комната “Жертвенные покрывала”

В Галерее "Декоративно-прикладное искусство хантов нижней Оби" представлены одежда, обувь, бытовые вещи и предметы культа части северных хантов, проживающих по рекам Сыня, Куноват и на участках, прилегающих в этом регионе к Оби (Шурышкарский район Ямала-Ненецкого автономного округа). Материалы для размещения в музее любезно предоставлены кандидатом исторических наук Антониной Макаровой Сязи.

При разработке концепции музейного Интернет-портала мы пошли на создание сложной наукоемкой системы, включающей в свой состав: музейные коллекции, полнотекстовую библиотеку, биографическую базу данных, географическую информационную систему, археолого-этнографический словарь, новостную ленту RSS, систему ссылок на мировые и российские музейные ресурсы, кинозал с этнографическими фильмами и аудио [Холюшкин., Воронин, Ильиных, 2005] (Рис.57-58).

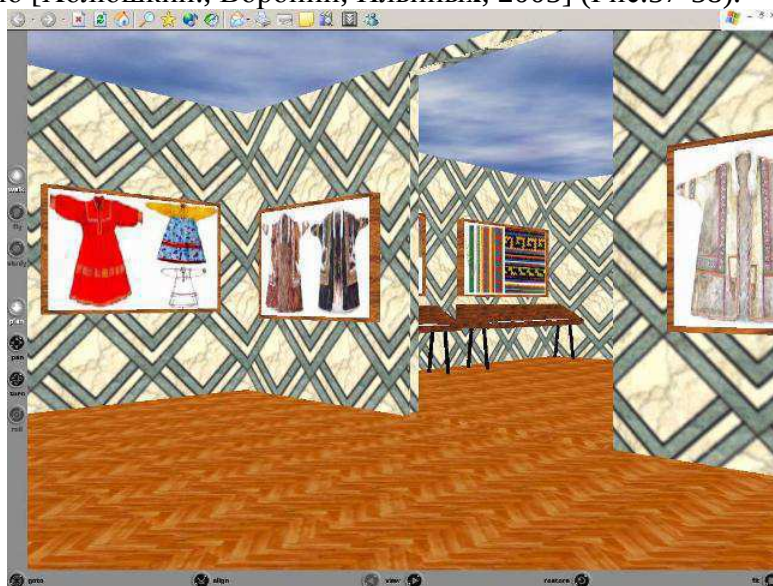


Рис.56. Комната “Искусство хантов Нижней Оби”

Фотография и название объекта являются гиперссылками, нажав на одну из которых, пользователь получит всю имеющуюся информацию об этом объекте в базе данных (рис. 59).



Рис. 57. Кинозал виртуального музея.

В результате этих поисков пользователь отбирает таким способом нужные экспонаты, чтобы из них составить виртуальную экспозицию. Для наиболее наглядного представления этой коллекции в поисковой системе портала используется генератор залов. Для того чтобы запустить процесс генерации залов для последующего путешествия по ним пользователя требуется, чтобы к моменту этого запуска на клиентском компьютере был установлен браузер Cortona.

Генератор динамически создаёт на языке VRML готовые к просмотру залы с экспозициями, размещает в них отобранные экспонаты и обеспечивает удобный и комфортный доступ пользователя к их содержанию. Эта часть программы выполняет задачи:

- генерация залов с виртуальными экспозициями, в том числе создание стен, пола, потолка, других элементов интерьера в каждом зале; выбор цвета для них происходит

случайным образом из набора заготовленных текстур;

– размещение экспонатов в залах (расчёт местоположения экспонатов на стенах в залах с экспозициями); при этом если, к примеру, экспонат не помещается на стене, то его размеры в разумных пределах будут уменьшены с тем, чтобы он все же смог бы поместиться на стене; однако если и разумные пределы помещения не позволяют разместить в очередном зале все оставшиеся неразмещенные экспонаты, происходит генерация нового зала и его соответствующее наполнение;

– свободный вход и путешествие пользователя по сгенерированным залам;

– просмотр информации об экспонатах в каждом сгенерированном зале.

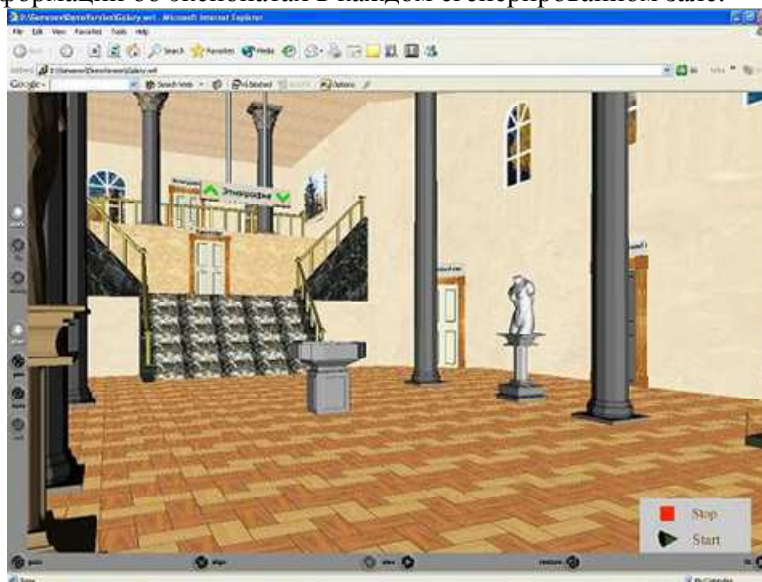


Рис. 58. Вестибюль Виртуального музея

Сгенерированные залы создают полную иллюзию помещений реального музея. С помощью графического интерфейса пользователю предоставляются мощные средства навигации, обеспечивающие возможности свободного и комфортного путешествия по залам и доступа к размещенным в них экспонатам (Рис. 60-61).

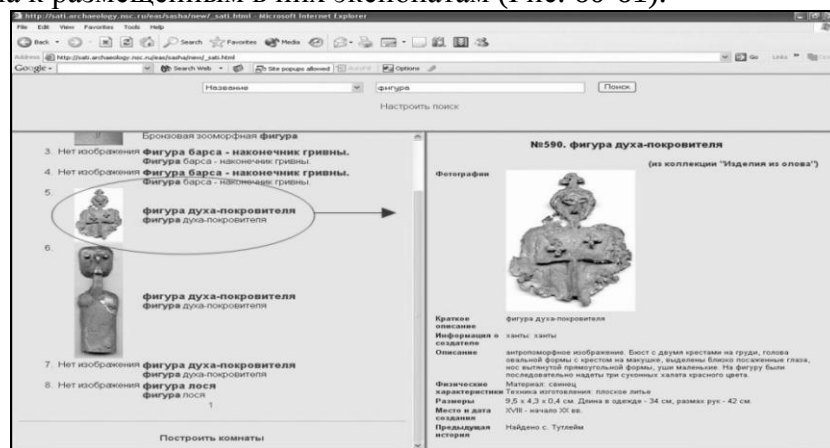


Рис. 59. Вывод полной информации об экспонате.

Компьютерные трёхмерные реконструкции, созданные по технологии виртуальная реальность и снабженные гипертекстовыми и мультимедийными структурами, могут помочь формированию значительно более целостного, наглядного и детального представления, дадут возможность погрузить музейные объекты в историко-культурный контекст, предоставят возможность самостоятельного обучения и исследования.

В мультимедийных CD-ROM наличие виртуальной реальности стало уже чуть ли не стандартом. Современный уровень CD-ROM-технологий уже сейчас позволяет организовать 360-градусный обзор в любой плоскости без ограничений.

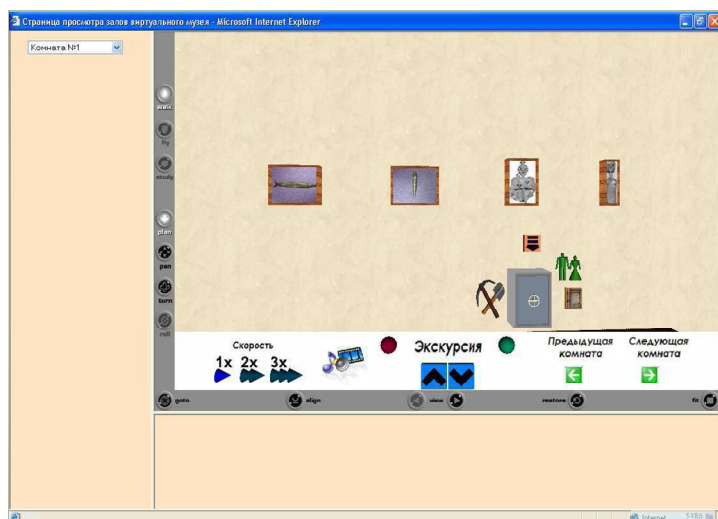


Рис.60. Генератор залов.

Создание же собственными силами 3D-моделей авторскими средствами разработки CD-ROM на персональных компьютерах нереально, т.к. требует также немалых затрат на приобретение готовых инструментальных пакетов.

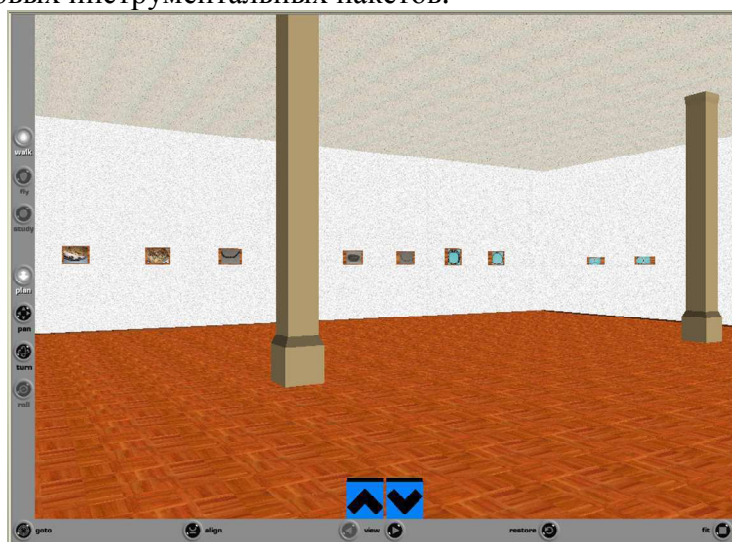


Рис. 61. Сгенерированный зал

К тому же модель по технологии CD-ROM нельзя опубликовать в Internet. Заметный вклад в реализацию программы сыграли В.М. Семенов, Е.В. Горбунов, А.Ю. Подчасов [О некоторых подходах..., 2006: с.31-40]

2.10. ГИС в музейных исследованиях

В последнее время в связи с развитием возможностей информационных технологий, наблюдается расширение приложений информационных систем в исторических исследованиях. Помимо традиционного текста с рисунками появилась возможность использовать электронные карты, планы, снимки, таблицы. Примером таких продвинутых программных средств, адекватных задаче описания исторического развития территории, являются геоинформационные системы.

В подобных проектах для работы создаются междисциплинарные коллективы, состоящие из социологов, историков, программистов и представителей других смежных областей наук.

Информация, представленная в виде фотографий и их описаний, текстовых документов полученных в ходе археологических изысканий, чертежей, исторических текстов, видеоматериалов, объединяется на основе ГИС, которая обеспечит привязку данных по месту к электронной карте и их упорядочивание во времени.

Рассмотрим технологические решение и пути дальнейшего развития проекта, реализованного в ИАЭТ СО РАН (рис.63)

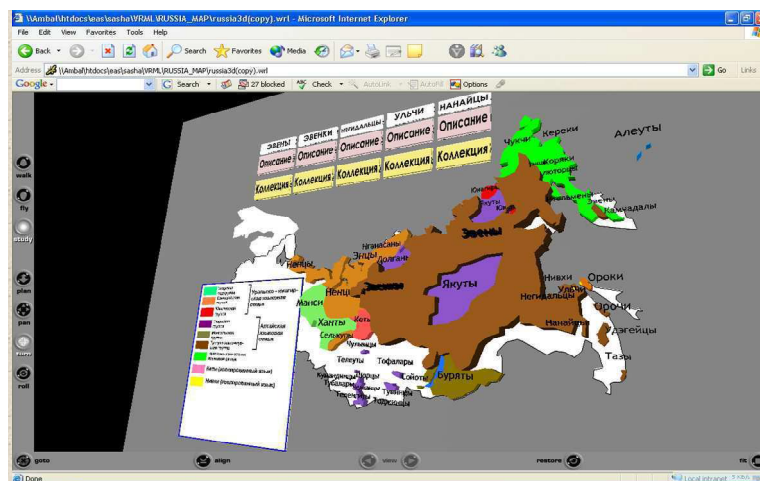


Рис. 62. Карта расселения народов Сибири (по А.Ю. Подчасову)

При разработке проекта ГИС-систем “Духовная культура” и “Материальная культура”, в качестве исходного материала, были использованы снимки карт с CD-ROM “Атлас Мира; Atlas Deluxe of Rand McNally” и растровые карты, отсканированные с «Историко-этнографического атласа Сибири» (рис.64).

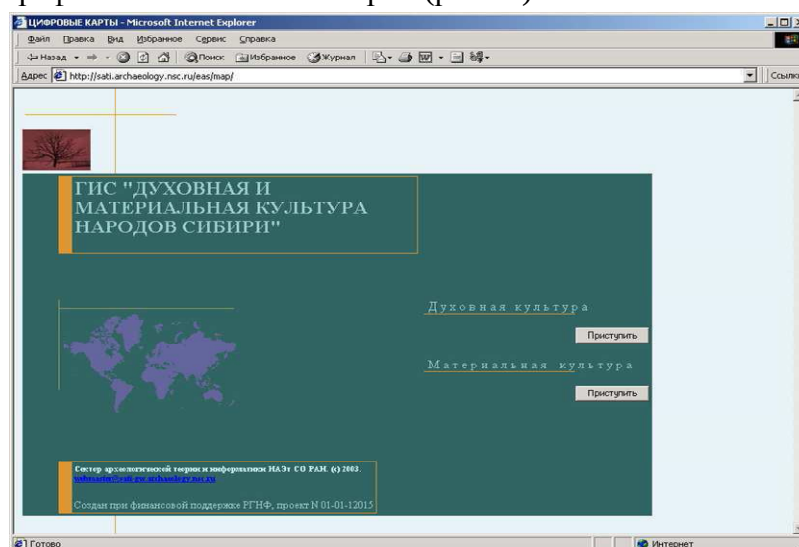


Рис. 63. главная страница ГИС

Для создания полноценной интерактивной картографической справочной системы, карт растрового формата недостаточно, для этого необходима векторизация имеющихся карт. Для решения данной проблемы, был использован, так называемый векторизатор Easy Trace.

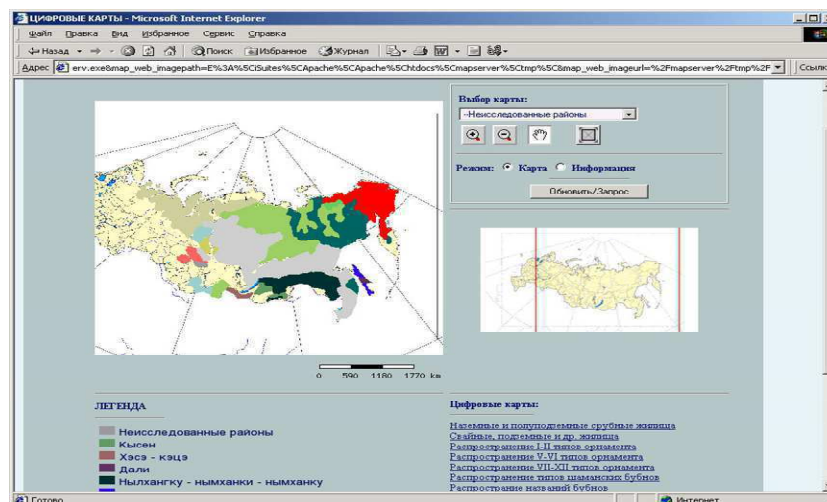


Рис 64. Территориальное распределение шаманских бубнов

Easy Trace является пакетом программ для полуавтоматической интерактивной векторизации цветных и черно-белых растровых изображений, работающим под Windows 95/NT. Он предназначен для переноса графической информации с бумажных носителей в компьютер и ориентирован, прежде всего, на обработку картографических материалов.

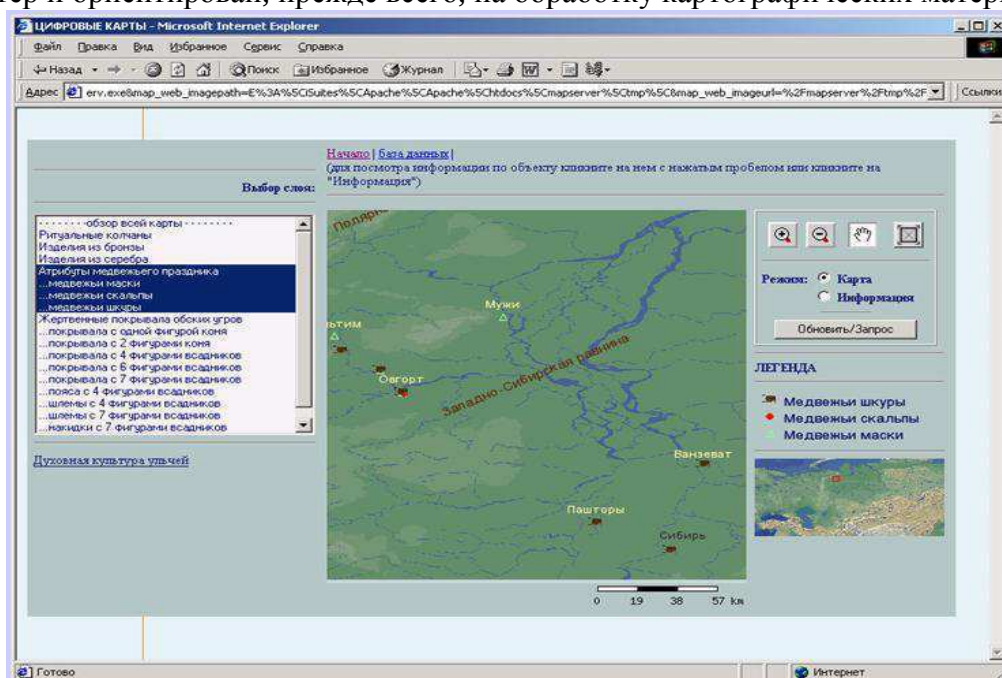


Рис. 65. Поисковая страница ритуальной атрибутики манси (Западная Сибирь)

Отметим, что данный пакет позволяет создавать проектные файлы, подключать к ним растровые слои, производить преобразования растровых подложек и т.д.

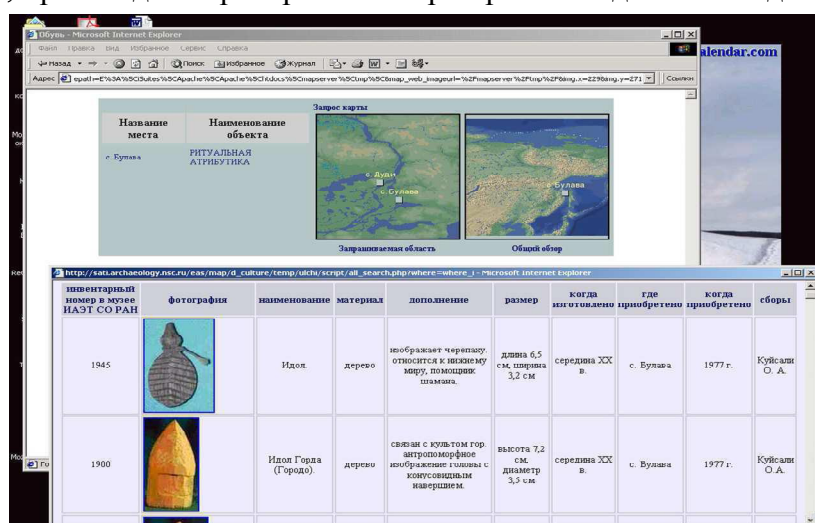


Рис. 66. Извлечение базы данных по духовной культуре ульчей.

Помимо замены растровых изображений на векторные примитивы с последующей передачей их в ГИС (САПР), программа Easy Trace позволяет в процессе векторизации решать следующие задачи:

- Минимизация числа векторных примитивов (две пересекающиеся линии разных слоев должны остаться двумя линиями, а не четырьмя, сошедшимися в одной точке);
- Восстановление информации, частично утраченной или искаженной из-за износа бумажного носителя, дефектов чертежных инструментов, дефектов исполнения, погрешностей сканирования;
- "Расслоение" изображения по его смысловому содержанию (например, карта может содержать слои рельефа, автодорог, коммуникаций и т.д.);

- Введение атрибутивной информации для графического объекта (например, напряжение линии электропередач, диаметр трубопровода, площадь земельного участка, его собственник и т.п.) в формате DBF;

- Построение корректной топологической структуры информации, соответствующей требованиям конечной ГИС или САПР.

Таким образом, с помощью программы-векторизатора можно создавать файлы векторных и атрибутивных данных, несущие в себе гораздо больше информации, чем исходный бумажный материал, и превышающие его по точности.

Несмотря на вышеперечисленные достоинства векторизатора Easy Trace, использование данного программного комплекса позволяет нам только подготовить шаблоны векторных карт, с последующей их обработкой в среде ArcView GIS Version 3.0 компании ESRI.

ArcView предоставляет более широкий спектр по созданию и редактированию справочных картографических систем. Это и добавление атрибутивной информации по соответствующим объектам, и собственно создание объектов. Также ArcView - мощный, легкий в использовании инструмент для обеспечения доступа к географической информации. Данный программный комплекс дает широкие возможности для отображения, изучения, выполнения запросов и анализа пространственных данных..

После проведения необходимых преобразований и редактирования данных в среде ArcView, полученные шейп-файлы были опубликованы в Internet, с помощью картографического сервера MapServer 4.0, первоначальные версии которого разрабатывались в the University of Minnesota (UMN). Дальнейшее совершенствование и поддержку данного продукта проводит открытое сообщество OpenSource.

Отметим, что главная часть MapServer - это CGI-приложение для публикации GIS и снимковых данных в Интернет (WWW), для просмотра которых достаточно иметь браузер Internet Explorer. Пакет содержит несколько отдельных программ для визуализации карт, создания масштабных линеек и легенд, для более удобной работы со справочной системой.

Таким образом, MapServer представляет собой программную оболочку, позволяющую даже неискушенному пользователю, получить необходимую информацию по интересующему его объекту, без использования особых навыков.

Отметим, что структура MapServer-а такова, что атрибутивные данные хранятся в формате dbf, т.е. для внесения изменений, в существующую базу данных, требуется использование специальных программных средств, таких как ArcView. Это несколько затрудняет сопровождение, и обновление справочной системы в режиме реального времени.

Для решения этой проблемы была осуществлена следующая связка:

MapServer – JavaScript – Php

Опишем вкратце эту технологию.

База данных объектов, привязанных к географическим местам, уже была некоторым образом реализована в формате XML, т.е. попутно встала задача состыковки данных выводимых системой MapServer и данных для объектов из XML-файла.

Визуализация данных из этой базы данных непосредственно осуществляется через PHP-скрипты. Забегая вперед, отметим, что, посредством PHP-скриптов производится также администрирование этой базы данных и её редактирование.

Используя программные особенности картографической справочной системы MapServer, при запросе пользователем данных по интересующему его объекту, посредством JavaScript-а производится отправка ключевых параметров этого запроса к PHP-скрипту. Т.е. происходит дублирование и отправка запроса параллельно запросу MapServer-у.

В свою очередь, PHP-скрипт направляет, сконструированный определенным образом, параллельный запрос к базе данных формата XML и выводит запрошенные пользователем

данные на экран монитора. Также, пользователь может при желании просмотреть другие материалы по данному объекту. Это и статьи, и фото-видеоматериалы.

Как видим, пользователь не замечает, того, что обрабатывается XML-файл, а не только база данных MapServer-a. Конечно, для подобной реализации потребовалось привести данные формата XML и dbf-файлы, к определенному виду. Впрочем, в настоящее время, единственное, что требуется для работы этой системы, это наличие ключевого поля в dbf-данных. Всё остальное модернизируется независимо от MapServer-a, в PHP-скриптах.

Администратор базы данных, не прибегая к помощи специализированных программных средств редактирования ГИС, может производить обновление базы данных ГИС-систем “Духовная культура” и “Материальная культура” в режиме реального времени. Отметим следующую особенность реализованной системы, что, по-нашему мнению, является положительным элементом доступности и открытости, обновление базы данных может осуществляться из любой точки земного шара, где есть доступ к Internet и компьютер, редактирование данных может быть произведено прямо через любой браузер.

Стоит отметить и тот факт, что открытость системы достигается за счет использования xml-формата хранения данных. Как известно, в настоящее время xml-формат, позиционируется в сети как единый формат хранения данных, одновременно содержащий механизм, позволяющий определить структуру документа и описывать содержащиеся в нем элементы.

Язык XML предназначен для создания новых языков разметки. С его помощью можно описать целый класс объектов данных, называемых XML - документами, ориентированными на конкретную предметную область. XML позволяет определить допустимый набор тэгов, их атрибуты и внутреннюю структуру документа. Тэги (подобно тэгам в HTML) представляют специальные инструкции, предназначенные для формирования в документах определенной структуры и четких отношений между различными элементами этой структуры.

Можно выделить следующий круг задач, связанных с созданием и обработкой структурированной информации, для решения которых может использоваться XML:

- Разработка сложных информационных систем, с большим количеством приложений, связанных потоками информации самой различной структуры. XML - документы исполняют роль универсального формата для обмена информацией между отдельными компонентами большой программы.
- XML является базовым стандартом для нового языка описания ресурсов, RDF, позволяющего упростить многие проблемы в Web, связанные с поиском нужной информации, обеспечением контроля за содержимым сетевых ресурсов, создания электронных библиотек и т.д.
- XML может использоваться в обычных приложениях для хранения и обработки структурированных данных в едином формате.
- XML позволяет описывать данные произвольного типа и используется для представления специализированной информации.
- XML может служить мощным дополнением к HTML для распространения в Web "нестандартной" структурированной информации
- XML-документы могут использоваться в качестве промежуточного формата данных в трехзвенных системах при поиске информации в удаленных базах данных. Сегодня на рассмотрение W3C предложена спецификация нового языка запросов к базам данных XQL.
- Информация, содержащаяся в XML-документах, может изменяться, передаваться на машину клиента и обновляться по частям. Разрабатываемые спецификации XLink и Xpointer позволяют ссылаться на отдельные элементы документа, с учетом их вложенности и значений атрибутов.

- Использование стилевых таблиц (XSL) позволяет обеспечить независимое от конкретного устройства вывода отображение XML- документов и фильтрацию данных.

Таким образом, хранение данных в XML-формате, позволяет значительно расширить круг использования данных в различных приложениях.

На текущий момент среда разработки MapServer поддерживает технологию MapScript, благодаря которой возможен доступ к функциям MapServer из различных скриптовых языков программирования, в частности это может быть PHP.

Благодаря использованию технологии MapScript, возможно расширение функциональных возможностей MapServer-a.

В частности, это могут быть:

- Автоматизация создание слоев на базе имеющихся объектов;
- Обработка расширенных запросов пользователя;
- Вычисления расстояний между интересующими объектами.

В данной реализации этих двух ГИС-систем для вывода слоев необходимо, прописывание формата объектов заданному слою в файле расширения *.map. Слои формируются строго в вышеуказанном файле, т.е. описываются статически, что значительно сужает количество определенных слоев. Избавится от этого ограничения, позволяет модуль MapScript.

Применение MapScript технологии дает возможность строить динамические слои исходя из потребностей пользователя. Что, в конечном счете, позволит значительно сократить сроки разработок новых ГИС, обновления существующих, и упростит процедуру её размещения в сетях Intranet и Internet. Любой пользователь может смоделировать слой, совмещающий в себе именно те объекты, которые интересуют только его. Это позволит построить систему, удовлетворяющую большинству её пользователей.

Таким образом, совмещая XML, PHP, MapServer и MapScript можно получить достаточно гибкую и открытую систему которая позволит работать пользователю с системой не испытывая особых затруднений.

Использование новых технологий, позволяет нам получить открытую систему, которая может постоянно развиваться, в зависимости от качественного изменения потребностей пользователей данных систем.

Источники:

1. <http://mapserver.gis.umn.edu/index.html>
2. <http://mapserver.gis.umn.edu/doc.html>
3. <http://www.easytrace.com/work/russian/news.html>
4. <http://www.xml.org/>

Электронный архив — система структурированного хранения электронных документов, позиционируемая как основа документооборота, обеспечивающая надежность хранения, конфиденциальность и разграничение прав доступа к документам.

Для преодоления логического препятствия археологам страны необходимо начать совместные разработки по упорядочению понятий в предметной области археологии. Под предметной областью (ПО) понимается совокупность объектов, «рассматриваемых с точки зрения некоторого предмета исследования – совокупности существенных свойств (атрибутов) и отношений объектов исследования, описываемых в некоторой системе понятий предметной области. Предмет исследования может быть задан онтологией предметной области – специфицирующей в некотором формальном языке множество рассматриваемых объектов, связи между ними, систему понятий, и свойства объектов. Предмет исследования и онтология определяют "взгляд", "точку зрения", с которой рассматриваются (описываются в системе понятий) объекты предметной области, отношения и их свойства» [Витяев, 2010: с. 9-16].

Таким образом, онтология в археологии представляет собой систему, описывающую структуру определенной проблемной области, и состоящую из множества классов понятий, связанных отношениями, их определений и аксиом, задающих ограничения на интерпретацию этих понятий в рамках данной проблемной области. Онтология, как пример общего соглашения о семантике области, способствует установлению корректных связей между значениями элементов этой области, тем самым создавая условия для их совместного использования.

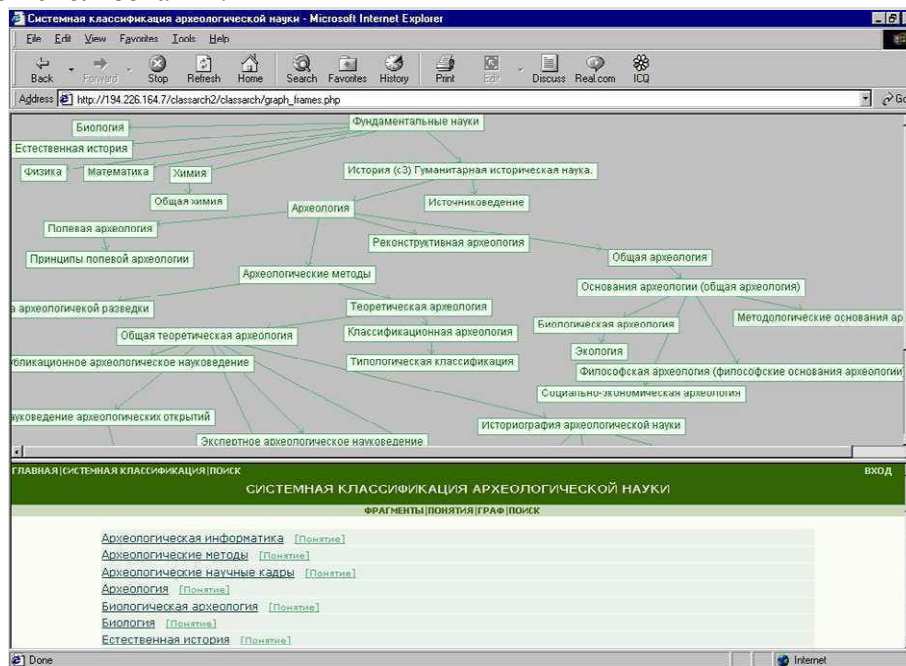


Рис. 67. Граф фрагментов системной классификации (по Андреевой, ... 2004, с.39-44).

Суммируя приведенные выше определения онтологии, можно сказать, что онтология представляет собой точное описание (модель) некоторой части мира применительно к конкретной области интересов [Андреева, Сергеев, Холюшкин, 2004, с. 39-44; Андреева и др., 2006: с. 832-840].

Таким образом, онтология – это четверка вида $\langle C, D, R, A \rangle$, где:

C – множество понятий конкретной предметной или проблемной области;

D – множество определений понятий;

R – множество отношений (связей) между понятиями;

A – множество аксиом.

Попытки построения фрагментов онтологий в археологии уже предпринимались в виде унификации полевых работ [Медведев, Алаев, Алаева, 1978], классификаций артефактов [Медведев, 1981] и каталогизации музейных коллекций [Асеев, Поднозова, Шер, 1980]. Однако в то время они не получили широкого распространения, поскольку задача онтологизации науки еще не созрела. Сознвая эти трудности, Медведев писал: «создание хотя бы одного определенного массива терминологии разовым порядком с целью придания ему законодательного ранга невозможно» [Медведев, 1981; с. 18]. Идея интеграции научного документооборота в виде общих представлений была отражена в ряде публикаций сектора археологической теории и информации (САТИ) [см. Холюшкин, Воронин, 2005]. По мере реализации были созданы информационные ресурсы, описываемые ниже.

2.11. Система знаний информационного портала по археологии и этнографии

В 2000-х гг., на фоне быстрого роста числа и объема тематических ресурсов, рассматривались концепции, расширяющие понятие баз знаний до уровня web-порталов знаний. В этот период в изданиях ИТГИ появился ряд публикаций математиков ИСИ СО РАН и Ю.П. Холюшкина, в которых анализировались возможности создания

интеллектуальных систем представления, обработки и анализа данных для решения задач поиска и извлечения информации из ресурсов Интернета⁴.

Портал знаний. Одним из главных направлений использования системной классификации археологического знания является портал знаний по археологии и этнографии, разработанный сотрудниками сектора в содружестве с Институтом систем информатики СО РАН. Его создание ориентировано на наиболее адекватный способ представление и структурирование предметных знаний по археологии и этнографии в глобальной сети. Решение этой задачи предполагает не только обеспечение целостного представления предметной области археологии и смежным дисциплин, но и навигацию по релевантным ей информационным ресурсам и легко настраиваемый на выбранную область знаний контент. Концептуальная и информационная основа портала в форме онтологии предметной области включает четыре базовых иерархии (иерархия разделов, иерархия объектов, иерархия методов исследования, и иерархия научных результатов).

Основой для построения иерархий понятий онтологии по археологии и этнографии послужила предложенная в (Холюшкин, Гражданников, 2000; Холюшкин, 2004) и развиваемая в настоящее время системная классификация археологической науки, фиксирующая явные и неявные связи между используемыми в ней понятиями. Привязка понятий системной классификации к онтологии портала производится таким образом, что понятия системной классификации используются для построения классов и доменов онтологии и для создания экземпляров понятий онтологии.

Так, некоторые понятия системной классификации объявляются классами онтологии и образуют упрощенную иерархию наследования. Это могут быть научные результаты (гипотеза, археологический факт, научное достижение и открытие), объект исследования (артефакт, комплекс), методология (методики и методы).

Информационную основу портала составляет онтология [Guariano, Giaretta, 1995.], подразделяющаяся на универсальную онтологию науки, служащую для представления понятий, необходимых для описания научной деятельности и научного знания в целом, и онтологию предметной области, представляющую описание конкретной научной дисциплины. Последняя онтология определяет систематизацию знаний и информационных ресурсов, относящихся к данной предметной области (археологии), а, следовательно, удобство доступа к ним. Поэтому именно от нее, в конечном счете, зависит полезность портала для пользователей, описанных выше типов.

Требования к онтологии портала знаний

Процесс разработки онтологии (согласно методологиям, представленным в работах [Uschold, Gruninger, 1996; Gruninger, Fox, 1995; Fernandez-Lopez, Gomez-Perez, Pazos Sierra, 1999: 37– 46; Staab, Schunurr, Studer, Sure, 2001: 26– 34]) должен включать определение возможной области применения онтологии и задач, которые могут быть решены с ее помощью. Для эффективной поддержки поиска информации заданной тематики онтология должна обеспечивать:

- простую настройку портала на выбранную область знаний.
- интеграцию знаний и информационных ресурсов в единое информационное пространство;
- содержательный доступ и удобную навигацию по всему информационному пространству портала;

Для упрощения настройки портала на выбранную область знаний в онтологии портала необходимо выделить структуры, независимые от предметной области (ПО) портала.

⁴ Боровикова О.И. Онтология предметной области «Археология и этнография» для портала научных знаний // Информационные технологии в гуманитарных исследованиях. Вып. 12. Новосибирск, 2008. С. 20–28; Загорюлько Ю.А., Боровикова О.И., Холюшкин Ю.П. Построение предметной онтологии для археологического портала // Информационные технологии в гуманитарных исследованиях. Новосибирск, 2006. Вып. 10. С. 24–30; Марчук А.Г., Загорюлько Ю.А., Холюшкин Ю.П., Воронин В.Т. Концепция портала знаний по археологии и этнографии Сибири // Информационный Бюллетень Ассоциации «История и компьютер». 2004. №32. С. 74–75.

Чтобы обеспечить интеграцию знаний и информационных ресурсов в единое информационное пространство, онтология должна не только представлять формальное описание системы понятий проблемной и предметной областей портала, но на ее основе также должны описываться типы информационных ресурсов и их связи с другими понятиями онтологии.

Онтология должна обеспечивать такое представление понятий и отношений между ними, на основе которого можно было бы автоматически строить внутренние хранилища данных портала, осуществлять навигацию по информационному пространству портала и организовывать содержательный поиск.

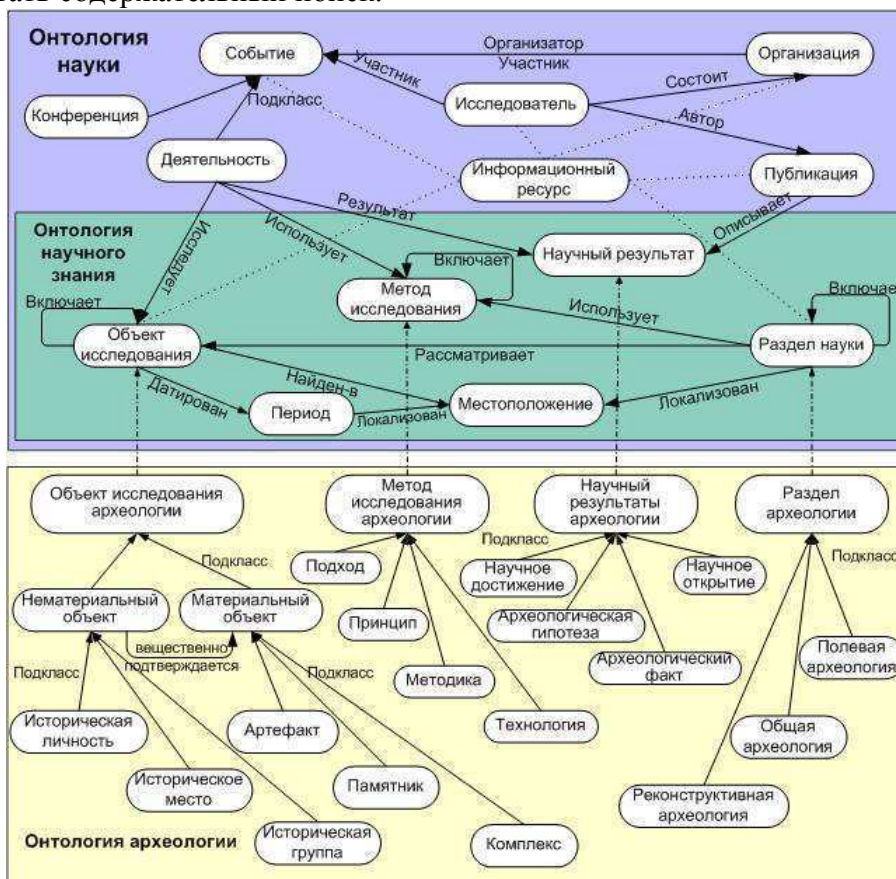


Рис.67. Фрагмент онтологии портала

Таким образом, при построении онтологии портала необходимо учитывать, что на ее основе:

- автоматически строится схема базы данных портала (генерация структуры БД и ее ограничений целостности должна выполняться по всем элементам онтологии);
- создаются формы для заполнения БД портала данными;
- определяется схема навигации по информационному пространству портала (по отношениям онтологии);
- генерируются формы поисковых запросов (по классам и отношениям онтологии).

В качестве базовых выбраны онтологии научной деятельности и научного знания [Zagorulko, Borovikova, Bulgakov, Sidorova, 2005: 45–56], которые не зависят от предметной области портала. Для описания ПО портала служит онтология ПО или предметная онтология.

Онтология научной деятельности включает базовые классы понятий, относящиеся к организации научной и исследовательской деятельности и служащие для описания (представления) субъектов и участников деятельности, такие как *Персона*, *Организация*, *Событие*, *Публикация*, *Местоположение*, *Информационный ресурс*. Назначение и список атрибутов базовых классов ранее подробно рассматривались в работе [Боровикова, Булгаков, Загорулько, Сидорова, Холюшкин, 2005: 33-39].

Онтология научного знания, по своей сути, является метаонтологией. Она содержит метапонятия и отношения, задающие структуры для описания рассматриваемой предметной области, такие как *Раздел науки*, *Метод исследования*, *Объект исследования*, *Научный результат*, *Период*, позволяющие выделить в данной науке значимые разделы и периоды времени, задать типизацию методов и объектов исследования, описать результаты научной деятельности.

Понятия онтологии научного знания связаны между собой и понятиями онтологии научной деятельности следующими ассоциативными отношениями:

"научное направление" – позволяет связывать события, публикации, организации, исследователей, информационные ресурсы с разделами науки;

"описывает" – задает связь публикации с научным результатом, объектом или методом исследования;

"использует" – связывает метод исследования с деятельностью или разделом науки;

"исследует" и "рассматривает" – сопоставляет соответственно деятельность и раздел науки с объектом исследования;

"результат деятельности" - связывает научный результат с деятельностью;

"ресурс" - связывает информационный ресурс с методами и объектами исследования;

"период локализован" - задает связь между периодом времени и географическим местоположением (например, Древняя Греция или античность).

Следует отметить, что выбор описанных выше ассоциативных отношений осуществлялся не только исходя из полноты представления проблемной и предметной областей портала, но и с учетом удобства навигации по его информационному пространству и поиска информации.

Кроме того, для организации иерархической структуры, понятия каждого из классов онтологии научного знания связаны транзитивным отношением "включает":

"объект включает" - позволяет описывать одни Объекты исследования в составе других;

"период включает" - организует хронологический порядок следования исторических периодов времени.

Онтология предметной области строится на основе базовых онтологий и отражает общие знания о предметной области, такие как иерархия классов понятий, семантические отношения на этих классах.

Построение онтологии предметной области

Онтология предметной области для научного портала знаний описывает научную дисциплину в целом как раздел науки. Она строится для организации эффективного доступа к знаниям и информационным ресурсам по определенной научной тематике (археологии) и, следовательно, также должна соответствовать описанным выше требованиям.

Важным моментом при разработке онтологии предметной области является построение иерархий понятий. При задании такой иерархии необходимо учитывать, не только то, насколько она полно описывает предметную область и связанные с ней информационные ресурсы, но и насколько удобно пользователю с ее помощью осуществлять навигацию по информационному пространству портала и вести содержательный поиск.

Разработанная нами онтология предметной области по археологии включает четыре базовых иерархии: иерархию разделов, иерархию объектов, иерархию методов исследования и иерархию научных результатов.

Основой для построения иерархий понятий онтологии по археологии и этнографии послужила предложенная в [Холушкин, Гражданников, 2000: 58 с.; Холушкин, 2004] и развиваемая в настоящее время системная классификация археологической науки, фиксирующая явные и неявные связи между используемыми в ней понятиями.

Взаимодействие понятий онтологии с системной классификацией имеет свои особенности. Понятия системной классификации используются для построения классов и доменов онтологии и для создания экземпляров понятий онтологии.

Так, некоторые понятия системной классификации объявляются классами онтологии и образуют упрощенную иерархию наследования. Это могут быть:

Научные результаты:

- “Археологическая гипотеза”
- “Археологический факт”
- “Научное достижение”
- “Научное открытие”

Объект исследования

- “Артефакт”
- “Комплекс”

Метод исследования

- “Археологические методики”
- “Археологические методы”

Эти классы в упрощенной классификации стали прямыми наследниками классов Научный результат, Объект и Метод исследования.

Другие понятия системной классификации становятся экземплярами понятий онтологии портала. В основном, к таким понятиям относятся экземпляры классов *Раздел науки* и *Метод исследования*. Связи между понятиями системной классификации в этом случае организуются на основе отношения "включает" между экземплярами классов понятий онтологии. Так, например, экземпляр класса *Раздел науки* понятие “Полевая археология” включает “Пространственную археологию” и “Стратиграфическую археологию”. А следующий экземпляр класса *Метод исследования* “Принципы общей методологии полевой археологии” включает среди прочих “Принцип сенсационности”, “Принцип подлинности”, “Принцип всесторонности”.

Некоторые понятия системной классификации используются также и при построении доменов базовых и предметной онтологий. К примеру, понятие “Археологические организации” используется при задании значения домена, определяющего типологию археологических организаций – музеев и институтов. А понятие “Исследователи-археологи” нашло свое отражение при задании значений домена, определяющего ученую степень в области истории и философии.

Онтология предметной области содержит следующие классы понятий, построенные на основе понятий Онтологии научного знания:

Раздел науки Археология. Этот класс отражает иерархию направлений научной деятельности. В частности, он может извлекаться из системной классификации науки, образуя свою иерархию наследования. К разделам археологии относятся, например, такие направления как *Искусствоведческая, Вещеведческая, Технологическая археология*, и др. Эти общие направления также подразделяются на более частные, например, *Полевая искусствоведческая, Реконструктивная искусствоведческая, Полевая вещеведческая, Реконструктивная вещеведческая археология*.

Метод исследования Археологии. Данный класс служит для описания методов исследования, применяемых в археологии к определенному типу археологических объектов. В данном классе были выделены подклассы *Подходы, Принципы, Методики, Технологии*, а также группа методов отдельных наук, используемых в археологии, образующая соответственно подклассы *Биологические методы, Физические методы и Химические методы*.

Объекты исследования Археологии. Данный класс понятий определяет объекты исследования в археологии, атрибутами которых являются количество и описание состава. Данный класс содержит подкласс *Нематериальных объектов*, разделяющийся в свою очередь на подклассы *Историческая группа людей, Историческая личность*,

Историческое место, и подкласс *Материальных объектов*, наследниками которого являются подклассы *Артефакт*, *Комплекс*, *Памятник*.

Научный результат Археологии. Этот класс служит для описания результатов научной деятельности археологии, таких как открытия, новые законы, теории, исторические факты. Свойства понятий этого класса представлены атрибутами: описание результата, дата получения и тип результата. Данный класс понятий разделяется на следующие подклассы: *Археологическая гипотеза*, *Археологический факт*, *Научное достижение*, *Научное открытие*.

Археологические периоды. Данный класс служит для датирования объектов исследования. Периоды образуют иерархию вложенности и исторического следования и задаются временным интервалом.

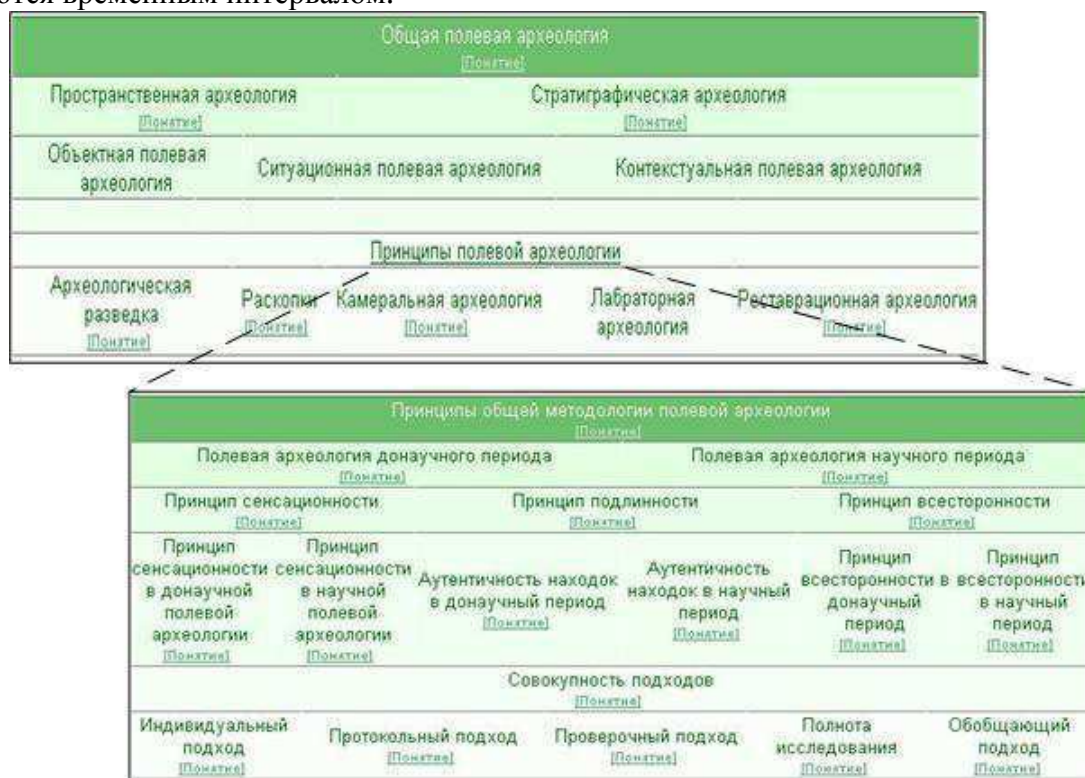


Рис.68. Фрагмент системной классификации археологической науки, предложенной Ю.П. Холюшкиным

Иерархии классов понятий ПО связаны между собой и с понятиями базовых онтологий посредством ассоциативных отношений, семантика которых определяется при задании отношений базовых онтологий.

Так, например, иерархия методов исследования, применимых к определенному типу объектов исследования, связана с иерархией объектов посредством отношения "применяется к классу объектов". Данное отношение отражает свойственное археологическим методам обобщение, когда, набор конкретных Методов исследования может применяться к определенному классу объектов исследования или разделов науки, например, "Метод датирования" применяется к классу *Памятники*.

Иерархия научных результатов, служащая для типизации и описания результатов научной деятельности, связана с деятельностью с помощью отношения "результат деятельности".

Иерархия объектов исследования связана с публикациями, в которых описывается объект, посредством отношения "описывает объект". Понятия объектов исследования связаны между собой с помощью отношения "подобен", задающего степень схожести (подобия) Объектов между собой, и отношения "вещественно подтверждается", задающего связь нематериального археологического объекта исследования с материальным объектом исследования, который подтверждает его существование.

Связь иерархии разделов науки с используемыми и рассматриваемыми в них методами исследования организована соответственно с помощью отношений "раздел использует метод" и "рассматривает метод". А использование отношений "раздел изучает период" и "раздел локализован в" позволяет установить хронологическое и географическое положение раздела науки.

2.12. Сибирский книжный археологический портал.

Последними разработками в ИАЭТ СО РАН, размещенных в вып. 17 ИТГИ были "Сибирский книжный археологический портал" и сайт серийного издания "Информационные технологии в гуманитарных исследованиях" [Холюшкин, Рыбинцев, 2012: с. 119-120].

В ходе работ 2012 года завершены работы над "Сибирским книжным археологическим порталом". С точки зрения пользователя, портал является тематическим Интернет-ресурсом, обеспечивающим возможность поиска и просмотра информации в рамках заданной предметной области (археологии, этнографии и антропологии). Портал реализован на CMS (системе управления содержимым сайта) WordPress с открытым исходным кодом (используемой по лицензии GNU GPL), написанной на PHP и использующей в качестве базы данных MySQL (Рис.69).

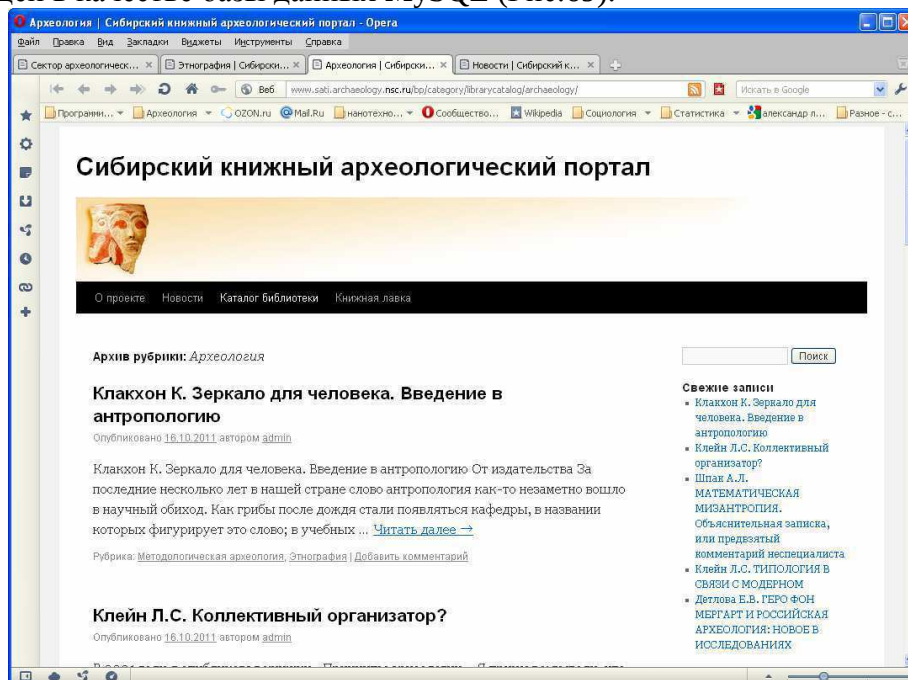


Рис. 69. "Сибирский книжный археологический портал".

За основу дизайна портала была взята тема Twenty Ten, которая была адаптирована под потребности проекта данного проекта. Проект рассчитан на заполнение контента группой людей, которые могут добавлять в существующую структуру новые записи в соответствии со своими правами. Каталог библиотеки включает разделы: археология, естественные науки, этнография, антропология. Описание предметной области основывается на системной классификации археологической науки, предложенной Ю.П. Холюшкиным и Е.Д. Гражданниковым и развиваемой в настоящее время. Эта классификация включает более 500 рубрик, представленных в виде иерархии.

Поддерживается разнообразный контент: текстовые файлы, PDF, документы MS Word, большое количество форматов изображений, видео и аудио файлы.

Права по изменению структуры могут быть делегированы нескольким пользователям. Также имеется возможность модерации новых записей, в настоящий момент эта функция не используется. Имеется поддержка канала Atom семейства RSS⁵.

⁵ RSS — семейство XML-форматов, предназначенных для описания лент новостей, анонсов статей, изменений в блогах и т. п. Информация из различных источников, представленная в формате RSS, может быть собрана, обработана и представлена пользователю в удобном для него виде специальными программами-агрегаторами. Протокол публикации

Предусматривалась реализация книжной продукции как в безвозмездном виде, так и путем продажи изданий в разделе книжная лавка. Реализован поиск статей по описаниям, как по всему portalу, так и по различным разделам. Предусмотрено было введение системы регистрации пользователей Сибирским книжным археологическим порталом.

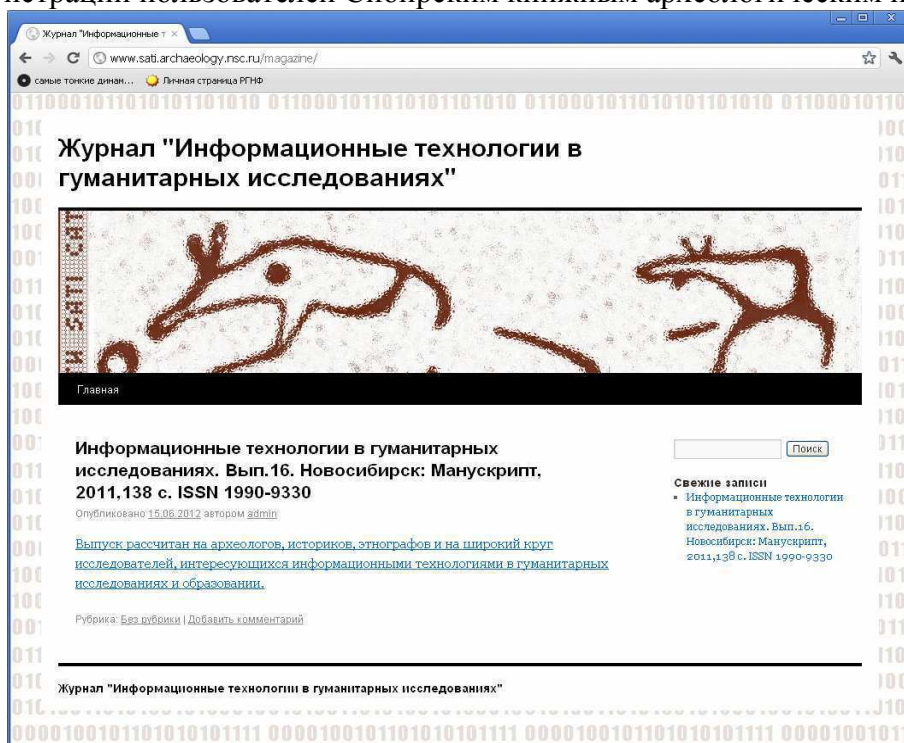


Рис. 70. Журнал "Информационные технологии в гуманитарных исследованиях"

В ходе реализации проекта в заключительном 2012 году был также создан сайт журнала "Информационные технологии в гуманитарных исследованиях ISSN 1990-9330" <http://www.sati.archaeology.nsc.ru/magazine/> (в настоящее время сайт не существует).

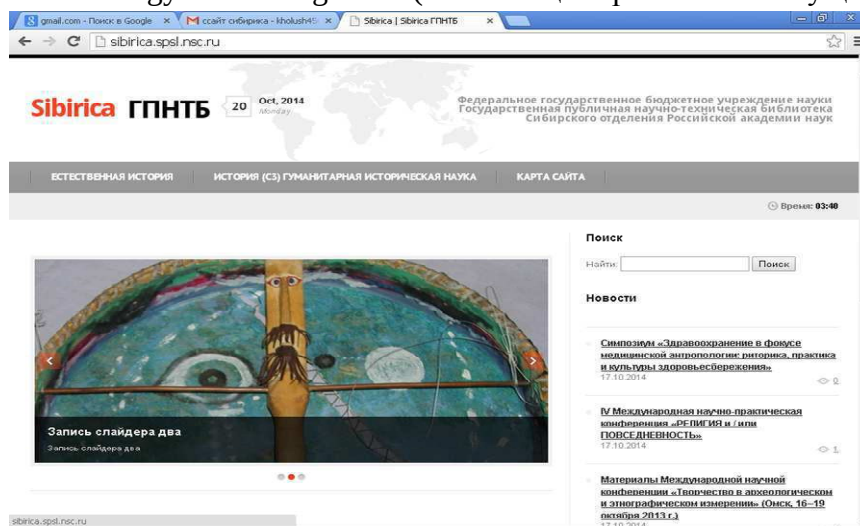


Рис.71. Портал "Sibirica"

Появилась возможность создавать статические страницы, например «о сайте», «обратная связь», а также мощный WYSIWYG текстовый редактор;

Предусматривается разместить на сайтах: онлайн-редактор фотографий, поддерживающий функции «обрезать», «отразить», «повернуть» и «масштабировать». Не Photoshop, конечно, но вещь порою очень полезная.

Atom (также AtomPub, от англ. Atom Publishing Protocol) основан на HTTP и позволяет создавать, изменять и удалять ресурсы, собранные в коллекции на веб-сайте (примером коллекции может служить блог). Содержимое коллекций описывается в формате Atom, а для управления им используются стандартные методы HTTP.

The screenshot shows a web browser window with the address bar displaying 'stibrica.spbl.nsc.ru/sitemap/'. The page content is a site map for 'stibrica' and is organized into a hierarchical tree structure. The root node is 'Стибрия', which branches into 'Естественные науки' and 'История'. 'Естественные науки' further branches into 'Естественно-историческая методология', 'Историческая география', 'Наука о космогенезе', 'Наука о мезоантропогенезе', 'Наука о палеоантропогенезе', 'Наука о филогенезе', 'Наука об антропогенезе', 'Общая естественная история', 'Палеоантропогенез', 'Палеогеография', 'Палеоэкология', 'Палеоэволюция', 'Палеофизиология', 'Палеопсихология', 'Палеосociология', 'Палеополитология', 'Палеоэкономика', 'Палеокультурология', 'Палеолингвистика', 'Палеоистория', 'Палеогеография', 'Палеоэкология', 'Палеоэволюция', 'Палеофизиология', 'Палеопсихология', 'Палеосociология', 'Палеополитология', 'Палеоэкономика', 'Палеокультурология', 'Палеолингвистика', 'Палеоистория'. 'История' branches into 'Историческая география', 'Историческая антропология', 'Историческая демография', 'Историческая экономика', 'Историческая социология', 'Историческая психология', 'Историческая философия', 'Историческая религия', 'Историческая культура', 'Историческая наука', 'Историческая литература', 'Историческое искусство', 'Историческое ремесло', 'Историческое производство', 'Историческое потребление', 'Историческое распределение', 'Историческое накопление', 'Историческое использование', 'Историческое управление', 'Историческое регулирование', 'Историческое ограничение', 'Историческое расширение', 'Историческое развитие', 'Историческое изменение', 'Историческое сохранение', 'Историческое восстановление', 'Историческое обновление', 'Историческое совершенствование', 'Историческое развитие', 'Историческое изменение', 'Историческое сохранение', 'Историческое восстановление', 'Историческое обновление', 'Историческое совершенствование'. The browser's address bar shows the URL 'stibrica.spbl.nsc.ru/sitemap/'. The browser's title bar shows 'stibrica'. The browser's status bar shows 'stibrica.spbl.nsc.ru/sitemap/'.

- Стибрия
 - Естественные науки
 - Естественно-историческая методология
 - Историческая география
 - Наука о космогенезе
 - Наука о мезоантропогенезе
 - Наука о палеоантропогенезе
 - Наука о филогенезе
 - Наука об антропогенезе
 - Общая естественная история
 - Палеоантропогенез
 - Палеогеография
 - Палеоэкология
 - Палеоэволюция
 - Палеофизиология
 - Палеопсихология
 - Палеосociология
 - Палеополитология
 - Палеоэкономика
 - Палеокультурология
 - Палеолингвистика
 - Палеоистория
 - История
 - Историческая география
 - Историческая антропология
 - Историческая демография
 - Историческая экономика
 - Историческая социология
 - Историческая психология
 - Историческая философия
 - Историческая религия
 - Историческая культура
 - Историческая наука
 - Историческая литература
 - Историческое искусство
 - Историческое ремесло
 - Историческое производство
 - Историческое потребление
 - Историческое распределение
 - Историческое накопление
 - Историческое использование
 - Историческое управление
 - Историческое регулирование
 - Историческое ограничение
 - Историческое расширение
 - Историческое развитие
 - Историческое изменение
 - Историческое сохранение
 - Историческое восстановление
 - Историческое обновление
 - Историческое совершенствование

Портал также реализован на CMS "WordPress", написанном на PHP и использующий в качестве БД MySQL. Встроенная система тем и плагинов в наибольшей степени подходит для реализации данного проекта и позволяет конструировать практически любые проекты, поддерживая разнообразный контент: текстовые файлы, PDF, документы Word, большое количество форматов изображения, видео и аудио файлы.

3.1. Вебометрия электронного ресурса ИТГИ (2013–2018 гг.)

В конце октября 2013 г. на библиотечном сайте Отделения ГПНТБ СО РАН были выставлены полнотекстовые электронные версии первых 17 выпусков нерегулярного продолжающегося издания «Информационные технологии в гуманитарных исследованиях» (ИТГИ). Этот сборник издаётся с 1998 года под редакцией академика РАН, д.и.н. Ю.П. Холушкина (URL: prometeus.nsc.ru/elibrary/infohum/). Последующее представление выпусков №№ 18–22 происходило по мере выхода их из печати и, таким образом, к началу 2019 г. на сайте представлены электронные версии всех выпусков ИТГИ за два десятилетия.

Таблица 16. Состав ресурса «Информационные технологии в гуманитарных исследованиях» на сайте prometeus.nsc.ru на 31 декабря 2018 года.

Название страницы	Веб-адрес	Дата выставления на сайте	Размер
индекс	/elibrary/infohum/	2013-10-31	24,9 к
вып. 01	1998-001 (ssi / pdf)	2013-10-31 / 2015-04-02	10,1 Кб / 25,1 Мб
вып. 02	2000-002 (ssi / pdf)	2013-10-31 / 2015-04-02	8,8 Кб / 26,6 Мб
вып. 03	2002-003 (ssi / pdf)	2013-10-31 / 2015-04-02	7,9 Кб / 22,4 Мб
вып. 04	2002-004 (ssi / pdf)	2013-10-31 / 2015-04-02	7,4 Кб / 15,1 Мб
вып. 05	2003-005 (ssi / pdf)	2013-10-31 / 2015-04-02	8,9 Кб / 24,6 Мб
вып. 06	2003-006 (ssi / pdf)	2013-10-31 / 2015-04-02	9,95 Кб / 24,4 Мб
вып. 07	2004-007 (ssi / pdf)	2013-10-31 / 2015-04-02	9,6 Кб / 28,7 Мб
вып. 08	2004-008 (ssi / pdf)	2013-10-31 / 2015-04-02	8,5 Кб / 16,9 Мб
вып. 09	2005-009 (ssi / pdf)	2013-10-31 / 2015-04-02	9,0 Кб / 24,1 Мб
вып. 10	2006-010 (ssi / pdf)	2013-10-31 / 2015-04-02	9,1 Кб / 30,7 Мб
вып. 11	2006-011 (ssi / pdf)	2013-10-31 / 2015-04-02	8,4 Кб / 28,8 Мб
вып. 12	2008-012 (ssi / pdf)	2013-10-31 / 2015-04-02	10,5 Кб / 28,1 Мб
вып. 13	2009-013 (ssi / pdf)	2013-10-31 / 2015-04-02	8,95 Кб / 34,2 Мб
вып. 14	2009-014 (ssi / pdf)	2013-10-31 / 2013-10-31	9,6 Кб / 2,2 Мб
вып. 15	2010-015 (ssi / pdf)	2013-10-31 / 2013-10-31	8,9 Кб / 24,5 Мб
вып. 16	2011-016 (ssi / pdf)	2013-10-31 / 2013-10-31	11,3 Кб / 7,3 Мб
вып. 17	2012-017 (ssi / pdf)	2013-10-31 / 2013-10-31	9,4 Кб / 4,7 Мб
вып. 18	2013-018 (ssi / pdf)	2013-11-15 / 2013-11-22	8,9 Кб / 3,8 Мб
вып. 19	2014-019 (ssi / pdf)	2014-09-30 / 2014-10-06	10,7 Кб / 4,0 Мб
вып. 20	2014-020 (ssi / pdf)	2014-11-13 / 2016-02-09	10,4 Кб / 7,5 Мб
вып. 21	2015-021 (ssi / pdf)	2015-12-11 / 2015-12-29	9,6 Кб / 5,2 Мб
вып. 22	2017-022 (ssi / pdf)	2017-12-29 / 2017-12-29	13,9 Кб / 10,3 Мб
Итого (по всем выпускам 1–22):			234,7 Кб / 399,2 Мб

Каждая страница в формате *html*, посвящённая отдельному выпуску, включает в себя изображение обложки (*gif*), полное библиографическое описание, подробное оглавление с указанием пагинаций и ссылку на полный текст (*pdf*). Для точности понимания отметим, что при дальнейшем упоминании указание на *html* (на сайте эти страницы имеют расширение *.ssi*) речь, как правило, идёт об *оглавлениях* выпусков, а сканированные полнотекстовые выпуски представлены файлами *pdf*. На 31.12.2018 г. общий размер ресурса составляет 80 файлов (24 *html* + 22 *pdf* + 33 *gif* + 1 *jpg*) объёмом в 400 мегабайт (Табл. 16).

По традиции материалы сборника включают в себя широкий круг вопросов, связанных с теорией и методологией науки, гуманитарной информатикой, разными аспектами использования информационных ресурсов в библиотечном, архивном и музейном деле, поднимают проблемы применения технологий знаний и «data Mining», компьютерного моделирования в различных гуманитарных отраслях, разработки программных продуктов (в том числе ГИС-приложений), вопросы наукометрии, вебометрии и пр. Политематические материалы сборника привлекают к себе внимание не только антропологов, археологов, историков, сотрудников библиотек и музеев, но и математиков, социологов, лингвистов, науковедов. Тематическое распределение публикаций, цитирующих ИТГИ, изученное на материалах платформы eLIBRARY,

показывает, что сборник активно цитируется в таких отраслях, как информатика, геодезия, картография, кибернетика, культурология и даже медицина⁶. Аналогичное распределение по ключевым словам подтверждает тематическое многообразие ИТГИ, так как цитирующие публикации описаны 496 терминами⁷.

Анализ поисковых запросов, обсуждений в соцсетях и ссылок в интернете указывает, что к целевой аудитории сборника примыкают также студенты и аспиранты, занимающиеся проблематикой научных классификаций и систематизацией в гуманитарных отраслях. Многочисленные ссылки на материалы ИТГИ зафиксированы на сайте Google Scholar, на зарубежных платформах Springer Link, ACM Digital Library⁸, в диссертациях и материалах различных конференций (IEEE и др.). На платформе Научной электронной библиотеки eLIBRARY присутствует ограниченное число материалов сборника – только 43 статьи (то есть 17% общего количества) из 18 разных выпусков. Тем не менее, здесь присутствует 139 ссылок на ИТГИ, включая 16 высокоцитируемых публикаций, имеющих от 10 до 32 ссылок.

Общее количество обращений к ресурсу устанавливается по данным статистической системы AWStats, обрабатывающей лог-файл библиотечного сервера. Согласно этим данным, за пять лет с небольшим (1887 суток) к электронной версии ИТГИ сделано около 15,5 тыс. доступов, то есть в среднем 8,18 доступов в сутки (Табл. 17).

Таблица 17. Посещаемость электронной версии ИТГИ за период с 1.11.2013 по 31.12.2018 г.
(по данным статистической системы AWStats)⁹.

Название страницы	Веб-адрес	Количество обращений
индекс	/elibrary/infohum/	7238
вып. 01	1998-001 (ssi)	363
вып. 02	2000-002 (ssi)	262
вып. 03	2002-003 (ssi)	242
вып. 04	2002-004 (ssi)	248
вып. 05	2003-005 (ssi)	240
вып. 06	2003-006 (ssi)	233
вып. 07	2004-007 (ssi)	280
вып. 08	2004-008 (ssi)	233
вып. 09	2005-009 (ssi)	245
вып. 10	2006-010 (ssi)	288
вып. 11	2006-011 (ssi)	256
вып. 12	2008-012 (ssi)	252
вып. 13	2009-013 (ssi)	266
вып. 14	2009-014 (ssi / pdf)	526
вып. 15	2010-015 (ssi / pdf)	407
вып. 16	2011-016 (ssi / pdf)	1174
вып. 17	2012-017 (ssi / pdf)	764
вып. 18	2013-018 (ssi / pdf)	878
вып. 19	2014-019 (ssi / pdf)	431
вып. 20	2014-020 (ssi / pdf)	307
вып. 21	2015-021 (ssi / pdf)	202
вып. 22	2017-022 (ssi)	101
Итого (за период с 1.11.2013 – 31.12.2018):		15436

AWStats имеет ту особенность, что, извлекая сведения из лог-файла, иначе говоря, из журнала доступов к серверу, эта система не умеет очищать их от спама. Поэтому можно предположить, что некоторые метрики, учтённые в Таблице 17, отражают «пустопорожный» трафик спам-ботов. С другой стороны, показатели скачивания более тяжёлых pdf-файлов (полных версий сборника) в гораздо меньшей степени зависят от

⁶ URL: https://elibrary.ru/title_profile_new_cit_rubrics.asp?id=38388&rand=0.566489747140918.

⁷ URL: https://elibrary.ru/title_profile_new_cit_keywords.asp?id=38388&rand=0.7004474044360953.

⁸ Marchuk, A.G., Kholushkin, Yu.P., Zagorul'ko, Yu.A., et al., Intellectual internet portal of knowledge for access to information resources on archeology and ethnography, Inf. Tekhnol. Guman. Issled., 2004, vol. 8, pp. 7–13. <http://www.prometeus.nsc.ru/elibrary/infohum/2004-008.pdf>

⁹ Особенности различных версий системы AWStats, в разные годы действовавших на сайте prometeus.nsc.ru, дали возможность обработать данные по скачиванию pdf только за 2015–2018 гг. (они представлены в Табл. 18). Новейшие версии софта AWStats выделяют эти метрики в отдельное меню «Downloads». Вместе с тем, часть данных по скачиванию pdf (выпуски 14–21) отразилась и в общих отчётах AWStats и присутствует здесь в Таблице 17.

загрязнения статистики. Видно, что они значительно выше, чем при обращении к *html*-страницам (Табл. 18).

Таблица 18. Загрузка полных версий сборника ИТГИ за 2015–2018 гг.
(по данным статистической системы AWStats).

Страница	2015	2016	2017	2018	Всего
1998-001.pdf	87	123	131	165	506
2000-002.pdf	72	96	136	80	384
2002-003.pdf	75	156	195	138	564
2002-004.pdf	108	137	185	132	562
2003-005.pdf	83	148	207	144	582
2003-006.pdf	87	111	110	87	395
2004-007.pdf	79	99	144	172	494
2004-008.pdf	76	175	211	164	626
2005-009.pdf	84	168	226	152	630
2006-010.pdf	91	119	218	170	598
2006-011.pdf	80	89	116	89	374
2008-012.pdf	88	177	179	81	525
2009-013.pdf	106	146	205	145	602
2009-014.pdf	293	328	517	509	1647
2010-015.pdf	97	188	281	227	793
2011-016.pdf	260	299	272	412	1243
2012-017.pdf	250	273	456	550	1529
2013-018.pdf	246	262	365	441	1314
2014-019.pdf	490	648	854	892	2884
2014-020.pdf	189	243	415	2982	3829
2015-021.pdf	38	598	804	1346	2786
2017-022.pdf	–	–	5	3268	3273
Вып. 1–22:	2979	4583	6232	12346	26140

Даже самый поверхностный взгляд на показатели скачивания последних 5–7 лет демонстрирует значительное преимущество полнотекстовых файлов над чисто описательными (библиографическими) страницами. Не играет роли даже то, что *html*-оглавления намного лучше учтены поисковыми машинами и представляют из себя содержательный набор ключевых слов (*pdf*-тексты даже не распознаны). Тем не менее, следует учитывать, что именно проиндексированные оглавления направляют пользователей на скачивание того или иного выпуска.

Чаще всего скачивались новые выпуски ИТГИ №№ 19–22, которые за четыре года суммарно загружались около 12,8 тыс. раз. Почти треть из этого количества составили загрузки монографического выпуска № 20, посвящённого описанию веб-системы создания, использования и применения стратегий решения задач. По данным Google Analytics, процент выхода с сайта после просмотра оглавления этого выпуска (2014-020.ssi) – один из самых низких (13,95%), что означает, что после просмотра пользователи, как правило, обращаются к скачиванию полного текста (2014-020.pdf).

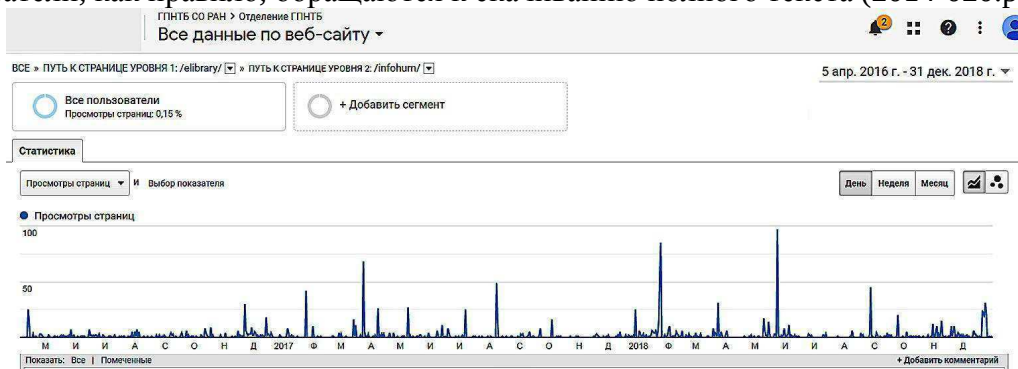


Рис. 73. Диаграмма количества просмотров ресурса ИТГИ за период 5.04.2016 – 31.12.2018 (по данным Google Analytics).

На противоположном конце присутствует выпуск № 11, замыкающий «рейтинг загрузок». Он скачивался в десять раз хуже лидера, но с точки зрения практики доставки знаний (веб-логистики), этот факт ни о чём не свидетельствует и, тем более, ни о каком содержательном качестве этого выпуска. Изучение электронных библиотек показывает,

что есть множество ресурсов, которые до поры до времени латентны, а затем проявляют себя, порою весьма ярко.

Качественную сторону обращений с точки зрения веб-логистики раскрывают данные Google Analytics (GA). Относительно достоверные цифры по ИТГИ имеются за период с 5 апреля 2016 г. по 31 декабря 2018 г. (это ровно 1000 суток). При этом нельзя утверждать, что система полностью свободна от загрязнений ботами. На диаграмме GA по числу просмотров за указанный период отчётливо видны всплески, обозначающие их активность (Рис. 73). Этим «пикам» соответствуют значения «просмотров» примерно от 25 до 100 страниц.

Тем не менее, можно опираться на факт, что веб-аналитика GA носит «относительно чистый» характер. В среднем в её отчёты попадает до трети данных AWStats. Так, по 2017-му году статистика AWStats показывает 1642 обращения к ИТГИ, а данные Google Analytics подтверждают 534 «просмотра» (32,5%). По 2018-му году – соответственно 2889 и 909 (31,5%). С качественной же стороны метрики GA рисуют намного более подробную картину использования ресурса, сообщая не только «очищенные» сведения об общем количестве всех просмотров и уникальных просмотров страниц, но и о средней длительности присутствия на страницах, показателях отказов (*bounce rate*) и доле завершённых сессий после просмотра конкретных страниц (Табл. 19). Сопоставление ключевых метрик (*key performance indicators*) между собой позволяет интерпретировать особенности внешнего трафика ресурса в целом и функционирования отдельных документов в веб-среде.

Таблица 19. Посещаемость электронной версии ИТГИ за период с 5.04.2016 по 31.12.2018 г.
(по данным статистической системы Google Analytics).

Страница	Просмотры	Уникальные просмотры	Средняя длительность просмотра	Показатель отказов	Процент выходов
/elibrary/infohum/	1089	608	00:01:21	74,71%	41,30%
2015-021.ssi	82	56	00:01:36	90,00%	26,83%
2017-022.ssi	57	39	00:03:29	54,55%	42,11%
2014-019.ssi	46	37	00:02:46	50,00%	17,39%
2006-010.ssi	44	31	00:01:36	58,33%	29,55%
2014-020.ssi	43	34	00:01:25	60,00%	13,95%
1998-001.ssi	42	38	00:01:43	66,67%	26,19%
2011-016.ssi	37	31	00:02:23	90,00%	37,84%
2004-007.ssi	32	27	00:01:49	16,67%	12,50%
2012-017.ssi	31	25	00:01:13	33,33%	12,90%
2013-018.ssi	29	22	00:00:51	100,00%	24,14%
2005-009.ssi	28	21	00:01:27	66,67%	25,00%
2008-012.ssi	28	21	00:00:54	0,00%	14,29%
2000-002.ssi	27	23	00:02:22	0,00%	25,93%
2009-013.ssi	25	22	00:00:43	100,00%	24,00%
2003-006.ssi	24	21	00:02:32	83,33%	37,50%
2006-011.ssi	24	19	00:01:30	50,00%	12,50%
2003-005.ssi	23	18	00:00:42	50,00%	13,04%
2010-015.ssi	23	21	00:01:01	100,00%	8,70%
2002-004.ssi	20	16	00:01:10	66,67%	15,00%
2002-003.ssi	18	17	00:00:37	0,00%	5,56%
2004-008.ssi	17	13	00:02:41	0,00%	0,00%
2009-014.ssi	16	13	00:00:16	0,00%	0,00%
Итого	1805	1173	00:01:26	73,05%	32,74%

В этой Таблице № 19 самыми любопытными представляются данные по отказам. Считается, что чем меньше уровень *bounce rate* конкретного веб-документа, тем он более привлекателен для пользователя. С просмотренной страницы посетитель, как правило, переходит к следующему документу, увеличивая долю своего присутствия на сайте. Правда, нужно помнить, что речь идёт о страницах описаний и оглавлений, после просмотра которых посетитель, как правило, заинтересован в полных текстах. С этой точки зрения самыми притягательными для пользователей, – если не брать во внимание не вполне понятные нули, выставленные Google Analytics выпускам №№ 2, 3, 8, 12, 14, – являются выпуски ИТГИ №№ 7, 17, 5, 11 и 19. Уровень отказов у этих html-страниц

намного ниже среднего показателя по сайту (79,18%). А самые сбалансированные метрики, что не удивительно, имеет самый последний выпуск № 22, где относительно приемлемый уровень отказов 54,55% сочетается с самым высоким показателем пребывания на странице – 3 мин. 29 сек. При этом среднее время пребывания по всему сайту составляет 1:54. Добавим, что по количеству скачиваний pdf выпуск № 22 находится на втором месте (см. Табл. 18), а по популярности («раскрутке») в Сети – на третьем, о чём свидетельствует метрика «входов» (Табл. 20).

Таблица 20. Страницы входа на сайт prometeus.nsc.ru за период с 5.04.2016 по 31.12.2018 г.
(по данным статистической системы Google Analytics).

Страница входа	Источники трафика			Действия		
	Сеансы	Новые сеансы	Новые пользователи	Показатель отказов	Страниц за сеанс	Средняя длительность сеанса
/elibrary/infohum/	431	66,13 %	285	74,71 %	2,33	00:01:58
2006-010.ssi	12	75,00 %	9	58,33 %	6,25	00:10:08
2017-022.ssi	11	72,73 %	8	54,55 %	2,73	00:05:18
2011-016.ssi	10	40,00 %	4	90,00 %	1,30	00:00:24
2015-021.ssi	10	60,00 %	6	90,00 %	4,80	00:01:21
index.ssi	11	20,00 %	3	80,00 %	1,20	00:00:04
1998-001.ssi	6	50,00 %	3	66,67 %	1,67	00:00:26
2003-006.ssi	6	33,33 %	2	83,33 %	1,33	00:00:02
2004-007.ssi	6	16,67 %	1	16,67 %	7,00	00:09:55
2014-020.ssi	5	100,00 %	5	60,00 %	1,60	00:01:04
2009-013.ssi	4	25,00 %	1	100,00 %	1,00	00:00:00
2013-018.ssi	4	50,00 %	2	100,00 %	1,00	00:00:00
2014-019.ssi	4	50,00 %	2	50,00 %	1,50	00:00:14
2002-004.ssi	3	33,33 %	1	66,67 %	1,67	00:00:21
2005-009.ssi	3	66,67 %	2	66,67 %	1,33	00:01:34
2012-017.ssi	3	66,67 %	2	33,33 %	2,33	00:02:22
2000-002.ssi	2	50,00 %	1	0,00 %	2,00	00:00:35
2003-005.ssi	2	0,00 %	0	50,00 %	2,00	00:00:04
2006-011.ssi	2	0,00 %	0	50,00 %	3,50	00:01:14
2009-014.ssi	2	0,00 %	0	0,00 %	6,00	00:09:17
2010-015.ssi	2	0,00 %	0	100,00 %	1,00	00:00:00
2005-009.ssi	2	100,00 %	2	50,00 %	2,00	00:00:19
Итого	541	62,66 %	339	73,01 %	2,43	00:02:07

Таким образом, из общего количества в 1805 просмотров ресурса, которые Google Analytics «насчитывает» для ИТГИ (см. Табл. 19) на долю прямых заходов приходится 541 (см. Табл. 20) или почти 30%. Другими словами, почти треть посещений сборника пользователи начинают с прямых заходов на сайт именно через ИТГИ. Конечно, источники этого трафика различны. Это может быть и органический поиск (поисковые выдачи Google, Яндекс и других поисковиков), и внешние ссылки на других сайтах, и упоминания в социальных сетях. Более подробное изучение данного вопроса носит трудоёмкий характер и, вероятно, станет предметом дальнейших исследований.

3.2. Основная проблематика изучения библиотечного веб-пространства. Вебметрия и веб-аналитика (обзор литературы)

Чёткое определение того, что такое «веб-пространство», до сих пор не существует. Несмотря на интенсивно растущее использование данного понятия, его нет ни в «Википедии», ни в современных словарях новой лексики. Как правило, термин применяют в синонимическом ряду со словами «веб», «всемирная паутина» и «интернет». Но с перечисленными понятиями обычно не используют атрибуты профессионализации или сегментирования, тогда как о *профессиональном веб-пространстве* (бизнес-коммуникативном, научно-образовательном или библиотечном) говорят довольно часто. В данном обзоре «библиотечное веб-пространство» используется в расширительном смысле. Под ним понимается не просто сообщество библиотечных сайтов, связанных гиперссылками и навигацией, но и совместная деятельность библиотек в интернете, а также единство принципов и подходов к этой деятельности.

Беглый поиск по НЭБу (научной электронной библиотеке eLIBRARY.Ru) даёт не менее 700 документов по разным отраслям знания, в которых так или иначе говорится о веб-пространстве. Сам термин, судя по всему, возник на рубеже XX-XXI вв., когда интернет-паутина уже уверенно опутала всю планету. В российских средствах массовой информации, а также в популярной и научно-технической литературе активное использование понятия «веб-пространство» началось когда завершился начальный этап массового сайтостроения, т. е. примерно с 2008 г. (Рис. 74).

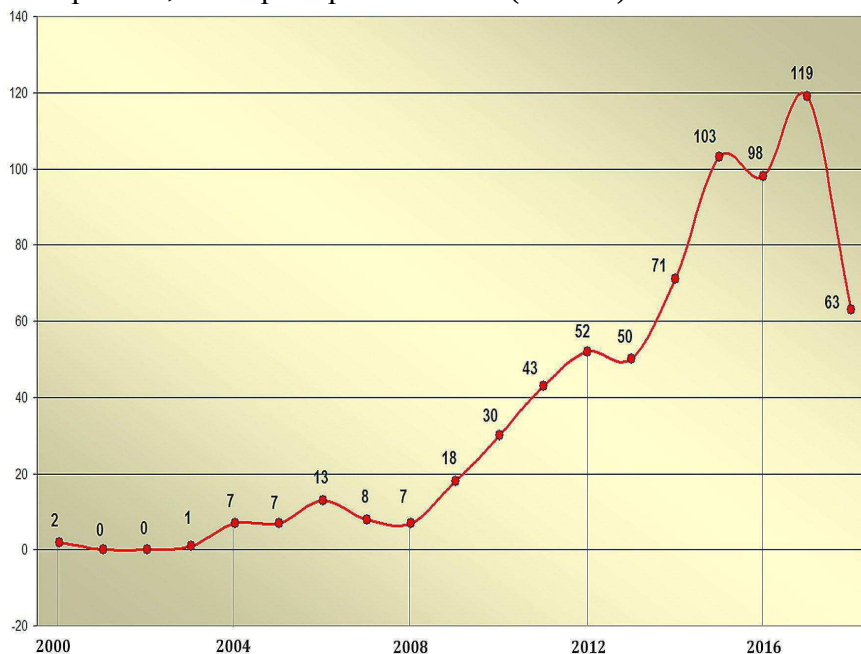


Рис.74. Количество публикаций 2000-2018 гг. в базе данных eLIBRARY.Ru, затрагивающих разные аспекты «веб-пространства». В диаграмму включено 692 документа, выявленных при поиске по заглавиям и ключевым словам.

Отечественное библиотечное веб-пространство имеет ряд специфических черт, выделяющих его из общего массива глобальной сети. В первую очередь, это один из самых почтенных, консервативных и устойчивых сегментов Рунета, обязанный своим положением тому, что интернет пришёл в библиотеки одним из первых. При вечном дефиците средств это отразилось и на библиотечных технологиях, которые в значительной мере остаются архаичными. Громоздкие и негибкие системы поиска библиотечной информации (каталоги), слабая автоматизация библиотечных процессов, практическое отсутствие вебометрики и веб-аналитических служб для оценки «обратной связи» – всё это следствие разного рода многолетних дефицитов – ресурсов, квалифицированных кадров и эффективных управленческих решений.

Вместе с тем, огромный вклад библиотек заключается в их следовании принципам открытости, бесплатности и свободы доступа к культурным ценностям и качественной научно-образовательной информации, на чём базируется всемирная концепция открытого доступа (*open access*). Задача библиотек – предоставить свободное и бесплатное право знакомиться, читать, загружать, копировать, распространять, распечатывать, искать и ссылаться на любые библиотечно-библиографические материалы, включая полнотекстовые. В условиях *www* традиционные ценности приобретают особое значение.

Во второй половине 1990-х и начале 2000-х годов библиотечное веб-пространство формировалось как часть общей коммуникативной среды науки и образования. В процессе становления принципиально новой инфраструктуры знания активно развивались и новые возможности доступа к информации, такие как электронные библиотеки, электронные журналы и книги, информационные системы, электронные коллекции, фактографические, библиографические и полнотекстовые базы данных. Лидером по числу компьютеров и пользователей, по объёму передаваемых данных, разнообразию и качеству услуг была Сеть передачи данных Сибирского отделения РАН (СПД СО РАН).

Органичной составной частью в неё входили ресурсы научных библиотек и музеев [Информационно-телекоммуникационные..., 2003; Сеть передачи..., 2005].

Аналогичным образом к подобной интеграции локальных электронных архивов и ресурсов с целью создать распределённую электронную библиотеку с единой точкой входа в систему российских и мировых научных ресурсов приступило и Дальневосточное отделение РАН [Корпоративная сеть..., 2007: 19; Ханчук, 2008: 302]. Веб-структуры информационного пространства носили разрозненный характер, отличались технологической неоднородностью и минимальной связностью друг с другом [Ханчук, 2009: 124].

Актуальность количественного и качественного изучения этих процессов сложилась в связи со стремительным развитием интернета и появлением в сети гигантского количества ресурсов, созданных научно-образовательными, а затем и библиотечными учреждениями. Возникла и укрепилась новая отрасль знаний – *вебометрия* – как отдельный раздел измерений в информатике, изучающий систему индикаторов и веб-данных, исследующий феномены гиперссылочных связей и социальных сетей. Целые коллективы исследователей обратились к анализу взаимодействий между сайтами разных типов и учреждений – институтов, университетов и научных библиотек [Печников, 2011; Шокин, 2011; Шокин, 2012; Печников, 2013а; Печников, 2014б; Печников, 2015].

Постепенное формирование инфраструктуры знаний (сайтов, порталов, корпоративных систем) привело к появлению целого ряда профессиональных «веб-пространств» и развитию соответствующих направлений в научных исследованиях. Большой пласт литературы отражает «*веб-пространство науки*» [Антопольский, 2017; Антопольский, 2018; Гуськов, 2016; Кабакова, 2014; Клименко, 2016а; Клименко, 2016б; Косяков, 2016; Кудряшова, 2007; Наумова, 2012; Наумова, 2014; Печников, 2013в; Печников, 2014а]; «*веб-пространство образования*» [Антопольский, 2012; Антопольский, 2018; Олефир, 2010]; «*библиотечное веб-пространство*» [Горячев, 2017; Зверевич, 2012; Колкова, 2015; Курганская, 2017; Миронов, 2011; Потехина, 2011; Шрайберг, 2018].

К началу 2010-х годов, в тесной связи с появлением новых моделей библиотечного обслуживания, таких как «Библиотека 2.0», само понятие библиотечного веб-пространства обогащается многочисленными *социальными сервисами*. Обнаруживается, что новая сущность библиотек обусловлена, в первую очередь, не столько автоматизацией и новой технологической средой, сколько принципиально иными возможностями (мобильными, сетевыми), позволяющими реализовать миссию библиотек в совершенно новых условиях. В новой электронной среде традиционные библиотечные услуги приобретают новые качества – доступности для неограниченного количества пользователей и оперативности предоставления [Серова, 2009: 27].

В первое время специфика новых свойств доставки и распространения библиотечной информации, своего рода «веб-логистика», была далеко не очевидна. При мониторинге и оценке веб-ресурсов исследователи смешивали информационные качества контента с функционалом его доставки, что явным образом присутствует в работах 2000-х гг. Так, например, в статье о методических подходах к оценке качества и эффективности электронных библиотек [Исмагилова, 2010] отсутствие чёткой границы между транспортной компонентой ресурсов и информационной стороной контента проступает в ошибочном противопоставлении разных видов и типов веб-ресурсов. Излагая результаты изучения эффективности создания электронных ресурсов сайта Научной библиотеки Таврического университета, автор статьи пишет, что полнотекстовые документы имеют «приоритет использования» по сравнению с библиографией [Исмагилова, 2010: 64-65].

Между тем, и те и другие виды ресурсов одинаково важны. Например, на сайте Отделения ГПНТБ СО РАН *prometeus.nsc.ru* наблюдается другая картина и библиографический сегмент сайта здесь играет решающее значение. Понятно, что полнотекстовые материалы будет просто невозможно найти без соответствующего библиографического подкрепления и ссылок. Сайт с библиографической информацией

может оказаться чрезвычайно полезным и эффективным для потребителя. При этом чем он лучше будет организован, тем выше окажется у него и показатель отказов (*bounce rate*). На коммерческих сайтах, напротив, высокий уровень отказов считался бы большим минусом.

С другой стороны, никакие ссылки на библиотечную миссию, авторские права или многоязычность (что, по большому счёту, можно рассматривать как вторичные критерии) не заменят чёткой структуры, оптимального размещения материалов и интуитивно понятной навигации. При отсутствии этих условий даже сайт с полнотекстовой информацией будет испытывать устойчивые проблемы посещаемости.

Смещение понятий логистической стороны распространения информации и качественной стороны контента сайтов наблюдается у многих авторов. Так, например, И.Филиппов не видит понятийной разницы между «неудобным / непродуманным интерфейсом» (функционал, доставка информации) и «неудачным / скучным контентом» (содержимое). Оба эти фактора, по словам автора, заставят посетителей «уходить с сайта вне зависимости от занимаемой им позиции в поисковой системе» [Филиппов, 2012: 59, 61].

Вместе с тем, Филиппов не обозначает и чётких правил, указывая, что любой сайт «сугубо индивидуален». В доказательство он приводит результаты исследований по трекингу глаз посетителей сайтов, согласно которым большинство из них фиксировали взгляд на меню и новостях слева. «Мы почти уверены, – пишет Филиппов, – что если радикально перестроить структуру нескольких сайтов, которые тестируются с помощью системы eye-tracking, то результаты исследований тоже существенно изменятся. Как сказал один из ведущих дизайнеров студии Артемия Лебедева, «люди смотрят туда, где им что-то показывают, а вовсе не куда-то туда всегда». Действительно, – подтверждает автор, – если на сайте, где размещено минимальное количество элементов, меню расположить в центре снизу (при условии, что веб-страница не требует прокрутки), то посетители все равно будут с ним взаимодействовать» [Филиппов, 2012: 62].

И ведь правда, если дверь в дом расположить не как обычно, по канону, а по центру снизу, возле фундамента, и горизонтально, то её всё равно будут искать, а кто-то даже попробует туда лазить. Наверняка, недолго и нечасто. В итоге, признаем правоту автора, что все практикующие веб-аналитики делятся на три категории и последняя из них объединяет «тех, кто сознательно или несознательно дискредитирует веб-аналитику в силу дилетантского подхода к самому процессу анализа веб-сайтов» [Филиппов, 2012: 60].

Возвращаясь к проблематике библиотечного веб-пространства, отметим, что со временем она стала сильно расширяться, сообразно усложнению всей мировой паутины. От первых прорывов и достижений, связанных с новыми информационными возможностями, библиотеки стали переходить к решению традиционных задач культурной и социальной работы. По словам Е.Д.Жабко, «электронная библиотека представляется очередным этапом развития библиотеки, а значит, на неё должно проектироваться общепринятое представление о традиционной библиотеке как социальном институте» [Жабко, 2009: 18]. По словам автора, эффективность информационных систем и, в частности, электронных библиотек зависит от 4-х уровней доступа к ресурсам – библиографического, аналитического, полнотекстового и универсального, предоставляющего комплекс виртуальных услуг и социальных сервисов [Жабко, 2009: 19]. Последний тип может включать различные веб-службы справочного, библиографического и рекомендательного характера. Его, на наш взгляд, было бы правильнее называть «сервисным» или «социальным».

Развивая тему новых возможностей ЭБ и инструментария для повышения социальной активности пользователей, Е.Д.Жабко называет электронную веб-среду «пространством общения» [Жабко, 2010: 105]. При этом она ссылается на Джека М. Манесса, писавшего о «социально насыщенном электронном пространстве» [Maness, 2006]. Манесс заметил, что

эгалитарная веб-среда уравнила пользователей и библиотекарей, позволяя всем быть проводниками единого информационно-библиотечного процесса.

«Интернет, – по словам Манесса, – очевидно становится всё более интерактивным и мультимедиа ведомым технологическим пространством» [Maness, 2006]. С помощью вики-технологий, персональных страниц, блогов и форумов все участники сетевого обмена, а не только библиотекари генерируют глобальный контент. «В некотором роде, – пишет Манесс, – для библиотек это и виртуальная реальность, и их реализация в сети как определённого места действия» [Maness, 2006]. Таким образом, новые технологии превращают всё пространство Интернета в общемировое пространство знаний.

Вместе с тем, известная часть библиотековедов осторожно пишет о необходимости более взвешенного подхода к самой концепции «библиотечного веб-пространства». «Библиотека 2.0, – отмечает Н.А. Миронов, – всего лишь концепция веб-интерфейса, позволяющего получать доступ к абстрактному массиву «электронных» документов в Сети, что библиотекой может являться только в самом примитивном понимании. Собственно, термин «библиотека» и применяется в определённых рамках – как обозначение, например, «класса объектов». Наверное, корректнее всего сказать, что «Библиотека» – это целая отрасль деятельности, делающей возможным использование современных средств и технологий» [Миронов, 2011: 33].

Автор подчёркивает, что эффективность библиотечно-информационной работы определяется не столько уровнем новейших технологий, сколько разумностью решений, принимаемых при их внедрении. Существующие преимущества «Библиотеки 2.0» могут быть использованы в библиотечной практике как составная часть социальных технологий, но «ограничить эту деятельность рамками только новых интернет-проектов не представляется возможным» [Миронов, 2011: 34].

Тем не менее, активное внедрение новых информационных технологий в библиотечную деятельность дало толчок к оптимизации всех технологических процессов на новых, более рациональных основаниях. Соответственно, и вебометрия, и веб-аналитика возникли в библиотеках не на пустом месте. Прежняя статистика, раньше нередко существовавшая как формальная часть библиотечных отчётов, особенно в небольших учреждениях, со временем превратилась в важный инструмент управления. В библиотеках сложился свой собственный серьёзный объект управления – *информационно-статистические системы*, которые с разных сторон описывали весь процесс взаимодействия библиотеки со своими пользователями и окружающим веб-пространством.

«Если рассматривать библиотеку с системной точки зрения как объект управления, – пишут авторы одного из обзоров, – то статистика необходима ей как инструмент оперативной обратной связи между элементами управляемой структуры, она нужна как средство быстрого обнаружения тенденций развития, поведения её внутренних и внешних связей. Статистика нужна как метод выбора совершенствования работы. Результаты анализа статистики являются базой планирования» [Зеленина, 2012: 16].

Этой новой роли статистики, включая и веб-метрики, противоречили старые взгляды. Укоренившееся формальное отношение прорывалось в высказываниях, что статистика это «самая точная из всех лженаук» [Зеленина, 2012: 12]. После посещения нескольких немецких библиотек, С.Ю. Волженина констатировала, что «имея в руках такой важный информационный ресурс для управления, как статистика, российские библиотекари по-прежнему не используют её потенциал» [Волженина, 2013: 8]. Сайты многих российских библиотек, даже федеральных, в отличие от их немецких коллег, стремились скрывать статистические показатели своей деятельности.

«Нужно ли говорить о том, – отмечала Волженина, – что открывая статистическую информацию, библиотеки формируют у потребителей представление о своей работе и информируют их о том, как потрачены средства налогоплательщиков» [Волженина, 2013: 8]. В статье содержался призыв, что время для воплощения в теории и практике

статистических подходов к оценке эффективности библиотек вполне назрело [Волженина, 2013: 8].

Включение в библиотечные процессы компьютерных технологий и новых возможностей, с ними связанных, в том числе для учёта и анализа данных, склонило подавляющее большинство библиотечной общественности к тому, что традиционная статистика и веб-метрики могут в значительной мере повысить эффективность диалога библиотек со своими пользователями. По мнению Т.Н.Пучининой, эти данные составляют «часть информационной составляющей фонда библиотеки. Обращаясь к ним, читатели удовлетворяют свои информационные потребности, используют их в научной, учебной, производственной работе. Библиотека, предоставляя доступ к этим ресурсам, может учитывать их как часть своего информационного потенциала и вести их учёт» [Пучинина, 2012: 24].

Овладение новыми веб-технологиями, однако, шло далеко неравномерно, преодолевая ряд ступеней. Поначалу сотрудники библиотек прикладывали немало сил для освоения неизвестного информационного пространства www, имеющего другой язык коммуникации, иные правила и принципиально другие ресурсы – как по форме, так и по содержанию. Большинство библиотек вышли в сеть с уникальными продуктами и услугами, нацеленными, в первую очередь, на краеведческие задачи [Левшина, 2012]. Затем новые возможности позволили подняться на более высокий уровень обслуживания, увеличить скорость поиска информации и предоставления пользователям. Были организованы сервисы виртуальной справки и электронной доставки документов.

Первые попытки ведения библиотечного учёта веб-материалов, как нетрудно догадаться, опирались на традиционные методы и, естественно, грешили ошибками. Например, отдел электронной информации Челябинской областной библиотеки организовал анализ обращения к сайтам по показателю «книговыдача» на основе листов учёта, обязав пользователей делать отметки о количественной и тематической стороне посещённых ресурсов. Сразу выяснилось, что пользователи предпочитают отмечать обращение не к сайтам, а к документам (веб-страницам), так как поисковая выдача Google, Яндекс и других систем предоставляет документы именно в таком виде. Часть пользователей вообще не хотели терять время на заполнение листов. «Процедура перестала вызывать негативную реакцию» только после соответствующих разъяснений сотрудников библиотеки [Пучинина, 2012: 24]. Дальнейшее наблюдение за читателями показало, что они вносят в листки учёта недостоверную информацию, и следующей заботой библиотеки стала необходимость разработки критериев оценки погрешности данных.

Знакомясь со всеми перипетиями учёта и освоения нового библиотечного веб-пространства, понимаешь, что ошибка библиотеки была связана с отвлечением внимания пользователей от их прямых задач на второстепенный, с их точки зрения, вопрос. На наш взгляд, более продуктивным был бы незаметный, объективный и обезличенный контроль за работой пользователей, под исключительную ответственность библиотеки.

Другая ошибка заключалась в том, что из статистики обращений заранее исключались доступы к социальному вебу – соцсетям, электронной почте, блогам, форумам и т. д., хотя уже тогда на социальный трафик приходилось до 30% запросов пользователей [Пучинина, 2012: 24]. Сотрудники заблуждались, считая, что нужно игнорировать эту статистику, «так как она не несёт информативной составляющей для образовательной и научной деятельности», или же отнести её к «досуговой деятельности (раздел ББК 77 "Досуг")» [Пучинина, 2012: 24]. Из библиотечного учёта мог быть исключён огромный пласт профильной, самообразовательной информации, количество которой в 2010-е гг. стало быстро нарастать. Сегодня библиотеки и библиотекари широко представлены во всевозможных социальных сетях и блогах.

В связи с потерей традиционного читателя, работа с социальным вебом становится всё более актуальной и, можно сказать, приоритетной частью библиотечной деятельности. Но

качество социального трафика, как выясняется, может носить как «позитивный», так и «негативный» характер [Комаров, 2017: 128]. Автор статьи рекомендует сосредоточиться на внешнем трафике, «ориентированном на сугубо научные, или, как минимум, научно-популярные, ресурсы академических библиотек», а также ресурсы, относимые к научным или научно-популярным: тематические группы (паблики) по поиску научной и учебной литературы, официальные или любительские группы научно-образовательных организаций и учреждений, сообщества студентов и аспирантов, сообщества научных и публичных библиотек, журналов и пр. [Комаров, 2017: 128]. Негативный трафик после проведения соответствующего анализа предлагается отсекаать из всех проектов и отчётов [Комаров, 2017: 134].

Этап осмысления, рефлексии и обратной связи, наступивший в связи с интенсивным развитием библиотечного интернета, строительством сайтов, созданием корпоративных сетей и пр., ещё больше усилил внимание к вебметрии как средству достижения самых эффективных результатов. В современных работах необходимость «регулярного и систематизированного сбора сведений о состоянии и развитии библиотеки путём регистрации конкретных данных, характеризующих количественную и качественную стороны библиотечных процессов» уже не подвергается никакому сомнению [Барышев, 2018: 17-18]. Двудеяная цель сбора данных состоит в том, чтобы регулярно отслеживать потребности пользователей и, наряду с этим, оценивать эффективность затрат на сопровождение библиотечных ресурсов и сервисов [Барышев, 2018: 18].

В этой же статье, посвящённой новым подходам к статистическим формам в цикле обслуживания читателей, отмечается, что «совершенствование статистических показателей вузовских библиотек, учитывающих изменения в структуре фондов, появление новых перспективных источников информации и новых запросов и потребностей пользователей, способствует своевременной выработке новых подходов к организации библиотечно-информационного обслуживания» [Барышев, 2018: 23-24]. Авторы считают, что «библиотечная статистика важна не только для библиотечного дела, но и для всей сферы культуры, т. к. позволяет сравнивать уровни развития учреждений этой сферы, выявлять новые тенденции, оценивать их перспективы» [Барышев, 2018: 23]. Этот подход особенно ярко проявляется в *библиотечной веб-аналитике* как самостоятельной отрасли электронного библиотековедения.

Кроме уже упоминавшейся статьи Филиппова [2012] в последнее время опубликовано немало работ, освещающих разные аспекты данной темы. В них исследуется *многообразие инструментов* [Васьковский, 2015; Редькина, 2017а; Скородумов, 2015; Ударцева, 2018в], *методологические подходы* к изучению показателей [Брумштейн, 2017; Булычева, 2018; Гуськов, 2015; Земсков, 2016; Косяков, 2016; Редькина, 2018; Рыхторова, 2018; Скородумов, 2016; Ударцева, 2018а]. О.М.Ударцева опубликовала литературный обзор, посвящённый истории и современным тенденциям развития вебметрии в библиотеках [Ударцева, 2018б; Udartseva, 2018]. Иными словами, тема становится «модной», хотя её сложность и запутанность от этого никак не уменьшается. Следует иметь в виду, что веб-аналитика – это всего лишь средство повышения эффективности библиотечной деятельности в интернете, но никак не самоцель. Сосредоточенность на проблемах библиотечной веб-аналитики не должна идти в ущерб работе по наращиванию ресурсов – справочных, библиографических, полнотекстовых и любых других.

По мнению Филиппова, главная задача веб-аналитики состоит не столько в сборе необходимых данных, сколько в их правильной интерпретации. При выработке стратегии повышения эффективности именно интеллектуальная обработка данных становится важной, ключевой задачей [Филиппов, 2012: 59]. С точки зрения автора, гораздо более продуктивно анализировать поведение посетителей сайта, чем заикливаться на анализе статистики посещаемости, нацеленном на «поисковую оптимизацию и корректировку входящего трафика» [Филиппов, 2012: 61].

Для решения стратегических задач и повышения эффективности сайта автор выделяет нескольких ключевых метрик, среди которых называются: общее количество уникальных посетителей, равное сумме уникальных cookie_id, число посещений (сумма сессий), источники трафика (рефереры и ключевые поисковые термины, небрендируемые ключевые слова), целевые показатели и поведенческие факторы (плотность кликов, популярные входные страницы, самые просматриваемые ресурсы) [Филиппов, 2012: 62-63].

Сбор и анализ статистических данных проводятся с помощью комплекса инструментов и систем веб-аналитики, имеющих разнообразный функционал. «Используя различные источники информации о тенденциях развития и возможностях различных счётчиков, сервисов и лог-анализаторов, – пишет Н.С.Редькина, – можно выбрать наиболее оптимальный инструмент для изучения сайта конкретной организации» [Редькина, 2017а: 9]. Здесь следует пояснить, что с помощью этих инструментов нельзя получить знание *о любом сайте*, а только о том, к статистике которого у вас есть доступ. Обычно веб-мастера ограничивают доступ посторонних к серверам и размещённой на них служебной информации (лог-файлам). Кроме того, более точно говорить не об изучении сайта, который строится своими же руками, а об использовании всех инструментов для его дальнейшего развития и повышения эффективности. Именно для этой цели и проводится вся веб-аналитическая работа. Что касается главного критерия для выбора инструментов веб-статистики и аналитики, то им являются *приоритетные цели сайта*.

Сбор показателей, анализ и выработка необходимых решений становятся условием эффективности библиотечных веб-технологий. Они способствуют реализации задач по обслуживанию читателей на новой технологической основе, привлечению новых пользователей и рационализации библиотечной технологии. При этом, замечает Н.С.Редькина, «недостаточно создать информационный ресурс или услугу и представить их в интернете. Важно, чтобы они были востребованными и приносили удовлетворение не только создателям, т. е. ресурс должен быть эффективным, с богатым набором сервисов, реализующим цели пользователей, учитывающим их предпочтения. Это может быть достигнуто посредством оценочного изучения веб-технологий с помощью различных инструментов, методов и средств, экспертной оценки на основании материала, полученного в ходе формализованного исследования (статистических данных с счётчиков или лог-файлов)» [Редькина, 2017б: 23-24].

Обратим особое внимание ещё на одну тему вебметрических рейтингов, возникших во 2-й половине 1990-х гг. Рейтинги составляются на основе разных критериев для сайтов научных и образовательных учреждений, вузов, научных журналов и т. д. – как в России, так и за рубежом. Критически важным показателем для создания таких рейтингов является так называемая ссылочная популярность (количество и качество внешних ссылок на сайты и отдельные документы), рассчитанная на базе алгоритмов крупнейших поисковых систем Google, Яндекс и др., большинство из которых являются закрытыми.

По мнению Ю.Е.Поляка, большинство других параметров, таких как дизайн и «юзабилити» сайта, а также популярность контента, выраженная в числе посещений, «в расчётах не участвуют и на результат не влияют» [Поляк, 2017: 22]. Точно таким же образом можно утверждать, что на рейтинги не сильно влияют и особенности пользовательской аудитории вебсайтов. Согласно исследованию, проведённому учёными СО РАН, такие KPI (ключевые показатели эффективности сайта), как общее число посетителей и сессий, а также количество просмотров страниц принципиально не влияют на его место в рейтингах. Авторам не удалось обнаружить принципиальную зависимость вебметрических рейтингов от посещаемости. Ранговая корреляция оказалась умеренной, а веб-трафик с других сайтов по внешним ссылкам – слабым [Гуськов, 2015: 12, 26].

Данное исследование выявило ряд проблем, снижающих посещаемость сайтов научно-исследовательских учреждений. Среди них: отсутствие деления на тематические разделы, дефицит перекрёстных ссылок внутри сайта, затруднённая навигация и ряд др.

Что же касается структуры веб-трафика, то исследование выяснило природу прямых заходов, которые указывают на постоянную аудиторию сайта, состоящую преимущественно из сотрудников и их коллег по НИИ. В силу этого, контактные сведения и перечень трудов сотрудников с полными текстами имеют значительную ценность для привлечения трафика из поисковых систем. Одновременно было «показано, что ссылки с других сайтов имеют низкую ценность, поскольку привлекают относительно небольшую долю веб-трафика» [Гуськов, 2015: 25-26].

Выводы исследования, безусловно, вносят большой вклад в вебометрию, но работа, выполненная на материалах научно-исследовательских сайтов, не может заменить специализированное изучение сайтов библиотек с полным отражением их специфики. Среди особенностей электронных библиотек назовём только несколько. Это, во-первых, большие объёмы политематических веб-ресурсов разного типа, предназначенных преимущественно для внешнего пользователя; во-вторых, предоставление ресурсов на условиях *open access*; далее: борьба с засилием спам-ботов (проблема, малоизвестная для сайтов НИИ); специфические подходы к созданию различных библиотечных веб-продуктов (навигаторов и др.); сервисная веб-деятельность библиотек (консультирование, рассылки и пр.); активность библиотек в соцсетях и целый ряд других.

С середины 2000-х гг. лидером международных вебометрических проектов научно-образовательной сферы стал сервис «Webometrics Ranking of World Universities», разработанный в лаборатории киберметрики Высшего совета по научным исследованиям министерства науки и инноваций Испании (CSIC). Его результаты публикуются дважды в год в январе и июле. Теперь это один из самых авторитетных рейтингов для сайтов научных и образовательных учреждений во всём мире. В России вебометрические оценки страдают различными недостатками [Московкин, 2016: 7] и, в целом, затухают. По мнению Ю.Е.Поляка, «вебометрические исследования не вошли в число приоритетных направлений исследований; они развивались усилиями отдельных энтузиастов, а при изменении условий работы (организационные перемены, сокращение финансирования) прекращались» [Поляк, 2017: 26].

Составление рейтинга сайтов научных организаций СО РАН, разработанного в 2008–2014 гг. Институтом вычислительных технологий (ИВТ СО РАН) по образцу Webometrics [Шокин, 2012] приостановилось в сентябре 2014 г. Аналогичная судьба постигла и другие рейтинги, созданные в Карельском научном центре РАН, Сибирском федеральном университете и Дальневосточном геологическом институте ДВО РАН. Проект «Российский вебометрический индекс научных и образовательных учреждений», который в 2012-2013 гг. вела лаборатория вебометрики Института научной и педагогической информации (ИНИПИ) РАО прекратил своё существование вместе с закрытием лаборатории. Аналогичных рейтинговых проектов для сайтов библиотечных учреждений не существует.

ЗАКЛЮЧЕНИЕ

К сожалению, не все направления были отражены в данном издании, поскольку они требуют отдельного рассмотрения. В первую очередь это касается трудов В.Е. Ларичева по палеоастрономии.

Не отражена была и методика наукометрического анализа в области археологии, которая ранее рассматривалась в ряде статей на страницах издания "Информационные технологии в гуманитарных исследованиях". В первую очередь, это цикл статей Н.А. Мазова, Ю.П. Холюшкина и В.С. Костина [Холюшкин, Костин, 2010: с.66–71; Костин, Холюшкин, 2011: с.37-57; они же, 2012: с.87-92]

Авторы предлагали целый спектр не только статистических, но и других методов и подходов, которые используют в наукометрических исследованиях, например, анализ нейронных сетей [Гарскова, 2018: с.213].

На основе результатов исследований был создан граф, отражающий связи между авторами издания в форме "незримых научных коллективов". Этот графы строится на основе таблицы сопряженности между авторами по количеству совместно написанных работ. Так была выявлена локальная региональная сеть, в которую входят в основном сотрудники Новосибирского научного центра СО РАН, публиковавшихся в ИТГИ. Эта сеть отличается очень высоким процентом работ в соавторстве. Центральными фигурами здесь являются Ю.П. Холюшкин, В.Т. Воронин, А.П. Деревянко, С.А. Федоров (рис.75).

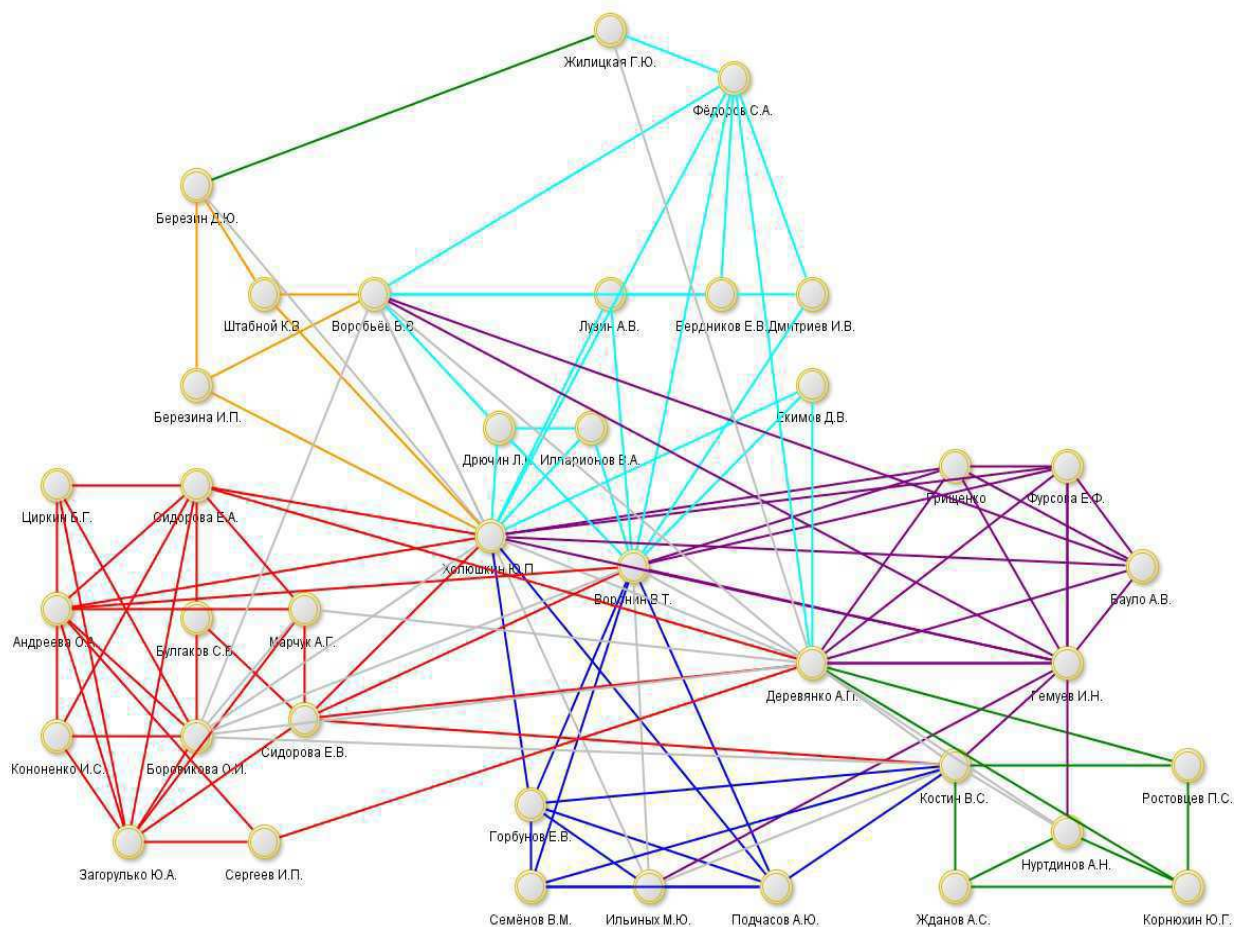


Рис.75. Локальная региональная сеть научных направлений, отраженная в публикациях ИТГИ.

Такой подход позволяет изучать формирование как формальных, так и "виртуальных" научных коллективов. Эти коллективы можно представить графически в виде подгрупп на

графах научных сетей, изучать их динамику, концентрацию, центральные фигуры, вокруг которых формируются эти группы. Связывая информацию о группах с тематическими рубриками публикуемых работ, можно выявлять специфику научных интересов, сходство и различия научных течений и школ.

Тематика "Интеллектуальные информационные системы" отражена в 11 статьях, опубликованных сотрудниками ИАЭТ и ИСИ СО РАН. В левой нижней части графа представлена группировка с тремя вершинами ребер (Ю.П. Холюшкин, В.Т. Воронин, А.П. Деревянко). В публикациях этого цикла информационные системы нового поколения авторами выделялись как распределенные системы, опирающиеся на множественные базы "осмысленных" данных, содержащие неспецифические общие данные, неспецифические приватные данные, специфические для информационной системы общие и частные данные, спецификации модели мира, предметной области и задачи. Совместимость данных и описаний должна в этих системах обеспечиваться общей методологией, едиными стандартами. На основе системной классификации археологии и данной технологии был создан портал археологических знаний.

В центральной нижней части рисунка расположена подсеть авторов – создателей виртуального музея в составе дипломников ВКИ НГУ с тремя вершинами ребер (Ю.П. Холюшкин, В.Т. Воронин, В.С. Костин).

Коллективом этой группы были выполнены проектно-технические решения по созданию интерфейса поисковой системы к базе данных портала «История народов Сибири и Дальнего Востока». Основная цель этой разработки состояла в том, чтобы встроить в поисковую систему портала удобные схемы наиболее современных технологий наглядного графического представления результатов поиска и доступа к данным портала.

Все эти схемы были интегрированы в две большие относительно самостоятельные части проекта:

- поисковая система базы данных портала;
- генератор его виртуальных залов.

Наибольший удельный вес и первую позицию по значимости имеют группа авторов статей, посвященных разработке и применению математических и статистических методов в археологии – 27 (21.8%). Группа расположена в нижней правой части графа. Как отметил авторитетнейший специалист в области исторической информатики, профессор МГУ Л.И. Бородкин, авторы статей ИТГИ успешно развивали потенциал новосибирской школы, получившей широкую известность в области применения современных методов многомерного статистического анализа в гуманитарных исследованиях.

Им же отмечены, что статьи издания отражают передовые позиции сибирских ученых в разработке тематических электронных ресурсов по археологии, этнологии и истории, создании электронных библиотек, энциклопедических словарей и обеспечении доступа к ним с использованием метаданных. Эта группа исследователей занимает верхние центральную и правую часть графа.

Достаточно значимыми по числу статей представлены разделы, посвященные методологии археологической науки и науковедческим исследованиям, а также палеоастрономии.

Таким образом, проблемно-тематическое обозрение статей позволяет утверждать, что серийное издание «Информационные технологии в гуманитарных исследованиях» рельефно фиксирует "эпицентры" исследовательских интересов археологов, этнографов, математиков, специалистов в области информатики и отражает их тесные взаимосвязи. В основном данное издание помогает ввести в научный оборот новые данные и способствует эффективному приращению и продвижению вперед научного знания в сфере археологической теории и исторической информатики.

Абдулганеев М.Т., Владимиров В.Н. Программа FUZZYCLASS: новые возможности археологического исследования (типология поселений Алтая раннего железного века) // Круг идей: новое в исторической информатике. М. 1994. С. 121–128;

Абдулганеев М.Т., Владимиров В.Н. Типология поселений Алтая 6–2 вв. до н.э. Барнаул, 1997.

Аврамова М.Б. Библиотечная статистика: в поисках оптимальных решений // Библиотековедение. 2016. Т. 65. № 4. С. 386–392.

Адлер Р., Эвинг Дж., Тейлор Р. П. Статистики цитирования // Игра в цифры, или как теперь оценивают труд ученого (сборник статей о библиометрике). М.: МЦНМО. 2011. С. 6–38.

Айгистов Р.А. Цифровые ресурсы в зарубежных библиотеках / Айгистов Р.А. // Библиография. Научный журнал по библиографоведению, книговедению и библиотековедению. 2016. №1(402). С. 61–94.

Аленова Е.М. Государственный каталог Музейного фонда РФ, Москва «К вопросу о создании Государственного каталога Музейного фонда Российской Федерации». Доклад на международной конференции EVA2003 Москва, 1 – 5 декабря 2003г., Государственная Третьяковская галерея. www.evarussia.ru

Андреева О.А., Боровикова О.И., Булгаков С.В., Загоруйко Ю.А., Сидорова Е.А., Циркин Б.Г., Холюшкин Ю.П. Археологический портал знаний: содержательный доступ к знаниям и информационным ресурсам по археологии // КИИ-2005. Т.3. С. 832–840.

Андреева О.А., Сергеев И.П., Холюшкин Ю.П. Информационная система «Системная археология»// Информационные технологии в гуманитарных исследованиях. Выпуск 7. Новосибирск: РИЦ НГУ. 2004. С. 39–44.

Антопольский А.Б. Вопросы оптимизации информационного пространства общественных наук // Информация и инновации. 2017. № S1. С. 16–21.

Антопольский А.Б. Измерение присутствия в Интернете образовательных учреждений / Антопольский А.Б., Поляк Ю.Е., Усанов В.Е. // Проблемы современного образования. 2012. № 4. С. 117–131. URL: <https://cyberleninka.ru/article/n/izmerenie-prisutstviya-v-internete-obrazovatelnyh-uchrezhdeniy>

Антопольский А.Б. Информационные ресурсы академического сектора исторической науки как источник единого российского электронного пространства знаний // Историческая информатика. 2018. № 1(23). С. 14–27. URL: http://nbpublish.com/library_read_article.php?id=26005

Антопольский А.Б. К вопросу о едином электронном пространстве знаний / Антопольский А.Б., Ефременко Д.В. // Вестник Российской академии наук. 2018. Т. 88. № 2. С. 163–170.

Асеев Ю.А., Поднозова И.П., Шер Я.А., Каталогизация музейных коллекций и информатика // Современный художественный музей. Проблемы деятельности и перспективы развития. Л., 1980.

Афанасьева О.Г. Выставка-инсталляция в пространстве библиотеки и веб-пространстве / Афанасьева О.Г. // Библиотечное дело. 2012. №16(178). С. 20–22.

Афанасьев Г.Е., Чернышев А.В. Применение ГИС-технологии в археологических исследованиях // Картография на рубеже тысячелетия. Доклады I Всероссийской науч. конф. по картографии. М. 1997.

Барышев Р.А., Бабина О.И., Сергиенко Т.В., Захаров П.А. Новые подходы к разработке статистических форм в цикле обслуживания читателей // Научные и технические библиотеки. 2018. № 2. С. 16–25.

Беленький К.Г., Витяев Е.Е., Костин В.С., Холюшкин Ю.П. WEB-портал статистической обработки археологических данных // ИТГИ. – Вып. 16. Новосибирск : Изд-во НГУ. 2011. С. 71–77.

Бенедетти Б., Маши М.Э. База данных он-лайн по археологии: Помпеи и проект "Fortuna Visiva".// Электронные библиотеки. 2006. Том 9. Выпуск 2. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part2/BM>

Березина И.П., Березин Д.Ю., Холюшкин Ю.П. Немецко-русский и русско-немецкий археологический словарь с общенаучной лексикой. Новосибирск: РИЦ НГУ, 2003.

Борисенко А.Ю., Бериков В.Б., Ведерников Ю.А., Лбов Г.С., Худяков Ю.С. Перспективы математического выявления причинно-следственных связей при изучении типологических изменений в процессе развития панцирных доспехов средневековых кочевников Центральной Азии // ИТГИ. Вып. 12. Новосибирск : Изд-во ООО Прайс-Курьер, 2008. С. 90–98.

Боровикова О.И. Онтология предметной области «Археология и этнография» для портала научных знаний / Информационные технологии в гуманитарных исследованиях. Вып. 12. Новосибирск, 2008. С. 20–28.

Боровикова О.И., Булгаков С.В., Загорюлько Ю.А., Сидорова Е.А., Холюшкин Ю.П. Система знаний информационного интернет-портала по археологии и этнографии // Информационные технологии в гуманитарных исследованиях. Вып. 9. – Новосибирск: Изд. НГУ, 2005. С. 33-39.

Боровикова О.И., Загорюлько Ю.А. Организация порталов знаний на основе онтологий // Труды международного семинара Диалог'2002 “Компьютерная лингвистика и интеллектуальные технологии”. Протвино. 2002. Т.2. С. 76– 82.

Боровикова О.И., Загорюлько Ю.А., Сидорова Е.А. Подход к автоматизации сбора онтологической информации для интернет-портала знаний // Труды междунар. конф. Диалог'2005 “Компьютерная лингвистика и интеллектуальные технологии”Звенигород, 1-5 июня 2005, М.: Наука. 2005. С. 65-70.

Бородкин Л.И. Историческая информатика начала XXI века: спрос на специальные алгоритмы и технологии // Информационный Бюллетень Ассоциации «История и компьютер». 2004. №32. С. 3–6.

Бородкин Л.И. Историческая информатика в СССР/России: ретроспектива, состояние, перспективы // История и компьютер: новые информационные технологии в исторических исследованиях и образовании. St. Katharinen, 1993. С. 251–273.

Беседин В.И. Об эффективности и надежности метода размытой классификации // Информационный Бюллетень Ассоциации «История и компьютер». 1996. №17. С. 80–81.

Беседин В.И. Применение метода размытой классификации для морфологического анализа древней керамики // Информационный Бюллетень Ассоциации «История и компьютер». 1995. №14. С. 76–77.

Браккер Н. В. Оцифровка, доступ и сохранение цифрового культурного и научного наследия: новые инициативы Европейской комиссии// Доклад на Международной научной конференции “Румянцевские чтения–2006: Библиотеки, музеи, архивы в формировании интеллектуального и информационного пространства”, 12 апреля 2006 г., Российская государственная библиотека, Москва.

Брумштейн Ю.М. Анализ вебметрических показателей основных сайтов, агрегирующих политематическую научную информацию / Брумштейн Ю.М., Васьковский Е.Ю. // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2017. № 11. С. 16-32.

Булычева О.С. Концептуальные положения и предпосылки создания вебметрической системы цифрового пространства библиотек / Булычева О.С., Сюнтюрено О.В. // Цифровые проекты в современной информационной среде: наука и практика: сб. науч. тр. Сер. «Электронная библиотека». СПб., 2018. С. 19-31.

Васьковский Е.Ю. Системный анализ функциональных возможностей счетчиков посещаемости сайтов / Васьковский Е.Ю., Брумштейн Ю.М. // Прикаспийский журнал: управление и высокие технологии. 2015. № 3. С. 96-113.

Виртуальный музей СО РАН / Ю.И.Шокин, А.М.Федотов, В.А.Ламин, В.Б.Бараннин, И.В.Бычков, А.Е.Гуськов, О.А.Клименко, Ю.В.Леонова, Н.А.Мазов, В.В.Москвичев, Ю.И.Молородов, Б.Н.Пищик, В.П.Потапов, Е.В.Рычкова, А.З.Фазлиев, Ю.П.Холюшкин, И.В.Шабальников// Распределенные информационно-вычислительные ресурсы: X Российская конференция с участием иностранных ученых. - Новосибирск: ИВТ СО РАН, 2005. - С.51.

Витяев Е.Е. Извлечение информации из данных // Информационные технологии в гуманитарных исследованиях. Новосибирск. 2010. Вып. 15. С. 84-87.

Витяев Е.Е., Костин В.С. Естественная классификация, систематика, онтология. Информационные технологии в гуманитарных исследованиях, Вып. 13, ИАЭТ СО РАН, Новосибирск, 2009: с. 65-75

Витяев Е.Е., Костин В.С. Естественная классификация как закон природы // Интеллектуальные системы и методология: Материалы науч.-практич. симпозиума

"Интеллектуальная поддержка деятельности в сложных предметных областях", Новосибирск, 7-9 апреля 1992. - Новосибирск, 1992. - Вып. 4. - С. 107-115.

Витяев Е.Е., Логвиненко А.Д. Обнаружение законов на эмпирических системах и тестирование систем аксиом теории измерений // Социология: 4М (методология, методы, математическое моделирование), 1998, №10: 97-121.

Витяев Е.Е., Москвитин А.А. ЛАДА – программная система логического анализа данных // Методы анализа данных (Вычислительные системы, вып. 111). Новосибирск, 1985: 38-58.

Витяев Е.Е., Москвитин А.А. Введение в теорию открытий. Программная система DISCOVERY // Логические методы в информатике (Вычислительные системы, вып. 148). Новосибирск, 1993. С. 117 -163.

Витяев Е.Е., Москвитин А.А. Введение в теорию открытий. Программная система DISCOVERY. // Логические методы в информатике (Вычислительные системы, вып. 148), Новосибирск, 1993, с.117-163.

Волженина С.Ю. Немецкая библиотечная статистика: размышления российского библиотекаря // Библиотечное дело. 2013. № 1(187). С. 6-8.

Гарден Ж.-К. Теоретическая археология. М.: Прогресс. 1983. 296 с.

Гарскова И.М. Историческая информатика: методологические и историографические аспекты развития. Дисс. На соискание учёной степени доктора исторических наук. М. 2018. 618с.

Гарскова И.М. Сетевой анализ историографии: динамика формирования региональных центров исторической информатики // Историческая информатика. № 3. 2017.

Гарфилд Ю. Можно ли выявлять и оценивать научные достижения и научную продуктивность? // Вестник АН СССР. 1982. № 7. С. 42-50.

Гендина Н.И. Библиотека в едином информационном пространстве: необходимость создания электронных путеводителей по интернет-ресурсам / Гендина Н.И., Колкова Н.И. // Научные и технические библиотеки. 2018. № 7. С. 43-59.

Гинзбург Э.Х., Горенштейн Н.М., Ранов В.А. Статистико–математическая обработка шести мусьерских памятников Средней Азии.// Палеолит Средней и Восточной Азии. История и культура Востока Азии. – Новосибирск, 1980: 7 –31.

Городцов В.А. Археология. Т.I. Каменный период. М.–Л., 1923: 398 с.

Горячев Д.Н. Методы элиминации масштабной сетки в процессе обработки графических полевых документов // Методология и методика археологических реконструкций. Новосибирск, 1994: с. 100-101.

Горячев С.А. Из опыта работы ДВГНБ в веб-пространстве / Горячев С.А. // Вестник Дальневосточной государственной научной библиотеки. 2017.№2(75). С. 70-76.

Гражданников Е.Д. Метод построения системной классификации наук. – Новосибирск, 1987.

Гуськов А.Е. Альтернативная вебометрика: исследование веб-трафика сайтов научных организаций / Гуськов А.Е., Быховцев Е.С., Косяков Д.В. // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. 2015. № 12. С. 12-28.

Гуськов А.Е. О присутствии академических библиотек в научном веб-пространстве / Гуськов А.Е., Косяков Д.В., Быховцев Е.С. // Книга. Культура. Образование. Инновации («Крым-2016»): материалы 2-го Международного профессионального форума. 2016. С. 422-429.

Деревянко А. П., Фелингер А. Ф., Холюшкин Ю. П. Методы информатики в археологии каменного века. Новосибирск, 1989. 268 с.

Деревянко А.П., Холюшкин Ю.П., Бердников Е.В., Воронин В.Т. ГИС «Палеолит Северной Азии» // Информационные технологии в гуманитарных исследованиях. Вып. 6. Новосибирск, 2003: с. 21–29.

Деревянко А.П., Холюшкин Ю.П., Воронин В.Т. Современные информационные технологии и проблемы археологической информатики // Математические методы в археологических реконструкциях (методология и методика археологических реконструкций). Новосибирск, 1995: с. 19–31.

Деревянко А.П., Холюшкин Ю.П. Воронин В.Т., Екимов Д.В. О некоторых принципах создания экспертных систем в археологии // Методы естественных наук в археологических реконструкциях. – Новосибирск, 1995: с. 27-31.

Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Екимов Д.В. Концепция и архитектура информационного центра Института археологии и этнографии СО РАН // Компьютер и историческое знание. Барнаул, 1994. С. 36–54.

Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Екимов Д.В. Концепция информационного центра Института археологии и этнографии СО РАН // Методология и методика археологических реконструкций. - Новосибирск, 1994: с. 43-51.

Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Костин В.С. Корреляция среднепалеолитических индустрий Ближнего и Среднего Востока и Кавказа. Часть II. Типология. Новосибирск: РИЦ НГУ, 2004.

Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Костин В.С. Статистическое исследование среднепалеолитических индустрий Средней Азии и Казахстана. Новосибирск: РИЦ НГУ, 2005.

Дремайлов А.В. Вливание музейно-архивных электронных каталогов в единое информационно-поисковое пространство культурных ресурсов. Доклад на международной конференции EVA2003 Москва, 1 – 5 декабря 2003г., Государственная Третьяковская галерея. URL: www.evarussia.ru

Ершов Ю.Л., Деревянко А.П., Диев В.С., Зыбарев Ю.М., Холюшкин Ю.П., Розов Н.С., Воронин В.Т. Формирование региональной проблемно-ориентированной среды для информационной поддержки гуманитарных наук, образования и культуры / Ю.Л.Ершов, А.П.Деревянко, В.С.Диев, Ю.М.Зыбарев, Ю.П.Холюшкин, Н.С.Розов, В.Т.Воронин // Информационные технологии в гуманитарных исследованиях. - Новосибирск: НГУ, 1998. - Вып.1. - С.5-11.

Жабко Е.Д. Виртуальные социальные сервисы в среде электронных библиотек // Библиотечное дело. 2009. № 15(105). С. 18-19.

Жабко Е.Д. Среда электронных библиотек: новые возможности повышения социальной активности пользователей // Вестник Санкт-Петербургского государственного университета культуры и искусств. 2010. № 1(5). С. 105-108.

Жабко Е.Д. Электронные библиотеки как объект научных исследований / Жабко Е.Д. // Вестник Санкт-Петербургского государственного университета культуры и искусств. 2017. № 4(33). С. 54-60.

Жданов А.С., Костин В.С. Значимость и устойчивость автоматической классификации в задаче поиска оптимального разбиения // Информационные технологии в гуманитарных исследованиях. - Новосибирск: НГУ, 2002. - Вып. 3. - С. 36-42.

Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. ОИГТМ СО РАН, Новосибирск, 2004: 361 с.

Загорюлько Ю.А., Боровикова О.И., Холюшкин Ю.П. Построение предметной онтологии для археологического портала // ИТГИ. – Вып. 10. – Новосибирск : Изд-во НГУ, 2006: с. 24–30.

Зайцева О.Ф. Анализ использования электронных образовательных ресурсов // Опыт работы Научной библиотеки Воронежского государственного аграрного университета им. императора Петра I. 2015 год. Воронеж, 2016. С. 36-45.

Зверевич В.В. Пространство современной библиотеки: «реальное» и «виртуальное» / Зверевич В.В. // Научные и технические библиотеки. 2012. № 11. С. 7-17.

Зеленина Г.Н. О чём молчит статистика? На помощь приходит «Руслан» / Зеленина Г.Н., Поникаровский К.В., Артемьев М.М. // Библиотечное дело. 2012. № 15(177). С. 12-17.

Земсков А.И. Библиометрия в библиотеках / Земсков А.И., Колосов К.А. // Научные и технические библиотеки. 2016. № 11. С. 5-23.

Информационно-телекоммуникационные ресурсы Сибирского отделения РАН / Шокин Ю.И., Федотов А.М., Чубаров Л.Б., Жижимов О.Л. // Информационно-телекоммуникационные ресурсы СО РАН: материалы выездного заседания координационного научного совета СО РАН по целевой программе / РАН, Сибирское отделение; Иркутский научный центр. 2003. С. 150-163. URL: <http://www.ict.nsc.ru/ws/elpub2003/6154/rep6154.pdf>

Исмагилова А.Х. Оценка качества электронных библиотек / Исмагилова А.Х. // Научные и технические библиотеки. 2010. № 5. С. 60-66.

Кабакова Е.А. Веб-сайт научно-исследовательского учреждения: наполнение, посетители, развитие / Кабакова Е.А., Усков В.С. // Вопросы территориального развития. 2014. Вып.3(13). С. 1-12. URL: <https://cyberleninka.ru/article/n/veb-sayt-nauchno-issledovatel'skogo-uchrezhdeniya-napolnenie-posetiteli-razvitiye>

Кендалл М., Стьюарт А. Теория распределений. М.: Наука, 1966. **Клейн Л.С.** Новая археология. Донецк: ДОНУ, 2009: 393 с.

Клейн Л.С. Принципы археологии. СПб. 2001. 152 с.

Клименко О.А. Модели представления академического веб-пространства // Информационные и математические технологии в науке и управлении. 2016а. № 2. С. 103-110.

Клименко О.А. Проблемы изучения процессов самоорганизации в веб-пространстве / Клименко О.А., Фёдоров Р.Ю. // Образовательные ресурсы и технологии. 2016б. № 2(14). С. 168-173.

Кокорина Ю.Г., Лихтер Ю.А. Опыт использования программы нечеткой классификации FUZZYCLASS для исследования семантики изображений на скифском оружии // Круг идей: междисциплинарные подходы в исторической информатике. М., 2008. С. 301–309.

Колкова Н.И. Реализация функции региональной памяти центральных библиотек регионов РФ в условиях веб-пространства: состояние и проблемы / Колкова Н.И., Гендина Н.И., Алдохина О.И. // Создание официальных сайтов учреждений культуры и образования: теория и практика: сб. науч. тр. / под общ. ред. Н.И.Гендиной, Н.И.Колковой. СПб., 2015. С. 100-107.

Комаров С.Ю. «Позитивный» и «негативный» внешний трафик из социальных сетей к сайтам научных библиотек: анализ доступов к сайту Отделения ГПНТБ СО РАН (2014-2017 гг.) / С.Ю.Комаров // Информационные технологии в гуманитарных исследованиях. Вып.22: Очерки по науковедению. Новосибирск: ГПНТБ СО РАН, 2017. С. 128-134.

Коптева Е.В., Щипунов В.В. Применение методов кластерного анализа в археологии// Методология и методика археологических реконструкций. Новосибирск, 1994: с. 55-59.

Корпоративная сеть Дальневосточного отделения РАН / Ханчук А.И., Сорокин А.А., Наумова В.В., Нурминский Е.А., Смагин С.И., Ворошин С.В., Казанцев В.А. // Вестник Дальневосточного отделения РАН. 2007. № 1(131). С. 3-19. URL: <https://cyberleninka.ru/article/n/korporativnaya-set-dalnevostochnogo-otdeleniya-ra-n>

Костин В.С. Статистика для сравнения классификаций // Информационные технологии в гуманитарных исследованиях: Новосибирск, 2003. Вып. 6. С. 57-65.

Костин В.С., Корнюхин Ю.Г. Построение обобщенной классификации // Информационные технологии в гуманитарных исследованиях: Сб. тр. - Новосибирск, 2003. - Вып. 6. - С. 65-72.

Костин В.С., Постнов А.В., Хасанов С.А., Диконская Е.К., Кисарова В.П. Подготовка и генерация отчета по археологическим полевым работам средствами информационных технологий: анализ проблем и постановка задач // Информационные технологии в гуманитарных исследованиях. Вып. 15. Новосибирск, 2010. С. 60-65.

Костин В.С., Холюшкин Ю.П. Некоторые подходы к библиометрическому анализу взаимоссылаемости // Информационные технологии в гуманитарных исследованиях. Вып. 16. Новосибирск, 2011. С. 37–57;

Костин В.С., Холюшкин Ю.П. Практическое применение библиометрического анализа цитирования и самоссылаемости журналов археологического и этнографического профиля // Информационные технологии в гуманитарных исследованиях. Вып. 17. Новосибирск. 2012. С. 87–92.

Косяков Д.В. Академические институты России в зеркале вебометрики / Косяков Д.В., Гуськов А.Е., Быховцев Е.С. // Вестник Российской академии наук. 2016. Т. 86. № 11. С. 1015-1025.

Кудряшова Э.Е. Фрактальная модель информационного пространства / Кудряшова Э.Е., Копылова М.Ю., Чистов Д.А. // Успехи современного естествознания. 2007. № 10. С. 96-97.

Кузьмин Е.И., Куйбышев Л.А., Бракер Н.В. «Проект MINERVA PLUS и его реализация в России». Доклад на ежегодной конференции РБА, 17 – 21 мая 2004г., Новосибирск, 2004.

Курганская Л.М. Библиотечное пространство: понятие, виды и модернизация / Курганская Л.М., Кубаев А.А. // Наука Искусство Культура. 2017. № 3(15). С. 168-173.

Ларичев В.Е., Сазонов В.И. Интерпретация образов искусства древнекаменного века Сибири: Инновационные методы и реконструктивные технологии анализа геометрии скульптур из бивня мамонта и числовых записей на их поверхностях. Новосибирск: Изд. ГЕО. 2016: 356 с.

Левшина Н.Н. Сайт библиотеки как инструмент эффективного обслуживания виртуальных пользователей // Библиотечное дело. 2012. № 15(177). С. 20-22.

Лихтер Ю.А. Классификация древних материалов с использованием программы нечеткой классификации FUZZYCLASS // Информационный бюллетень Ассоциации «История и компьютер». 2000. №26/27. С. 76–82

Мазов Н.А., Жижимов О.Л., Федотов А.М. Классификация, определение и семантика музейных категорий данных, используемых при каталогизации информационных ресурсов // URL: http://conf.cpic.ru/upload/eva2004/reports/doklad_208.rtf

Марчук А.Г., Загорулько Ю.А., Холушкин Ю.П., Воронин В.Т. Концепция портала знаний по археологии и этнографии Сибири // Информационный Бюллетень Ассоциации «История и компьютер». 2004. №32. С. 74–75.

Медведев Г.И., К проблеме морфологического анализа каменного инвентаря палеолитических и мезолитических ансамблей Восточной Сибири// Описание и анализ археологических источников. Иркутск. 1981.

Медведев Г.И., Алаев С.Н., Алаева Т.В., Опыт применения полевой фиксационной карточки на палеолитических местонахождениях//Нач.-техн. конф. Археология, этнография Восточной Сибири Иркутск. 1978.

Миронов Н.А. Современное Web-пространство: о сущности и дефинициях // Библиотечное дело. 2011. № 8(146). С. 33-34.

Михайлова А.В. Анализ веб-страниц библиотек институтов культуры как инструмента реализации внешнего направления интеграции информационно-образовательных ресурсов // Вестник культуры и искусств. 2017. № 3(51). С. 49-59.

Московкин В. Университетский рейтинг Webometrics: технические проблемы в России // Троицкий вариант. 2016, № 216 (1 ноября). С. 7.

Музалевская И.М. Камис и ас "музей". наши комментарии // URL: <http://www.future.museum.ru/part03/03010301.htm>

Наумова В.В. Интегрированная информационная среда Дальневосточного отделения РАН и Дальневосточного федерального университета: текущее состояние и возможности развития / Наумова В.В., Владимиров Л.Г. // Открытое образование. 2014. № 6(107). С. 40-48.

Наумова В.В. Мониторинг информационного пространства Дальневосточного отделения РАН // Вестник Дальневосточного отделения Российской Академии наук. 2012. № 2(162). С. 117-122. URL: <https://cyberleninka.ru/article/n/monitoring-informatsionnogo-prostranstva-dalnevostochnogo-otdeleniya-ran>

Наумова В.В. Сервис вебометрического анализа структур научных сайтов / Наумова В.В., Платонов К.А. // Вестник Отделения наук о Земле РАН. 2013. Т. 5. NZ3001. URL: <https://onznnews.wdcb.ru/publications/v05/2013NZ000117/2013NZ000117.pdf>

Несговорова Г.П. Виртуальный музей – новая реальность // Информационные технологии в гуманитарных исследованиях. Вып. 10. Новосибирск, 2006. С. 96–100.

Олефир С.В. Статья информационным лидером. Школьная библиотека в веб-пространстве / Олефир С.В. // Библиотечное дело. 2010. № 19(133). С. 38-39.

О некоторых подходах к созданию музейного портала по археологии и этнографии Сибири и Дальнего Востока / Ю.П.Холушкин, В.Т.Воронин, В.С.Костин, М.Ю.Ильиных, В.М.Семенов, Е.В.Горбунов, А.Ю.Подчасов // Информационные технологии в гуманитарных исследованиях. - Новосибирск: НГУ, 2006. - Вып.10. - С.31-40.

Печников А.А. Анализ гиперссылок веб-пространства Петрозаводского государственного университета / Печников А.А., Дербенева О.Ю. // Учёные записки Петрозаводского государственного университета. 2013а. № 2(131). С. 100-106.

Печников А.А. Исследование взаимосвязей между веб-сайтами научных библиотек университетов России // Дистанционное и виртуальное обучение. 2011. № 7. С. 13-24.

Печников А.А. О некоторых тенденциях изменения связности российского академического веба // Информационные ресурсы России. 2014а. № 2(138). С. 16-20.

Печников А.А. Об измерениях вебометрических индикаторов // Международный журнал экспериментального образования. 2013б. № 10-2. С. 400-404.

Печников А.А. Применение вебометрических методов для исследования информационного веб-пространства научной организации (на примере Карельского научного центра РАН) // Труды Карельского научного центра РАН. 2013в. № 1. С. 86-95. URL: <https://cyberleninka.ru/article/n/primenenie-vebometricheskih-metodov-dlya-issledovaniya-informatsionnogo-veb-prostranstva-nauchnoy-organizatsii-na-primere-karel'skogo>

Печников А.А. Сравнение вебометрического ранжирования институтов РАН и их ссылочной популярности в русской Википедии / Печников А.А., Таракановский Н.А. // Научно-образовательная информационная среда XXI века: матер. IX Всерос. науч.-практ. конф. 2015. С. 148-150.

Печников А.А. Сравнительный анализ связности веб-графов научных учреждений // Современные проблемы науки и образования. 2014б. № 3, ст.77. С. 1-7.

Поляк Ю.Е. Оценивание и ранжирование веб-сайтов. Вебометрические рейтинги // Научный редактор и издатель. 2017. Т. 2, № 1. С. 19-29.

Постнов А.В., Вергунов Е.Г. Применение спутниковых навигационных приёмников при проведении археологических исследований // Информационные технологии в гуманитарных исследованиях. Вып. 5. Новосибирск, 2003. С. 67–99;

Постнов А.В., Вергунов Е.Г. Применение элементов ГИС-технологий при комплексных археологических исследованиях памятников // Информационные технологии в гуманитарных исследованиях. Вып. 10. Новосибирск, 2006. С. 66–71.

Потехина Ю.В. Управление официальным сайтом библиотеки как предмет научных исследований: постановка проблемы / Потехина Ю.В. // Библиосфера. 2011. № 4. С. 85-89.

Пучинина Т.Н. Информационные ресурсы как элемент библиотечной статистики: к постановке вопроса // Библиотечное дело. 2012. № 6(168). С. 23-24.

Редькина Н.С. Направления развития инструментов веб-аналитики // Научно-техническая информация. Сер. 2: Информационные процессы и системы. 2017а. № 5. С. 5-10.

Редькина Н.С. Оценка деятельности библиотеки в социальных сетях инструментами веб-аналитики / Редькина Н.С. // Научные и технические библиотеки. 2018. № 3. С. 16-23.

Редькина Н.С. Перспективные клиентоориентированные технологии обслуживания пользователей библиотек // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий: II междунар. науч. конф. / Белорус. с.-х. б-ка, Нац. акад. наук Беларуси; науч. ред. и сост. И.Б.Стрелкова. 2016. С. 342-347.

Редькина Н.С. Эффективные веб-технологии в деятельности библиотеки // Научные и технические библиотеки. 2017б. № 3. С. 15-24.

Родионов М.Д. Федеральный закон о Национальной электронной библиотеке и начало его реализации / Родионов М.Д., Сахаров Н.А. // Библиотековедение. 2017. Т. 66. № 1. С. 7-12.

Ростовцев П.С. Черно-белый анализ связи переменных // Анализ и моделирование экономических процессов переходного периода в России. Новосибирск: ИЭиОПП, 1996. С. 264-286.

Ростовцев П.С. Статистическое согласование мер связи в анализе социально-экономической информации // Экономика и математические методы. 1991. Т. 27. Вып. 1: с. 150-156.

Ростовцев П.С. Алгоритмы анализа структуры прямоугольных матриц "пятна" и "полосы" // Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. - С. 203-214.

Ростовцев П.С., Костин В.С., Корнюхин Ю.Г., Смирнова Н.Ю. Анализ структур социологических данных. Устойчивость // Анализ и моделирование экономических процессов переходного периода в России. – Новосибирск: ИЭиОПП: 1997. С. 174-208.

Ростовцев П.С., Корнюхин Ю. Г., Костин В.С., Смирнова Н.Ю. Алгоритмы анализа структуры и типологического группирования в анализе археологических данных // Методология и методика археологических реконструкций. Новосибирск, 1994: с. 59-68.

Рыхторова А.Е. Сегментирование пользователей сайта для продвижения библиотечных ресурсов и услуг / Рыхторова А.Е., Ударцева О.М. // Библиосфера. 2018. № 3. С. 59-67.

Серова О.В. Качество услуг в электронной среде и новые сервисы // Библиосфера. 2009. № 1. С. 27-32.

Сеть передачи данных Сибирского отделения РАН (СПД СО РАН) – Сеть Интернет Сибирского отделения РАН: информ. материалы науч.-коорд. совета целевой программы «Информационно-телекоммуникационные ресурсы СО РАН» / ИВТ СО РАН. Новосибирск, 2005. 79 с. URL: http://www-sbras.nsc.ru/net/files2/Reports/Telecom_up_to_2005.pdf

Скородумов П.В. Анализ подходов и инструментальных средств анализа статистики посещения веб-сайта научной организации / Скородумов П.В., Холодев А.Ю. // Вопросы территориального развития [сетевое издание]. 2015. № 9(29). С. 1-13. URL: <http://vtr.vscac.ru/article/1650> ; <https://cyberleninka.ru/article/n/analiz-podhodov-i-instrumentalnyh-sredstv-analiza-statistiki-posesheniya-veb-sayta-nauchnoy-organizatsii>

Скородумов П.В. Анализ популярности веб-сайта научной организации с помощью различных систем сбора статистических данных / Скородумов П.В., Холодев А.Ю. // Вопросы территориального развития [сетевое издание]. 2016. № 1(31). С. 1-13. URL: <http://vtr.vscac.ru/article/1772>

Тарасенко В.П. Некоторые проблемы формализации гуманитарных знаний (на примере археологии) // Информационные технологии в гуманитарных исследованиях. Вып. 2. Новосибирск. 2000. С. 84-87.

Третьякова Н.В. Библиотека в виртуальном пространстве: роль веб-ресурсов в работе библиографа / Третьякова Н.В. // Моргенштерновские чтения - 2018. Информационное обслуживание библиотек в меняющейся социальной среде: тенденции, новации, перспективы: материалы Всерос. науч.-практ. конф. Челябинск, 2018. С. 104-110.

Третьякова О.В. Возможности использования вебометрического анализа в оценке сайта научного института / Третьякова О.В., Кабакова Е.А. // Вопросы территориального развития [сетевое издание]. 2014. № 2(12). С. 1-10. URL: <http://vtr.vssc.ac.ru/article/1389> ; <https://cyberleninka.ru/article/n/vozmozhnosti-ispolzovaniya-vebometricheskogo-analiza-v-otsenke-sayta-nauchnogo-instituta>

Ударцева О.М. Веб-аналитика в библиотечной практике // Румянцевские чтения - 2018: библиотеки и музеи как культурные и научные центры: историческая ретроспектива и взгляд в будущее: к 190-летию со времени основания Румянцевого музея: материалы междунар. науч.-практ. конф. / Рос. гос. б-ка, Библ. Ассамблея Евразии. 2018а. С. 178-182.

Ударцева О.М. Вебометрия в библиотеках: история и современные тенденции развития (обзор) / Ударцева О.М. // Научно-техническая информация. Сер. 1: Организация и методика информационной работы. 2018б. № 9. С. 13-21.

Ударцева О.М. Использование инструментов веб-аналитики в оценке эффективности способов продвижения библиотечных ресурсов / Ударцева О.М., Рыхторова А.Е. // Библиосфера. 2018в. № 2. С. 93-99.

Федоров-Давыдов Г.А. Статистические методы в археологии. – М.: Высшая школа, 1987: 216с.

Фелингер А.Ф. Статистические алгоритмы в социологических исследованиях. Новосибирск: Наука, СО АН СССР, 1985.

Филиппов И. Web-аналитика: что это такое? / И. Филиппов // Современная библиотека. 2012. № 1(21). С. 58-63. URL: https://modern-lib.ru/assets/upload/journals/2012/1_2012_web.pdf

Фролов А.В., Фролов Г.В. Электронные конференции в Интернет// Газета "Капитал", рубрика "Три килобайта". 1996.

Ханчук А.И. Информационное пространство Дальневосточного отделения РАН / Ханчук А.И., Наумова В.В. // Вестник Дальневосточного отделения РАН. 2009. № 4(146). С. 122-129. URL: <https://cyberleninka.ru/article/n/informatsionnoe-prostranstvo-dalnevostochnogo-otdeleniya-ran>

Ханчук А.И. Корпоративная сеть ДВО РАН: высокотехнологичная интеграция научных подразделений / Ханчук А.И., Наумова В.В., Сорокин А.А. // Вестник Российской академии наук. 2008. Т. 78. № 4. С. 298-303. URL: https://elibrary.ru/download/elibrary_9953169_98311898.pdf

Холюшкин Ю.П. Системная археология. Новосибирск: НГУ, 2010.

Холюшкин Ю.П. Системная классификация науковедения (археологическое экспертное науковедение) // Научный альманах = Science Almanac. Тамбов, 2016. N 4-4(18); с.118-123. – DOI: 10.17117/na.2016.04.04.118. - URL: <http://ucom.ru/doc/na.2016.04.04.pdf>

Холюшкин Ю.П. Некоторые подходы к дифференцированной классификации цитирования / Ю.П.Холюшкин // Научные и технические библиотеки 2017. № 9. С. 104-113. URL: http://www.gpntb.ru/ntb/ntb/2017/9/NTB9_2017_A5_5.pdf

Холюшкин Ю.П. Введение в археологическое науковедение. Новосибирск: Изд-во НГУ. 2004. 58 с.

Холюшкин Ю.П., Березина И.П., Березин Д.Ю., Штабной К.В. О концепции создания Web-энциклопедий и словарей по археологии и этнографии Западной Сибири// Информационные технологии в гуманитарных исследованиях. Информационные технологии в гуманитарных исследованиях. №1. Новосибирск: НГУ, 1998: С.71-74.

Холюшкин Ю.П., Витяев Е.Е., Костин В.С. Задачи археологии и методы их решения. Вып.18. Новосибирск: Манускрипт. 2013. 98 с.

Холюшкин Ю.П., Витяев Е.Е., Костин В.С. К вопросу о поисках закономерностей в археологии // Информационные технологии в гуманитарных исследованиях. №17. Новосибирск: НГУ, 2012: С. 4-39.

Холюшкин Ю.П., Витяев Е.Е., Мартынович В.В. Web-система создания, использования и применения стратегий задач археологии. Новосибирск, 2014: с.90-92.

Холушкин Ю.П., Воронин В.Т. Сектор археологической теории и информатики: итоги десятилетия. Новосибирск. 2005. 26 с.

Холушкин Ю.П., Воронин В.Т. Интернет – новая информационная среда и технология в археологии и этнографии // Интернет – новая информационная среда исторической науки. Барнаул, 1998. С. 47–50.

Холушкин Ю.П., Воронин В.Т. Некоторые концептуальные подходы к созданию интегрированной информационной системы представления археологических и исторических данных и знаний // Информационный Бюллетень Ассоциации «История и компьютер». 1996. №17. С. 20–24.

Холушкин Ю.П., Воронин В.Т., Воробьев В.В. Информационная система по подготовке годовых научных отчетов // Информационные технологии в гуманитарных исследованиях. Выпуск 7. - Новосибирск: Редакционно-издательский Центр НГУ. 2004. С. 72-74.

Холушкин Ю.П., Воронин В.Т., Воробьев В.В., Бердников Е.В., Федоров С.А., Жилицкая Г.Ю., Грищенко А.А., Лузин А.В. Электронный каталог научной библиотеки Института археологии и этнографии СО РАН (археология и этнография) // Информационные технологии в гуманитарных исследованиях. Выпуск 6. - Новосибирск: Редакционно-издательский Центр НГУ. 2003. С. 81-85.

Холушкин Ю.П., Воронин В.Т., Ильиных М.Ю. Виртуальный музей истории и культуры народов Сибири и Дальнего Востока // Информационные технологии в гуманитарных исследованиях. Вып.9. - Новосибирск: РИЦ НГУ, 2005, с. 48-57.

Холушкин Ю.П., Воронин В.Т., Ильиных М.Ю., Соловьева С.А. Разработка архитектуры системы музейного портала "История и культура Северной Азии и Дальнего Востока" // Информационные технологии в гуманитарных исследованиях. Вып.12. Новосибирск: ЗАО РИЦ "Прайс-Курьер". 2008. С. 29-36.

Холушкин Ю.П., Воронин В.Т., Костин В.С. Концептуальные подходы к созданию on-line статистического пакета анализа археологической информации с элементами картографирования на сайте «Sibirica» // Информационные технологии в гуманитарных исследованиях. Вып. 12. Новосибирск, 2008. С. 50–53.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Болдырев А.В. Разработка пакета статистического анализа археологических данных (пользовательский интерфейс) // Информационные технологии в гуманитарных исследованиях. Вып. 10. Новосибирск, 2006. С. 72–87.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Данник Е.В. Разработка пакета статистического анализа археологических данных (методы анализа структур) / Ю.П.Холушкин, В.Т.Воронин, В.С.Костин, Е.В.Данник // Информационные технологии в гуманитарных исследованиях. - Новосибирск: ЗАО РИЦ «Прайс-Курьер», 2008. - Вып.12. - С.71-75.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Ильчук А.Ю. Разработка пакета статистического анализа археологических данных (программный интерфейс) // Информационные технологии в гуманитарных исследованиях. - Новосибирск: ЗАО РИЦ «Прайс-Курьер», 2008. - Вып.12. - С.62-65.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Корнюхин Ю.Г. О проекте пакета статистического анализа археологических данных на сайте «Sibirica» // Информационные технологии в гуманитарных исследованиях. Вып. 12. Новосибирск, 2008. С. 54–61.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Мок А.В. Разработка пакета статистического анализа археологических данных (методы анализа связей) / Ю.П.Холушкин, В.Т.Воронин, В.С.Костин, А.В.Мок // Информационные технологии в гуманитарных исследованиях. - Новосибирск: ЗАО РИЦ «Прайс-Курьер», 2008. - Вып.12. - С.66-70.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Нуртдинов А.Н., Корнюхин Ю.Г. К методике анализа данных в археологических исследованиях // Информационный Бюллетень Ассоциации «История и компьютер». 2004. №32. С. 193–194.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Ильиных М.Ю. и др. О некоторых подходах к созданию музейного портала по археологии и этнографии Сибири и Дальнего Востока // Информационные технологии в гуманитарных исследованиях. Вып. 10. Новосибирск, 2006. С. 31–42.

Холушкин Ю.П., Воронин В.Т., Костин В.С., Семенов В.М. и др. Использование технологии виртуальной реальности в секторе археологической теории и информатики ИАЭТ СО

РАН // Информационные технологии в гуманитарных исследованиях. Вып. 11. Новосибирск, 2006. С. 15–20.

Холюшкин Ю.П., Воронин В.Т., Нуртдинов А.Н. ГИС «Материальная и духовная культура народов Западной Сибири» // Информационный Бюллетень Ассоциации «История и компьютер». Специальный выпуск. Материалы IX конференции Ассоциации «История и компьютер» (Подмосковье, 22–25 апреля 2004г.). – 2004. – №32. – С. 124–125.

Холюшкин Ю.П., Воронин В.Т., Семенов В.М. Виртуальный VRML-музей Института археологии и этнографии СО РАН // Информационные технологии в гуманитарных исследованиях. Вып.9. - Новосибирск: РИЦ НГУ, 2005, с. 57-61.

Холюшкин Ю.П., Воронин В.Т., Семенов В.М. Музейные объекты ИАЭТ СО РАН в виртуальной реальности (опыт применения VRML-технологии) // Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий: материалы годовой сессии Института археологии и этнографии СО РАН. Т.ХI, Ч.2. - Новосибирск: ИАЭТ СО РАН, 2005. - С.224-227.

Холюшкин Ю.П., Гемуев И.Н., Бауло А.В., Воронин В.Т., Нуртдинов А.Н., Ильиных М.Ю. Религиозно-мифологические представления народов Западной Сибири. // Информационные технологии в гуманитарных исследованиях. Новосибирск: РИЦ НГУ, 2003. С. 73-77.

Электронный каталог научной библиотеки ИАЭТ СО РАН / Ю.П.Холюшкин, В.Т.Воронин, С.А.Федоров, Е.В.Бердников, Г.Ю.Жилицкая // Информационные технологии в гуманитарных исследованиях. – Новосибирск: НГУ, 2002. Вып.3. С.15-20.

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: РИЦ НГУ, 2002. С. 16-20.

Холюшкин Ю.П., Гражданников Е.Д. Системная классификация археологической науки (элементарное введение в археологическое науковедение). Новосибирск: Изд-во ИДМИ Минобразования. Новосибирск. 2000. 58 с.

Холюшкин Ю.П., Костин В.С. Статистический анализ взаимоцитирования «новых археологов» // Информационные технологии в гуманитарных исследованиях. Вып. 15. Новосибирск, 2010. С. 66–71.

Холюшкин Ю.П., Костин В.С., Есина И.В. Разработка Web-пакета статистического анализа археологических данных (дискриминантный анализ) // Информационные технологии в гуманитарных исследованиях. Новосибирск: Прайс-Курьер, 2008. С. 75-84.

Холюшкин Ю.П., Костин В.С., Федотова Д.В. Разработка Web-пакета статистического анализа археологических данных (регрессионный анализ) // Информационные технологии в гуманитарных исследованиях. - Новосибирск: ЗАО РИЦ «Прайс-Курьер», 2010. - Вып.15. - С.83-89.

Холюшкин Ю.П., Рыбинцев Р.В. Новые ресурсы на сайте Сибирика /// Информационные технологии в гуманитарных исследованиях. Новосибирск, 2012. Вып. 17. С. 119–120.

Холюшкин Ю.П., Холюшкина В.А. Об оценке классификационной значимости признаков галечных орудий Самаркандской стоянки // Новое в археологии Сибири и Дальнего Востока. - Новосибирск: Наука, 1979. С.40-51.

Шокин Ю.И. Анализ связей между сайтами институтов Сибирского отделения РАН / Шокин Ю.И., Клименко О.А., Петров И.С. // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2011. Т. 9. № 4. С. 12-17. URL:

<https://cyberleninka.ru/article/n/analiz-svyazey-mezhdu-saytami-institutov-sibirskogo-otdeleniya-ran>

Шокин Ю.И. Исследование научного веб-пространства Сибирского отделения Российской Академии наук / Шокин Ю.И., Веснин А.Ю., Добрынин А.А., Клименко О.А., Рычкова Е.В., Петров И.С. // Вычислительные технологии. 2012. Т. 17. № 6. С. 85-98. URL: <https://cyberleninka.ru/article/n/issledovanie-nauchnogo-veb-prostranstva-sibirskogo-otdeleniya-rossiyskoy-akademii-nauk>

Шрайберг Я.Л. Формирование единого пространства знаний на базе сетевой информационной инфраструктуры в условиях становления и развития современной цифровой экономики: ежегод. докл. 4-го междунар. профес. форума «Крым-2018» / Шрайберг Я.Л. // Научные и технические библиотеки. 2018. № 9. С. 3-75.

Щапова Ю.Л. Археологическая эпоха: хронология, периодизация, теория, модель. М.: Комкнига, 2005,192с.

Binford L. R. An archaeological perspective. – NY, London, 1972: 464 p.

- Brezillon M.** La denomination des objets de pierre taillee. // *Materiaux pour un vocabulaire de prehistoriens de langue francaise.* – Paris, 1968: 413 p.
- Efron B.** Better bootstrap confidence intervals.//*J. Amer. Statist. Ass.*, 81, 1986
- Ford J.A.** The type concept revised // *American Anthropologist*. Vol. 56, 1954: p. 42-54.
- Clarke D.L.** Analytical archaeology. – L: Methuen, 1968: 684 p.
- Clarke D.L. (ed).** Models in archaeology. – L: Methuen, 1972: 1055 p.
- Doran, J.E. & Hodson, F.R.** Mathematics and Computer in archaeology.– Edinburg, 1975: 381 p.
- Fernandez-Lopez M., Gomez-Perez A., Pazos Sierra J.** Building a Chemical Ontology Using Methontology and the Ontology Design Environment // *IEEE Intelligent Systems*, 14(1). 1999. P. 37– 46.
- Ford J.A.** A quantitative method for derivinf cultural chronology. – Washigton, 1962.
- Gruninger M., Fox M.S.** Methodology for the Design and Evaluation of Ontologies // *Proceedings of IJCAI 1995 Workshop on Basic Ontological Issues in Knowledge Sharing*. 1995.
- Guariano N., Giaretta P.** Ontologies and Knowledge Bases. Towards a Terminological Clarification // *Towards Very Large Knowledge Bases / Ed. by N. J. I. Mars.*| Amsterdam: IOS Press. 1995.
- Kintigh K.** Measuring archaeological diversity by comparison with Simulated assemblages // *American Antiquity*, vol. 49, 1984: p. 44-54.
- Maness J.** Library 2.0 Theory: Web 2.0 and Its Implications for Libraries // *Webology*. 2006. Vol.3, № 2. URL: <http://www.webology.org/2006/v3n2/a25.html> ; https://digitalcommons.du.edu/libraries_facpub/13/
- PSPP - Summary** // <http://savannah.gnu.org/projects/pspp/>
- Ringrouse T.,** Bootstraping and correspondenting analysis in archaeology // *Journal of Archaeological Sciences*. Vol. 19, 1992: p.615-629.
- Spoulding A.S.** Statistical techniques for discovery of artifact types // *American antiquity*, 1953, v.18, № 4: p. 305–313.
- Staab S., Schunurr H-P., Studer R., Sure Y.** Knowledge processes and ontologies // *IEEE Intelligent Systems*, Special Issue on Knowledge Management 16(1) 2001. P. 26– 34.
- Udartseva O.M.** An overview of webometrics in libraries: history and modern development tendencies // *Scientific and Technical Information Processing*. 2018. V. 45. № 3. P. 174-181.
- Uschold M., Gruninger M.** Ontologies: Principles, Methods and Applications // *Knowledge Engineering Review* 11(2). 1996.
- Zagorulko Yu., Borovikova O., Bulgakov S., Sidorova E.** Ontology-based approach to development of adjustable knowledge internet portal for support of research activity // *Bull. of NCC. Ser.: Comput. Sci.* 2005. Is. P. 23: 45–56.

Подписано в печать 25.12.2018.
Формат 60х84/8. Усл.печ.лист 15.
Печать цифровая.
Тираж 100 экз. Заказ № 236.
Отпечатано в ООО «Офсет-ТМ»
630117, Новосибирск, ул. Арбузова, 4/27
т.:(383)332-82-32, e-mail: ofsetn@yandex.ru