

Министерство образования Республики Беларусь
Учреждение образования «Витебский государственный
университет имени П.М. Машерова»
Кафедра информационных технологий и управления бизнесом

ИСТОРИЧЕСКАЯ ИНФОРМАТИКА

Курс лекций

*Витебск
ВГУ имени П.М. Машерова
2024*

УДК 930:004(075.8)
ББК 63с51я73+16.3я73+32.972я73
И90

Печатается по решению научно-методического совета учреждения образования «Витебский государственный университет имени П.М. Машерова». Протокол № 2 от 20.12.2023.

Составитель: старший преподаватель кафедры информационных технологий и управления бизнесом ВГУ имени П.М. Машерова, магистр педагогических наук **Т.А. Денисенко**

Р е ц е н з е н т :
заведующий кафедрой истории и культурного наследия
ВГУ имени П.М. Машерова, кандидат исторических наук,
доцент *А.Н. Дулов*

Историческая информатика : курс лекций / сост. Т.А. Денисенко. – Витебск : ВГУ имени П.М. Машерова, 2024. – 47 с.

Данное издание содержит теоретические материалы по дисциплине «Историческая информатика» и предназначено для студентов, обучающихся по специальностям исторического образования. Курс лекций позволит оптимизировать процесс усвоения теоретических знаний при подготовке к лабораторным занятиям, а также всем формам контроля знаний.

УДК 930:004(075.8)
ББК 63с51я73+16.3я73+32.972я73

© ВГУ имени П.М. Машерова, 2024

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
ЛЕКЦИЯ 1. Предмет и задачи курса. История развития информационных технологий и исторической информатики	5
ЛЕКЦИЯ 2. Методология исторической информатики. «Цифровой поворот» в исторической науке	14
ЛЕКЦИЯ 3. Электронный текст исторических исследований: создание и хранение	27
ЛЕКЦИЯ 4. Базы данных в исторических исследованиях	31
ЛИТЕРАТУРА	45

ВВЕДЕНИЕ

Учебная дисциплина «Историческая информатика» занимает важное место в процессе подготовки будущих учителей истории в силу своей высокой практикоориентированности.

Целью учебной дисциплины является освоение студентами путей и способов применения современных информационных технологий в научно-исследовательской и педагогической деятельности учителя истории.

Теоретической основой курса лекций по исторической информатике является современная концепция информации (в том числе и социальная информация) и теоретическое источниковедение, а прикладной – информационные и компьютерные технологии.

Направления исторической информатики включают электронную публикацию исторических источников, разработку исторических баз данных и тематических интернет-ресурсов, исторических геоинформационных систем, виртуальных 3d реконструкций объектов культурного наследия, использование компьютеризованных методов и технологий анализа статистических, структурированных, текстовых, визуальных и других источников, компьютерное моделирование исторических процессов, а также создание специализированного программного обеспечения и применение информационных технологий в историческом образовании (включая обучение в режиме онлайн).

Задачи:

1) определить особенности исторической информатики как междисциплинарной области исторических исследований и важность ее использования в исторической науке и методике преподавания истории;

2) выявить возможности применения различных информационно-коммуникационных технологий и глобальной сети Интернет в области исторических исследований и методике преподавания истории;

3) овладеть соответствующими компетенциями по использованию компьютерных технологий и глобальной сети Интернет при проведении исторических исследований и преподавании истории;

4) усвоить компетенции в области использования современных информационных технологий для будущей профессиональной деятельности учителя истории с целью повышения эффективности педагогической деятельности.

В учебное издание включены лекционные материалы и список литературы. Может использоваться преподавателями, студентами, а также всеми, кто интересуется концепциями, методами и технологиями междисциплинарных исторических исследований.

ЛЕКЦИЯ 1. Предмет и задачи курса.

История развития информационных технологий и исторической информатики

1. Историческая информатика как междисциплинарная область исторических исследований и учебная дисциплина.
2. История развития информационных технологий.
3. Этапы развития исторической информатики.
4. Международная ассоциация «History and Computing».

1. Историческая информатика как междисциплинарная область исторических исследований

Историческая информатика – междисциплинарная область исследований исторической науки и прикладной информатики. В основе исторической информатики лежит совокупность теоретических и прикладных знаний, необходимых для создания, обработки и анализа цифровых версий исторических источников и электронных ресурсов.

Термин «историческая информатика» был введён в начале 1990-х гг. Л.И. Бородкиным. Основные концепции и подходы исторической информатики разрабатываются на кафедре исторической информатики исторического факультета МГУ имени М.В. Ломоносова под эгидой ассоциации «История и компьютер», созданной в 1992 в качестве российской ветви Международной ассоциации «History and Computing».

Историческая информатика изучает и развивает методы и процессы сбора, хранения, обработки, передачи, поиска, анализа и оценки исторической информации с применением компьютерных технологий. Теоретической основой исторической информатики является современная информатическая теория и теоретическое источниковедение, а прикладной основой – информационно-коммуникационные и компьютерные технологии.

Направления исторической информатики включают электронную публикацию исторических источников, разработку исторических баз данных и тематических интернет-ресурсов, исторических геоинформационных систем (ГИС), виртуальных трёхмерных (3D) реконструкций объектов культурного наследия, использование компьютеризованных методов и технологий анализа статистических, структурированных, текстовых, изобразительных, аудиовизуальных и других источников, компьютерное моделирование исторических процессов, а также создание специализированного программного обеспечения и применение информационных технологий в историческом образовании. В исторической информатике принято выделять ресурсную и аналитическую компоненты.

Предмет и задачи курса

Историческая информатика как учебная дисциплина имеет целью ознакомление студентов-историков с основными понятиями информатики

и математики, а также со спецификой их использования в исторических исследованиях. Рассматриваются основные аспекты исторической информатики: предмет, история становления и основные тенденции развития. Дается понятие о теории информации, рассматриваются семиотические аспекты информации. Значительное внимание уделяется характеристике основных этапов компьютеризованного исторического исследования.

2. История развития информационных технологий

Применение «ручных» информационных технологий (до второй половины XIX в.). Их инструментом в основном являлись канцелярские принадлежности и средства почтовой связи, обеспечивавшие пересылку писем, пакетов и бандеролей. Основная цель технологии – предоставление информации в нужной форме.

Использование «механических» информационных технологий (с конца XIX в.). В этот период к названному инструментарию добавляются средства оргтехники (пишущие машинки, телеграф, телефон, магнитофоны и диктофоны). Информационные коммуникации поддерживаются с помощью более совершенных средств доставки почты. Основная цель технологии – представление информации в нужной форме более удобными средствами.

Создание первых ЭВМ (с конца 40-х гг. XX в.). Появление «электрических» технологий, инструментом которых составляют: большие ЭВМ и программное обеспечение к ним, электрические пишущие машинки, настольные копиры, портативные диктофоны и т.п. В этот период развиваются и совершенствуются существующие информационные коммуникации, появляются телевидение, системы передачи данных по воздушным и безвоздушным линиям связи. Были распространены системы пакетной обработки. Акцент в информационной технологии начинает перемещаться с формы представления информации на формирование её содержания.

С появлением первых электронных компьютеров, историки начали осознавать потенциал этих машин для обработки больших объемов информации. В 1960-х годах начали появляться специализированные программы для обработки текстов и анализа данных, что способствовало развитию количественных методов в истории.

Изобретение микропроцессорных технологий и появление персонального компьютера (ПК) (70-е гг. XX вв.). Появление «электронных» технологий. Их основной инструментарий – большие ЭВМ с создаваемыми на их базе автоматизированными системами управления (АСУ) и информационно-поисковыми системами (ИПС). Очень популярны сетевые и иерархические модели данных. Появляются факсимильные средства передачи данных, компьютерные вычислительные и информационные коммуникации: локальные и междугородные вычислительные сети. Технологии ещё более ориентируются на формирование содержательной стороны информации

для управленческой среды различных сфер общественной жизни, особенно на организацию аналитической работы.

Расширение доступности компьютерных технологий для широкого потребителя (с середины 80-х гг. XX в. и до конца 1990-х гг.). Основным инструментом в этот период становится персональный компьютер. Для него создаётся множество различных программных продуктов и периферийных устройств. Появляются автоматизированные рабочие места (АРМ), в том числе локальные (на одном персональном компьютере) и системы поддержки принятия решений. Широкое распространение получили реляционные модели. Информационные коммуникации называют телекоммуникациями. Они включают локальные, региональные глобальные (международные) и иные компьютерные сети.

В этот период появились первые специализированные программы для историков, позволяющие анализировать и визуализировать исторические данные. В 1980-х годах началось создание цифровых архивов и библиотек, которые сегодня являются неотъемлемой частью исторической информатики.

С появлением доступных персональных компьютеров, новых устройств хранения информации и универсальных, достаточно простых программ для распространенных приложений связано оформление исторической информатики в самостоятельное научное направление.

Массовое внедрение информационных технологий во все сферы деятельности общества (с начала 2000-х гг.). Период формирования информационных обществ. Он характеризуется глобализацией информационных технологий и связанным с ними применением интернета и суперкомпьютеров. В области телекоммуникаций всё чаще используются оптические проводные и беспроводные системы, а также иные беспроводные коммуникации, развиваются облачные и нейросетевые технологии.

Появление веб-технологий открыло новые возможности для распространения и анализа исторической информации. Проекты по оцифровке и созданию цифровых архивов стали массовыми, и многие учреждения начали активно публиковать свои коллекции в интернете. В это время также начали развиваться методы визуализации данных, такие как картографирование исторических событий и создание интерактивных хронологий. Эти инструменты позволяют историкам более эффективно представлять свои исследования и делать их доступными для широкой аудитории.

Современные технологии, такие как машинное обучение и анализ больших данных, начали находить применение в исторических исследованиях. Эти методы позволяют обрабатывать огромные объемы информации и выявлять скрытые закономерности в исторических данных.

3. Этапы развития исторической информатики

Эра больших ЭВМ: 60-е – середина 80-х годов. Первые публикации советских историков, посвященные применению ЭВМ в исторических

исследованиях, появились в начале 1960-х годов. Первые опыты по применению ЭВМ были связаны с обработкой локальных историко-статистических данных по социально-экономической истории страны, с созданием программ поиска и идентификации семей по подворным описаниям крестьянских хозяйств. В конце 60-х годов при Отделении истории АН СССР была организована Комиссия по применению математических методов и ЭВМ в исторических исследованиях.

В 70-е годы спектр применения ЭВМ в исторических исследованиях существенно расширился. Этому способствовало, прежде всего, становление научного направления, получившего известность как «Применение количественных методов в исторических исследованиях» (более позднее название – квантитативная история). В начале 70-х годов в Институте истории СССР АН СССР была открыта лаборатория по применению математических методов и ЭВМ в исторических исследованиях, в конце 70-х такая же группа, позднее ставшая лабораторией, появилась в структуре кафедры источниковедения на историческом факультете МГУ; начались работы в этом направлении и на историческом факультете ЛГУ. В 80-е годы небольшие лаборатории или группы этого профиля были организованы в Московском историко-архивном институте (ныне – Российский государственный гуманитарный университет – РГГУ) и ряде других вузов, включая БГУ. В эти годы было положено начало формированию сообщества ученых, применяющих математические методы и ЭВМ в исторических исследованиях. Процесс стимулировался активизацией контактов специалистов в рассматриваемой области. С 1979 г. они регулярно собирались на заседаниях всесоюзного семинара «Количественные методы в исторических исследованиях», организованного Комиссией и историческим факультетом МГУ. Также проходили школы-семинары по новым методам исторических исследований, поддерживаемые МГУ, где собирались молодые ученые.

Работа над проблемами, возникающими на стыке «источник – ЭВМ» и составляющими основу исторической информатики, была все эти годы органичной частью исследований по квантитативной истории. Особенность развития этих работ в СССР определялась, в частности, тем, что в состав большинства лабораторий были приняты специалисты по прикладной математике и кибернетике – эта инициатива принадлежала руководству Комиссии. Взаимодействие между этими специалистами и специалистами по квантитативной истории обеспечивали достаточно высокий уровень использования сложных и эффективных математико-статистических методов и моделей – даже в условиях нехватки передовой вычислительной техники в 70–80-е годы.

До середины 80-х годов советские историки использовали в основном большие ЭВМ серии БЭСМ. Главная трудность заключалась в том, что доступные для историков ЭВМ находились в основном в крупных вычислительных центрах, что исключало возможность оперативной работы.

Ситуация несколько улучшилась лишь в начале 80-х годов, когда на историческом факультете МГУ была установлена пара дисплеев, соединенных с большой удаленной ЭВМ.

Вся обработка источников проводилась с помощью программ, созданных в упомянутых выше лабораториях, что потребовало глубокого понимания методических проблем со стороны историков и содержательных проблем со стороны математиков-программистов.

Можно выделить несколько типов программ, создаваемых историками в данный период:

- программы, ориентированные на создание и сопровождение архивов машиночитаемых данных (МЧД). В первой публикации, посвященной проблемам разработки архивов МЧД по данным массовых исторических источников, обобщался опыт, накопленный историками МГУ. К началу 80-х годов этот архив МЧД на магнитных лентах содержал около 50 массивов, в том числе данные первой переписи населения России 1897 г., данные сельскохозяйственной переписи 1917 г., тысячи анкет делегатов съездов Советов 1917–1936 гг. Существенную роль при использовании таких архивов МЧД играли программы, реализующие функции информационно-поисковых систем;

- программы, реализующие историко-демографические методики «восстановления истории семей» (ВИС). Это направление развивалось преимущественно эстонским историком Х. Палли, первая публикация которого в 1971 г. содержала описание программы, сопоставляющей сведения о крестьянских семьях из различных источников (метрических книг, ревизских сказок) в условиях вариации прозвищ – фамилий у эстонских крестьян в XVII и XVIII вв. не было;

- программы статистического анализа данных исторических источников. В течение 70–80-х годов в лабораториях академических институтов и исторических факультетов были созданы десятки программ, реализующих большинство математико-статистических методов. Две причины стимулировали такие разработки: специфика структуры данных исторических источников и задач их анализа и трудности доступа к использованию «продвинутых» статистических пакетов типа SPSS;

- программные комплексы выросли из крупных исследовательских задач-проектов, связанных с обработкой массовых статистических источников. Так, комплексы статистических программ различной ориентации были созданы в результате многолетних исследований писцовых книг XVII в., материалов крестьянской реформы в 1861 г., массовых источников по аграрному развитию России конца XIX – начала XX в., массовых источников по истории советского рабочего класса и т.д.;

- программное обеспечение для работы со средневековыми текстами. Здесь речь идет об исследованиях источниковедческого характера, посвященных атрибуции определению авторства – средневековых текстов нарративного (описательного) характера и изучению истории текста

памятников древнерусской письменности, имеющих многовековую традицию и множество рукописных копий (списков). Разработанные программы позволяли: построить генеалогическое древо списков, исходя из сопоставления разночтений; определить особенности авторского стиля средневековых текстов на основе структуры парных встречаемостей грамматических классов и использовать их в задачах атрибуции;

▪ программное обеспечение задач моделирования исторических процессов и явлений. Проблемы моделирования вызвали, пожалуй, наиболее острую дискуссию среди советских историков-клиометристов в 70–80-е годы. Ряд программ, реализующих различные имитационные и «аналитические» модели, был разработан в 70–80-е годы для моделирования Пелопонесских войн, динамики древнейших человеческих общин в условиях палеолита, процесса расслоения феодально-зависимых крестьян в Византии, динамики социальных революций, процесса иностранных инвестиций в России начала XX в., социальной динамики в период нэпа и т.д. Нередко создание программ для моделирования исторических процессов было связано с использованием сложного математического аппарата – теории игр, теории дифференциальных уравнений, теории марковских цепей.

Середина 80-х была отмечена резким повышением внимания в стране к проблемам информатизации общества, особенно в системе образования. В вузах был введен обязательный курс информатики, а преподаватели массово осваивали ЭВМ. На историческом факультете МГУ программа компьютеризации была реализована своими специалистами, имеющими достаточный опыт применения ЭВМ в исторических исследованиях.

В результате ускоренного прохождения курса компьютеризации в ряде университетов возникли полезные контакты между историками и специалистами по информатике и кибернетике; в последующие годы это привело к распространению новых методов исторических исследований. Часть историков смогла оценить потенциальные возможности применения новых информационных технологий в работе с историческими источниками. В системе образования возникла настоятельная необходимость приобретения нужного количества вычислительной техники нового поколения.

Микрокомпьютерная революция: конец 80-х – начало 90-х годов. Первые персональные компьютеры в распоряжении советских историков появились в начале 1988 г., лет на пять позднее, чем у коллег из стран Западной Европы и США. Это были три отечественных IBM-совместимых микрокомпьютера «Нейрон» с весьма скромными характеристиками, которые были приобретены историческим факультетом МГУ и в течение полутора лет являлись материальной базой для количественных исследований и учебного процесса. В 1990 г. эти машины были заменены шестью ПК класса IBM/PC AT; к этому времени машины такого уровня появились в Институте истории СССР АН СССР, а также у историков некоторых

других институтов и вузов страны. Открывались новые возможности для развития компьютерных технологий в исторических исследованиях и обучении историческим дисциплинам.

Реализация этих возможностей наиболее успешно проходила на историческом факультете МГУ, где в 1992 г. насчитывалось 25, а в 1996 г. – почти 60 ПК, что позволило установить их на многих кафедрах, существенно улучшить материальную базу лаборатории исторической информатики и ряда других лабораторий факультета, открыть компьютерные классы для проведения занятий по исторической информатике и самостоятельной работы – включая навигацию в Интернете – студентов и сотрудников факультета.

Различные модели курса исторической информатики развиваются также в РГГУ и ряде других университетов, включая БГУ.

Появление персональных компьютеров новых моделей заметно активизировало разработку программ, предназначенных для контроля знаний и обучения историческим дисциплинам.

Другой важный аспект микрокомпьютерной революции был связан с возрастанием интереса к созданию баз данных по материалам исторических источников. Программное обеспечение персональных компьютеров дало в руки историков весьма удобный и доступный каждому инструмент для хранения данных источника и информационного поиска. Исходя из необходимости координации и определенной унификации деятельности по разработке баз данных, представители исторического факультета МГУ, Института истории СССР АН СССР и МГИАИ выступили в 1989 г. с инициативой создания Банка (машиночитаемых) данных по российской истории. Эта идея получила финансовую поддержку Отделения истории АН СССР (а позже Российского гуманитарного научного фонда РГНФ), и с 1990 г. инициативная группа приступила к практической деятельности. Описания конкретных баз данных приводятся в выпусках Информационного бюллетеня «История и компьютер». Этот бюллетень, созданный в 1990 г. и выходящий в свет ежеквартально, стал основным каналом оперативного обмена информацией в среде специалистов в области исторической информатики и квантитативной истории

Появление компьютеров нового поколения стимулировало разработку историками не только баз данных, но и баз знаний, применение экспертных систем и других систем искусственного интеллекта. Активизировались работы по компьютерному анализу текстовых источников и изображений, ведущиеся в ряде лабораторий страны. Новые компьютерные технологии, основанные на использовании сканеров, лазерных дисков, мультимедийных технологий, существенно расширили возможности хранения и анализа изобразительных источников. В середине 90-х годов российские историки получили доступ к информационным ресурсам глобальной сети Интернет (в эту сеть сетей подключились компьютеры исторических факультетов МГУ и других учреждений).

Развитие исторической информатики в СССР вплоть до 1992 г. не пересекалось с развитием этого направления за рубежом.

Специалисты по исторической информатике ряда стран Западной Европы еще к середине 80-х годов пришли к выводу о необходимости институционализации международного научного сообщества, складывающегося в этом активно развивающемся направлении исследований.

К середине 80-х годов тенденции развития квантитативной истории сделали очевидными следующие обстоятельства: большинство статистических методов (включая наиболее сложные) историками-клиометристами освоены, и дальнейшее расширение их методических средств требует новых подходов; практически весь арсенал математической статистики реализован в стандартных статистических пакетах, и при наличии персонального компьютера историк-исследователь в принципе может обрабатывать свои данные самостоятельно без посредника-программиста; повышение требований к уровню аргументации результатов исторического исследования акцентирует вопрос о корректности использования данных источника; стремительный рост количества баз данных неизбежно ставит вопросы их стандартизации, систематизации, распространения, устранения возможного дублирования; эти вопросы должны решаться в рамках соответствующих банков данных; существующие стандартные пакеты программ пригодны для хранения, поиска и анализа далеко не всех исторических источников; есть немало источников, обладающих спецификой, которая не могла быть учтена создателями стандартных пакетов; микрокомпьютерная революция открыла возможности для использования историками новых информационных технологий, не связанных с вычислениями и обработкой статистических данных: оптический ввод текста с помощью сканера, компьютерное картографирование и анализ изображений, использование систем искусственного интеллекта, глобальных сетей.

4. Международная ассоциация «History and Computing»

В 1986 была создана Международная ассоциация «History and Computing» (АНС), которая является одной из самых представительных профессиональных ассоциаций историков и самой активной из них. С 1986 г. АНС проводит ежегодные международные конференции. С 1992 г. в состав АНС вошла Ассоциация стран СНГ «История и компьютер» (АИК), объединяющая специалистов по исторической информатике из России и стран СНГ. С 1989 г. АНС выпускает международный журнал «History and Computing».

К середине 90-х годов выявились дискуссии о содержании и приоритетах развития исторической информатики, в которых обсуждался вопрос о том, как далеко за пределы компьютерного источниковедения простирается

область исторической информатики. Компьютерное источниковедение обычно понимается как совокупность методов и технологий создания машиночитаемых исторических источников, а традиционные вопросы критики источника рассматривает классическое источниковедение. С самого начала компьютерное источниковедение рассматривалось как ядро исторической информатики; именно проблемы адекватной формализации и репрезентации информации источника, создания баз данных, учитывающих специфику исторических источников, были в центре внимания пионеров исторической информатики. В то же время на конференциях АНС, в ее публикациях всегда присутствовала и другая важная компонента – аналитическая, связанная с использованием компьютера для анализа данных машиночитаемого источника, получения содержательно значимых результатов. Эта проблема определяется в значительной мере различиями в оценках возможностей историко-ориентированного, с одной стороны, и проблемно-ориентированного подхода – с другой. Историческая информатика должна органично сочетать обе компоненты – источниковедческую и аналитическую.

В этом контексте естественным представляется следующее определение исторической информатики – это научная дисциплина, изучающая закономерности процесса информатизации исторической науки и образования; в основе исторической информатики лежит совокупность теоретических и прикладных знаний, необходимых для создания и использования в исследовательской практике электронных версий исторических источников всех видов.

Теоретической основой исторической информатики является современная концепция информации, включая социальную информацию и теоретическое источниковедение, а прикладной – информационные (компьютерные) технологии.

Область интересов исторической информатики включает разработку общих подходов к применению информационных технологий в исторических исследованиях, в частности – специализированного программного обеспечения; создание исторических баз и банков данных (знаний); применение информационных технологий представления данных и анализа структурированных, текстовых, изобразительных и других источников; компьютерное моделирование исторических процессов; использование информационных сетей (Интернета и др.); развитие и применение средств мультимедиа и других новых направлений информатизации исторической науки; а также применение информационных технологий в историческом образовании.

ЛЕКЦИЯ 2. Методология исторической информатики. «Цифровой поворот» в исторической науке

1. Основные понятия информатики.
2. Специфика применения Всемирной сети при проведении исторических исследований.
3. Обзор языков программирования и возможности их применения в исторической науке.

Место исторической информатики в системе исторических наук и ее связь с другими областями научного знания

Историческая информатика связана с такими историческими науками, как источниковедение, историография, археология, вспомогательные исторические дисциплины. Также историческая информатика связана с другими областями научного знания: информатика, математическая статистика, архивоведение, музееведение, библиотековедение, геоинформатика.

Компьютерные технологии позволяют оцифровывать исторические материалы, проводить количественный анализ и выявлять закономерности в больших объемах исторических данных, визуализировать и анализировать пространственные данные, создавать и поддерживать цифровые архивы и базы данных, связывать результаты исследований, создавать образовательные ресурсы и также соблюдать этику использования данных, включая вопросы конфиденциальности, авторских прав и сохранения культурного наследия.

Компьютерные технологии сохраняют исторические материалы, упрощают доступ к ним и их исследование, делают их доступными для анализа и исследования в цифровом формате.

Интернет предоставляет историкам доступ к огромному количеству цифровых архивов, библиотек и баз данных и обеспечивает взаимодействие между историками, исследователями и учреждениями по всему миру.

1. Предметная область информатики

Предметом информатики является информационный ресурс, как единство знания и информации – его сущность, законы функционирования, механизмы взаимодействия с другими ресурсами общества и воздействие на социальный прогресс.

Основные понятия информатики

Информация – сведения о чем-либо. История человечества – это история того, как человек стремился получить информацию об окружающем его мире, сохранить, использовать, передать ее следующим поколениям. Наскальная живопись, клинопись, устная речь, музыкальные звуки, нотные знаки для их записи, алфавит, телеграф, радио, телефон, телевидение,

компьютеры – вот лишь некоторые звенья цепи попыток совершенствовать способы получения, сохранения, обработки и передачи информации.

Применительно к компьютерной обработке данных под информацией понимают некоторую последовательность символических обозначений (букв, цифр, закодированных графических образов и звуков и т.п.), несущую смысловую нагрузку и представленную в понятном компьютеру виду.

Информационные процессы – действия, выполняемые с информацией: получение, хранение, обработка, передача.

Информационные технологии – технологии накопления, обработки и передачи информации с использованием определенных технических средств.

Информационные ресурсы – это идеи человечества и указания по их реализации, накопленные в форме, позволяющей воспроизводить их (книги, статьи, патенты, диссертации, научно-исследовательская и опытно-конструкторская документация, технические переводы, данные о передовом производственном опыте и др).

Понятие «цифровая история» (Digital history). Цифровая история – раздел цифровых гуманитарных наук (Digital Humanities), изучающий использование компьютерных технологий и цифровых медиа для исторического анализа, исследований и визуализации данных. Два основных применения данного направления – знакомство интернет-аудитории с цифровыми архивами, интерактивными картами, хрониками событий (то есть углубление исторических знаний пользователей Сети) и создание новых исследовательских инструментов для учёных-историков, помогающих им в развитии истории как науки. Историческая информатика, в отличие от «цифровой истории», не рассматривает оцифровку источников как необходимое условие использования информационных технологий в историческом исследовании.

Компьютерное источниковедение – это совокупность методов и технологий создания машиночитаемых исторических источников. Машиночитаемые источники – это источники, переведенные в электронную форму. Однако, поскольку в машиночитаемую часть переводится только часть информации, потенциально содержащейся в источнике, то более корректным и часто употребляемым является термин «машиночитаемые данные» (МЧД). Вместе с тем машиночитаемые версии источников могут рассматриваться и как новые источники – машиночитаемые источники.

История появления термина «компьютерное источниковедение»

Термин «компьютерное источниковедение» (до сих пор не имеющий однозначного определения) появился в отечественной исторической информатике в 1993 г., в связи с дискуссиями этого периода по теоретическим основаниям нового направления. В статье, где этот термин был предложен впервые, компьютерное источниковедение фактически отождествляется с исторической информатикой, понимаемой как «дисциплина, ориентированная, прежде всего, на разработку методов анализа исторических источников

с помощью ЭВМ, основных принципов их формализации и представления в машиночитаемой форме, а также общих подходов использования информационных и компьютерных технологий в исторических исследованиях» (при этом машиночитаемыми историческими источниками авторы считали исключительно базы данных). С другой стороны, некоторые авторы считают историческую информатику ничем иным, как «новым» источниковедением, или источниковедением XX–XXI вв.

В историографии встречаются и определения компьютерного источниковедения как «дважды междисциплинарной» области: «компьютерное источниковедение – это новое научное направление, возникшее под влиянием теории информации на стыке исторической информатики и источниковедения. Оно изучает как традиционные проблемы источниковедения средствами новых компьютерных технологий, так и источниковедческие проблемы, порождаемые теорией информации и развитием информационных технологий». К новым проблемам в данном случае отнесены «оценка информационного потенциала комплекса массовых источников, источниковедческие аспекты построения модели исторического источника, создание баз данных, «компьютерная археография», научное описание баз и файлов машиночитаемых данных, вторичное использование машиночитаемой информации и т.п.».

В отечественном учебнике 1996 года по исторической информатике компьютерное источниковедение определяется как совокупность методов и технологий создания машиночитаемых исторических источников, баз данных, формализации и репрезентации информации источника, но при этом подчеркивается важность аналитической компоненты для анализа машиночитаемых данных и получения содержательных результатов.

На основе идеи об информационном потенциале исторических источников и принципа источник-ориентированного подхода возникла специализированная концепция компьютерного источниковедения, ориентированная на задачи электронных публикаций и создание информационной структуры для презентации и анализа (источниковедческой критики) источников в семантических сетях. Она позволяет рассматривать компьютерное источниковедение как самостоятельный раздел классического источниковедения, в центре внимания которого находятся методические и технические аспекты источниковедческой критики на основе компьютерных технологий.

Сфера применения методов компьютерного источниковедения (источниковедческая эвристика, методико-аналитическое источниковедение)

Практика работы исследователя истории включает в себя поиск источников. Целью поиска источников является обеспечение исследования информацией, которая необходима для получения совокупности

исторических фактов. Иногда приходится приложить большие усилия, пересмотреть немало источников и исследований, побывать во многих библиотеках, архивах и музеях, чтобы уточнить какую-нибудь дату или расшифровать сомнительный инициал.

Поиск можно разделить на выявление и отбор источников. Выявление источников – это выяснение с наиболее возможной полнотой того, где и какие существуют источники, которые относятся к теме поиска, а также учет этих источников. Отбором называется определение того, какие источники следует изучать, чтобы получить оптимальный объем источниковой информации.

Выявление и отбор необходимых в исследовательской практике источников составляют раздел источниковедения – источниковедческую эвристику. В источниковедческой эвристике различают эвристику библиографическую и архивную.

Библиографическую эвристику можно охарактеризовать как теорию и методику библиографических разысканий, как совокупность теоретических и практических знаний, относящихся к приемам обнаружения какого-либо тиражированного книжного материала или отдельных его элементов. Архивную эвристику можно определить как теорию и методику поисков неопубликованных материалов по данной теме.

В свете этих характеристик можно рассматривать источниковедческую эвристику как раздел источниковедения, посвященный теории и методике сбора и обнаружения опубликованных и неопубликованных источников с учетом закономерностей их возникновения, существования и хранения.

Методико-аналитическое источниковедение включает в себя методику и технику изучения (критики) исторических источников.

Специфика компьютерного источниковедения и его влияние на работу историка

Специфика компьютерного источниковедения обусловлена методами и техникой изучения исторических источников на основе компьютерных технологий. Прикладной характер, технологическая (инструментальная) направленность компьютерного источниковедения в определенной степени лишает его теоретической и методологической самостоятельности: можно утверждать, что в настоящее время теоретическая и методологическая основа компьютерного источниковедения – традиционное теоретическое источниковедение, разрабатывающее основные принципы источниковедческого познания, теорию исторического источника, а также теорию познания источника. Однако не исключено, что в перспективе компьютерное источниковедение сможет внести свой весомый вклад в дальнейшую разработку учения о методах исторического источниковедения и теорию исторического источника в контексте более широкого учения об информации, тем более что первые шаги в этом направлении уже делаются.

Компьютерные технологии при работе с историческими источниками (репрезентация источников, их аналитическая и синтетическая критика)

Репрезентация источников в электронной форме позволяет эффективно использовать компьютерные технологии и методы и является приоритетной задачей компьютерного источниковедения для изучения исторических источников и подготовки электронных публикаций.

В источниковедческой практике давно сложился определенный алгоритм выделения (репрезентации) информации об исторических источниках, при этом каждый источник фактически рассматривается как многоуровневый объект. Прежде всего предоставляется «внетекстовая» информация об источнике – исторических условиях и обстоятельствах его возникновения, об авторе, месте хранения, о палеографических особенностях (внешней форме) источника. Далее выделяется собственно источник-информация, при этом происходит как бы многоуровневая репрезентация его информации: сначала – информация в рамках крупных логических фрагментов, отражающих внутреннюю форму (структуру) источника, затем – свидетельства источников о конкретных аспектах исторической реальности. Специально выделяется также информация о стилистических, лингвистических и других особенностях текстов. Выделение данных информационных аспектов является основой для интерпретации информации как отдельного источника, так и их совокупностей. В соответствии с алгоритмом репрезентации информации в традиционном источниковедении, в компьютерном источниковедении должна соблюдаться стадильность в представлении данных, а разметка должна быть многоуровневой и сочетать в себе самые различные параметры, которые могут представлять интерес для историка или источниковеда. В ходе разметки неизбежно происходит формализация информации источников, однако это не трансформирует оригинальный текст источника, поскольку формализация здесь проявляется как условное, стандартизированное и унифицированное, обозначение (маркировка) выделенных частей документа, то есть, в сущности, формализация является интерпретацией источника, «надстройкой» над его текстом.

Традиционно источниковедческий анализ, или критика источника, рассматривается как двухстадийный процесс: первый его этап – внешняя или аналитическая критика источника (иначе, критика текста, филологическая критика или критика происхождения), второй этап – внутренняя или синтетическая критика (иначе, историческая критика или критика содержания). Аналитическая критика изучает происхождение источников, предполагает установление их подлинности, авторства, времени, места, условий возникновения. Синтетическая критика заключается в усвоении целей, задач, смысла источников как документов классовой политической, партийной борьбы. Она состоит в выяснении научного значения источника

для изучения жизни и устанавливает, таким образом, не только фактическую ценность, достоверность сообщаемых фактов, т.е. соответствие содержания реальной исторической действительности, но и его классовую направленность. Синтетическая критика устанавливает также полноту источника, выявляет умолчания и искажения, их причины и цели.

Моделирование исторических процессов и явлений

Моделирование – метод построения моделей объектов, процессов или явлений с целью их подробного исследования. Под моделью имеется в виду объект, схожий с оригиналом, являющийся упрощением последнего и используемый в целях познания.

Компьютерное математическое моделирование в историческом исследовании

Цели моделирования. Математическое моделирование в истории решает различные задачи. Первой и главной задачей моделирования представляется построение таких моделей, которые позволяют выявить коренную суть изучаемых явлений и процессов в целом, т.е. рассмотреть их как определенные системы. Воссоздаются отсутствующие данные об объекте или процессе в какой-то промежуток времени.

Второй задачей моделирования является прогнозирование дальнейшего хода развития общественных систем.

Кроме того, математическое моделирование позволяет проводить изучение альтернатив исторического развития.

Этапы моделирования. Выделяются два этапа моделирования – сущностно-содержательный и формально-количественный.

На сущностно-содержательном этапе определяются конкретные признаки исторической реальности, подлежащие изучению. Происходит выявление источников, заключающих необходимые факты, и устанавливаются их достоверность и точность. После этого формируется совокупность научно-исторических фактов, которая должна быть представительной, т.е. создавать возможность для решения поставленной задачи.

Количественное моделирование состоит в формализованном выражении качественной модели посредством тех или иных математических средств.

Классы объектов моделирования

1. *Исторический материальный объект* – от элементарного орудия труда или предмета быта до древнего поселения. В целом это направление получило название исторические реконструкции.

2. *Исторические явления и процессы* – от кратковременных событий до целых их цепей, протяженных во времени, на основе которых возможно получение закономерностей развития общества.

3. *Исторический источник.* Проблема моделирования исторического источника становится особенно актуальной в связи с необходимостью перевода исторических источников в машиночитаемый формат – в таком случае моделирование является средством формального представления источника.

4. *Историческое знание*: понятийный и категориальный аппарат, теории и гипотезы, парадигма и т.п. Математическая формализация самого исторического знания является высшим уровнем математизации науки.

По способу представления модели могут быть предметными и знаковыми. Предметные модели воссоздают физические характеристики объекта-оригинала. Примером служит метод антропологической реконструкции облика на краниологической основе, то есть восстановление лица человека по внешним данным черепа, по его рельефу и структуре. Предметными моделями являются копии орудий труда древних людей, макеты их жилищ и кораблей.

Знаковые модели являются абстрактными и воспроизводят объект-оригинал в виде формул, таблиц, графических схем, сценариев альтернативных событий, в том числе в форме исторического повествования. Последнее представляет собой описательную модель. Например, придуманный литературный герой из художественного произведения реалистичного жанра, которое было создано в изучаемую эпоху, может стать для исследователя образцом описания мотивов поведения типичного представителя какой-либо социальной общности в определённой исторической обстановке, что допустимо, так как часто литературные герои представляют собирательный образ, имеющий реальных прототипов.

Типы моделей: модели-реконструкции исторических явлений и процессов; модели построения исторических альтернатив; моделирование нелинейных исторических процессов

Типология моделей исторических процессов и явлений, включает отражательно-измерительные и имитационно-прогностические модели.

Отражательно-измерительные модели представляют изучаемую реальность такой, какой она была в действительности, выявляя и анализируя статистические взаимосвязи в системе показателей, характеризующих изучаемый объект. Целью имитационных моделей является реконструкция отсутствующих данных о динамике изучаемого процесса на некотором интервале времени. Здесь возможен анализ альтернатив исторического развития и теоретическое исследование поведения изучаемого явления (или класса явлений) по построенной математической модели.

Выделяют два типа имитационных моделей: имитационно-контрфактические и имитационно-альтернативные модели исторических процессов. Обычно контрфактическое моделирование ассоциируется с произвольным перекраиванием исторической реальности, но, с другой стороны, оно может быть эффективным инструментом изучения альтернативных исторических ситуаций. Здесь находят применение аналитические и имитационные модели. Для первых характерна запись процессов функционирования рассматриваемой системы в виде функциональных соотношений (уравнений). Имитационные модели воспроизводят сам изучаемый процесс в его функционировании во времени.

В настоящее время разрабатываются модели, охватывающие как закономерные периодические явления и процессы, так и выходящие за рамки повторяющихся, в том числе нелинейные, состояния «хаоса», фазовых переходов, учитывающих стохастические явления и т.п. Растущий интерес историков к изучению переходных эпох, альтернатив исторического развития, роли случайностей в истории способствует внедрению синергетики в изучение истории, которая рассматривается сегодня как методологическая база изучения неустойчивых исторических процессов. В соответствии с синергетической парадигмой длительные стабильные состояния – аттракторы – сменяются короткими периодами хаотического поведения – бифуркациями. Традиционные методы больше подходят для изучения устойчивых состояний, в то время как синергетика дает аппарат для изучения нестабильных периодов в истории, выявления закономерностей развития, периодизации, а также изучения альтернатив.

Клиометрия и клиодинамика

В 1960-х гг. начало складываться научное направление, связанное с применением математических методов в истории. Оно имело несколько различных названий: квантитативная история (от quantitative – количественный), клиометрика (или клиометрия) и первоначально означало одно из основных направлений исследований по экономической истории, основанное на применении математико-статистических методов и моделей. Иными словами, это «историческая экономика», «экономика истории». В отечественной науке вместо термина клиометрия использовалось понятие «количественная история».

Еще одним новым направлением в моделировании исторических процессов стала клиодинамика. Сторонники применения данного метода используют математические модели для изучения динамических процессов в истории: сокращения или увеличения численности населения, роста или упадка экономики, укрепления или развала государств. Для этого исследователи анализируют четыре основных переменных: численность населения, социальную структуру, государственную мощь и политическую нестабильность.

Методы моделирования (дифференциальные уравнения, вероятностное моделирование, табличные модели, базы данных, ряд Фибоначчи) и их практическое применение

Системы дифференциальных уравнений задают динамику процесса и позволяют интерполировать и экстраполировать существующие данные. Такие модели могут быть простейшими или содержать неопределенные параметры, изменение которых существенно влияет на поведение системы. Простейшая модель демографической динамики, не учитывающая войны, климатические катаклизмы и т.п., описана для землевладельческого общества, однако даже такая модель позволяет сделать на ее основе определенные выводы о динамике популяции.

Любым историческим процессам свойственен вероятностный характер, с другой стороны, история традиционно изучает зависимости и причинно-следственные связи различных факторов, поэтому использование аппарата теории вероятностей и математической статистики является достаточно традиционным в построении и анализе моделей – как в качестве основного метода, так и наряду с другими. Для описания стохастических моделей в дискретной форме используются цепи Маркова. Примером применения вероятностных методов является выявление закона распределения по данным стрельб и его использование для реконструкции параметров метательного оружия. Статистический анализ распределения мифологических мотивов может использоваться для построения модели эволюции мифологии и изучения корреляции мотивов с формами социальной организации и прочими факторами

Табличные модели наиболее простые в построении, применяются для систем с конечной совокупностью элементов, которые можно разместить по какой-либо иерархической схеме; до определенного времени табличные модели считались единственно реально полезными для истории. Такие модели позволяют решать различные классы задач, в том числе упорядочивания и классификации информации и статистические задачи. В настоящее время для создания и анализа табличных моделей используются электронные таблицы, статистические пакеты и базы данных.

Показано, что схема основных этапов развития человечества в археологическую эпоху может быть представлена с помощью ряда Фибоначчи, который отражает количественно наблюдаемое археологами и историками качественное явление – «ускорение исторического времени».

2. Специфика применения Всемирной сети при проведении исторических исследований

Сетевой сервис под названием WWW с самого начала создавался как среда для оперативных научных публикаций, связанных между собой гиперссылками. Что же касается вычислительных сетей в целом, и Интернета в частности, то задача, стоящая перед ними, была и есть намного серьезнее – организация произвольных коммуникаций между людьми, программами и данными. При этом потенциальные, лишь частично реализованные возможности сети в историческом источниковедении и, в целом, в исторических исследованиях гораздо шире. Общая логика развития Всемирной паутины диктует необходимость перехода к территориально распределенным исследованиям, в качестве актуальных задач ставит разработку специализированных семантических сетей для репрезентации и изучения больших комплексов исторических источников и создание семантических публикаций памятников письменности и других типов источников.

Всемирная сеть и исторические исследования в 90-е гг. XX в. Реальные возможности распределения компьютерных вычислений среди

множества машин, соединенных через сеть, появились только в конце 90-х годов XX века в связи со стремительным развитием Интернет-технологий. Способ распределенных исследований стал применяться для задач, требующих больших вычислительных мощностей либо использующих значительные объемы территориально распределенных данных. Технология организации распределенных вычислений в сети Интернет получила название Grid-технологии, а само направление развития научных исследований с использованием распределенных вычислений стало называться e-Science. В рамках Grid появился термин «виртуальная организация», обозначивший комплекс аппаратных, программных и организационных решений, а также человеческие ресурсы, обеспечивающие распределенные вычисления. В то же время возникли так называемые «виртуальные лаборатории», сначала для компьютерных экспериментов в учебных целях, затем – и для научных исследований, особенно в области биоинформатики. Основой таких лабораторий были специализированные программы, запускаемые через web-интерфейс и реализующие алгоритмы статистического анализа, сравнения последовательностей, построения филогенетических деревьев и т.д.

Виртуальные лаборатории быстро обрастали сопутствующими ресурсами, такими как тезаурусы, словари, библиотеки. С появлением технологий Web 2.0 стали создаваться инструменты для совместного редактирования и аннотирования научных текстов, обмена библиографическими описаниями и ссылками. Появился и широко распространился термин «виртуальная среда для научных исследований» (Virtual Research Environment, VRE). Методы и технологии организации VRE стали составной частью e-Science.

Выход e-Science за рамки Grid-технологий привел к тому, что сначала в США и Великобритании, а затем и в других странах стали делаться попытки использования методологии e-Science в социальных, а затем – и в гуманитарных науках. В этих науках тоже возникают задачи, требующие широкомасштабных распределенных вычислений и обработки больших массивов оцифрованных данных, например, коллекций текстов в корпусной лингвистике, растровых изображений и электронных текстов письменных источников в историческом источниковедении, фотографий предметов, привязанных к планам раскопок, в археологии, аудиозаписей в устной истории и фольклористике. В данных случаях обычно применяются Grid-технологии.

Основные направления применения Всемирной сети в исторических исследованиях на современном этапе

Доступ к архивам и источникам. Многие архивы и библиотеки оцифровали свои коллекции, что позволяет исследователям получать доступ к редким документам, манускриптам и другим историческим материалам из любой точки мира.

Онлайн-базы данных: Существуют специализированные базы данных, содержащие исторические документы, статистику, карты и другие ресурсы, которые облегчают поиск и анализ информации.

Сетевые сообщества и коллаборации. Ученые могут взаимодействовать друг с другом через онлайн-платформы, участвовать в форумах, семинарах и конференциях, что способствует обмену знаниями и идеями.

Использование цифровых технологий для анализа исторических данных: текстовый анализ (синтаксический анализ, парсинг), визуализация данных, географические информационные системы.

Образовательные ресурсы. Интернет предоставляет доступ к онлайн-курсам, лекциям и образовательным материалам, что способствует популяризации исторических знаний и повышению уровня образования в этой области.

Социальные сети и блоги. Историки могут делиться своими исследованиями и находками через блоги и социальные сети, что помогает привлечь внимание широкой аудитории и повысить интерес к истории.

Краудсорсинг. Некоторые проекты по оцифровке и анализу исторических данных используют краудсорсинг, позволяя добровольцам участвовать в исследовательских процессах.

Интерактивные платформы. Создание интерактивных карт, временных линий и других визуальных инструментов помогает лучше понять исторические события и их контекст.

Влияние использования глобальной сети Интернет на качество исторических исследований

Положительное влияние.

Доступ к информации. Интернет предоставляет доступ к огромному количеству исторических источников, архивов и библиотек, что позволяет исследователям находить редкие документы и материалы, которые ранее были недоступны.

Упрощение поиска. Поисковые системы и специализированные базы данных позволяют быстро находить нужную информацию, что экономит время и усилия исследователей.

Междисциплинарные исследования/ Возможность взаимодействия с учеными из других областей (например, социологии, антропологии, информатики) способствует более глубокому и многогранному анализу исторических событий.

Цифровизация и сохранение. Оцифровка исторических документов и материалов помогает сохранить их для будущих поколений и облегчает доступ к ним.

Визуализация данных. Использование графиков, карт и других визуальных инструментов позволяет лучше понять сложные исторические процессы и связи.

Обратная связь и обсуждениею Онлайн-платформы и социальные сети позволяют исследователям получать обратную связь от коллег и широкой аудитории, что может улучшить качество работы.

Отрицательное влияние.

Качество источников: В интернете можно найти как высококачественные, так и недостоверные источники. Исследователи должны быть осторожны и критически оценивать информацию, чтобы избежать распространения мифов и неточностей.

Переизбыток информации. Огромное количество доступной информации может привести к затруднениям в отборе релевантных данных и увеличению времени на анализ, трудностям в обработке и анализе информации, а также к снижению качества выводов.

Проблемы с авторскими правами. Оцифровка и использование материалов могут сталкиваться с юридическими ограничениями, что может затруднить доступ к некоторым источникам.

Изменение методов исследования. Некоторые исследователи могут полагаться на интернет-ресурсы в ущерб традиционным методам, таким как работа с первоисточниками в архивах, что может снизить глубину анализа.

Рост плагиата как отдельная проблема применения Всемирной сети в исторической науке

Доступность информации. Интернет предоставляет легкий доступ к огромному количеству исторических материалов, статей и исследований. Это создает соблазн для некоторых исследователей копировать и вставлять информацию без должного цитирования.

Анонимность и безнаказанность. В сети сложно отследить авторство, что может способствовать плагиату. Некоторые исследователи могут считать, что они не будут пойманы, если используют чужие работы без указания источника.

Недостаток образовательных программ. Не все учебные заведения уделяют достаточное внимание обучению студентов правильному цитированию и этике научной работы. Это может привести к тому, что молодые исследователи не осознают серьезность плагиата.

Автоматизированные инструменты. Существуют программы и онлайн-сервисы, которые могут помочь в обнаружении плагиата, но не все исследователи используют их. Это может привести к тому, что плагиат остается незамеченным.

Влияние на научное сообщество. Плагиат подрывает доверие к научным исследованиям и может негативно сказаться на репутации исторической науки в целом. Это также может затруднить доступ к качественным исследованиям, если публикации будут подвергаться сомнению.

Правовые и этические последствия. Плагиат может привести к юридическим последствиям и утрате академической степени или должности. Это подчеркивает важность соблюдения этических норм в научной деятельности.

3. Обзор языков программирования и возможности их применения в исторической науке

XML – язык разметки, используется для хранения и передачи данных. Является стандартом представления данных, ориентированным, в частности, на обмен информацией между независимыми участниками и структурную, а не только «оформительскую» разметку информации. Он был разработан для упрощения обмена данными между различными системами и приложениями. При создании цифровых архивов и библиотек все чаще используются технологии XML.

Для компьютерного источниковедения особое значение имеет то, что технология XML сохраняет целостность исторического источника как отдельного текстового файла, доступного для редактирования и обработки с помощью универсального программного обеспечения. Технология XML особенно эффективна, если возникает необходимость интеграции комплекса исторических источников в единую базу данных с целью проведения операций источниковедческого анализа и синтеза. В данном случае база данных – это единый XML-файл, интегрирующий совокупность источников и представляющий информационную модель комплекса источников.

XPath – язык запросов, используемый для навигации и поиска информации в XML-документах. Он позволяет точно указать путь к элементам, атрибутам и текстовым данным, которые нужно найти в структуре XML файлов и получить данные при необходимости.

XSLT – язык преобразования, который используется для преобразования XML-документов в другие форматы, такие как HTML, текст или другие XML-документы. Он основан на XPath и использует его для выборки узлов.

XQuery – язык запросов, предназначенный для извлечения и обработки данных из документов, представленных в формате XML. Он был разработан для работы с XML-данными и позволяет выполнять сложные запросы, манипулировать данными и генерировать новые XML-документы. XQuery более мощный и гибкий, чем XPath.

Потенциал и перспективы использования искусственного интеллекта в научно-исследовательской и педагогической деятельности историка

Pythia – нейросетевая модель разработки DeepMind для восстановления текстовых данных. Разработчики протестировали модель на задаче восстановления текстов на древнегреческом языке.

Ithasa построена на основе системы восстановления текста Pythia. Ithasa не просто помогает историкам восстанавливать текст, но и способна определять его происхождение, а также дату создания. Ithasa уже использовалась для разрешения спора о корректной датировке нескольких древнеафинских указов.

Технология проекта Российского исторического общества и Сбера «Digital Пётр» позволила расшифровать трудно читаемые из-за плохого почерка рукописи Петра I. Эта технология может быть применена для работы с манускриптами Древнего Египта.

ЛЕКЦИЯ 3. Электронный текст исторических исследований: создание и хранение

1. Электронный текст исторических исследований.
2. Обзор программ создания электронных текстов.

1. Электронный текст исторических исследований

В основном электронные исторические источники определяются как «результат трансформирования (отражения) исторических источников в электронной форме». Поэтому основная проблема – это вопрос об адекватной передаче накопленной человечеством информации с традиционных носителей (прежде всего бумаги) на электронные. Сопутствующими являются вопросы установления идентичности электронной версии и оригинала исторического документа (источника), проблема цитирования, проблема статуса электронных источников и др.

Текст является для исторической науки и образования и средством (например, учебники), и предметом (например, исторические источники) изучения, и в этом также состоит его особая роль, несвойственная другим формам представления информации. В том случае, если источник имеется в электронном виде, может происходить интеграция средства обучения, объекта и предмета исследования в рамках одного конкретного текста. Так, за счет использования языков разметки или гиперссылок в текст исторического источника может быть заложен алгоритм его изучения или модель его интерпретации, в том числе вариативные. С использованием элементов разметки (HTML, XML, TEI) текст приобретает качественно новые возможности и значение: традиционное понятие текста расширяется за счет введения в него структуры, нелинейности, установления перекрестных связей, гипертекст позволяет реализовать изучение механизма образования ассоциативных связей, задать алгоритм его прочтения, что особенно важно при изучении студентами исторических источников. Процесс такой разметки текста является элементом моделирования информации, представленной в тексте, он важен как с точки зрения изучения рассматриваемого текста, так и с точки зрения формирования соответствующего умения.

Электронный текст, в отличие от многих других информационных технологий, используемых историками, таких как базы данных и мультимедиа, напрямую воспроизводит традиционный печатный текст. При этом

электронная форма, сохраняя смысл и предназначение текста как такового, допускает качественно новые возможности работы, позволяет историку использовать исследовательские инструменты, не доступные ранее, не отрицая при этом традиционные технологии работы с текстом «классического» источниковедения.

Электронный текст позволяет плавно, поэтапно, прозрачно сформировать предметную информационную компетентность в области истории.

Формирование умений работы с электронным текстом становится одной из важнейших задач исторического образования в области информационных технологий. Освоение текстовых технологий студентами-историками включает традиционные составляющие: создание, хранение, визуализация, (представление) и анализ текста, которые реализуются в рамках курсов информатики и исторической информатики.

Выделяется два основных вида электронных исторических источников: «вторичный», возникший первым и представляющий собой дубликат исторических источников-оригиналов, первоисточников (сканирование, перепечатка), и «первичный» электронный исторический источник.

2. Обзор программ создания электронных текстов

Блокнот (Notepad) – это простое текстовое приложение, встроенное в операционную систему Windows. Оно предназначено для создания и редактирования текстовых файлов и является одним из самых базовых инструментов для работы с текстом. Блокнот имеет минималистичный и интуитивно понятный интерфейс, что делает его доступным для пользователей любого уровня. Работает с текстовыми файлами в формате .txt, а также может открывать и редактировать файлы с другими расширениями, такими как .log и .csv. Поддерживает стандартные функции редактирования, такие как копирование, вставка, вырезание, отмена и повтор действий. Встроенные инструменты для поиска и замены текста позволяют быстро находить нужные фрагменты и заменять их. Поддерживает кодировки Unicode, что позволяет работать с текстами на различных языках и с символами, выходящими за пределы стандартного ASCII.

Notepad++ – бесплатный текстовый редактор с открытым исходным кодом, который предназначен для работы на платформе Windows. Он является мощным инструментом для программистов, веб-разработчиков и всех, кто работает с текстовыми файлами. Поддерживает множество языков программирования и обладает широким спектром функций, которые делают его незаменимым помощником в повседневной работе. Этот редактор часто используется как альтернатива стандартному Блокноту, предлагая значительно больше возможностей и гибкости.

Notepad++ был разработан с целью предоставить пользователям удобный и функциональный инструмент для редактирования текста и кода.

Он поддерживает работу с различными кодировками, что делает его универсальным решением для работы с текстовыми файлами на разных языках. Кроме того, Notepad++ активно развивается и обновляется, что гарантирует его соответствие современным требованиям и стандартам.

Notepad++ поддерживает более 50 языков программирования, что делает его универсальным инструментом для разработчиков, работающих с различными технологиями. Кроме того, Notepad++ позволяет пользователям создавать свои собственные определения синтаксиса. Можно настроить редактор под свои нужды, добавив поддержку новых языков или изменив существующие настройки.

Одной из ключевых функций Notepad++ является подсветка синтаксиса. Это позволяет легко различать ключевые слова, комментарии и другие элементы кода, что значительно упрощает чтение и редактирование файлов. Подсветка синтаксиса делает код более наглядным и структурированным, что особенно важно при работе с большими проектами. Подсветка синтаксиса настраивается для каждого языка программирования отдельно.

Многооконный режим в Notepad++ позволяет открывать и работать с несколькими файлами одновременно, а также поддерживает разделение окна на несколько частей. Notepad++ поддерживает использование регулярных выражений для поиска и замены текста.

Notepad++ имеет встроенный менеджер плагинов, который позволяет легко устанавливать, обновлять и удалять плагины. Это значительно расширяет функциональность редактора и позволяет адаптировать его под конкретные задачи.

Notepad++ отлично справляется с обработкой больших файлов, что делает его идеальным инструментом для анализа логов и других объемных данных.

WordPad – текстовый редактор, встроенный в операционные системы Windows. Он предлагает более продвинутые функции по сравнению с Блокнотом, что делает его подходящим для создания и редактирования текстовых документов с некоторым уровнем форматирования.

WordPad может открывать и сохранять документы в различных форматах, включая .rtf, .txt (с потерей форматирования) и .docx. Это делает его более универсальным по сравнению с Блокнотом.

WordPad позволяет изменять шрифт, размер, цвет и стиль текста (жирный, курсив, подчеркивание, зачеркивание, надстрочный, подстрочный). Можно также применять различные стили абзацев (отступы, маркированные/нумерованные списки) и выравнивание текста, создавать комбинированные документы, содержащие текст, графику, числовые данные в электронных таблицах и даже мультимедийные файлы небольшого размера.

Microsoft Word – один из самых популярных текстовых процессоров, созданный корпорацией Microsoft. Он является частью офисного пакета

Microsoft Office. Предоставляет широкие возможности для создания, редактирования и форматирования текстовых документов.

По своим функциям Word приближен к издательским программам верстки. Это значит, что в этом редакторе можно полностью подготовить к печати газету, книгу, изготовить интернет-страницу.

Основные функции и возможности Microsoft Word.

- Пользовательский интерфейс Word прост и удобен, что делает его доступным для пользователей любого уровня.
- Создание нового документа с помощью шаблонов
- Одновременное открытие и работа с большим количеством документов.
- Автоматическая коррекция наиболее часто повторяющихся ошибок.
- Расширенные возможности форматирования документа.
- Word допускает выравнивание документа по обоим краям, многоколоночную верстку.
- Использование стилей для быстрого форматирования документа.
- Автоматическая проверка орфографии, грамматики и стилистики при вводе документа.
- Удобные механизмы работы с ссылками, сносками, колонтитулами.
- Включение в текст элементов, созданных в других программах Microsoft Office – графических изображений, электронных таблиц и графиков, звуков, видеоизображений и т.д.
- Подготовка простых электронных таблиц и гипертекстовых документов Internet.
- Работа с математическими формулами.
- Автоматическое создание указателей и оглавления документа.
- Отправка готового документа непосредственно из Word на факс и по электронной почте (необходимость оснащения модемом).
- Встроенный мастер подсказок и объемная система помощи.
- Автоматизация ввода повторяющихся и стандартных элементов текста.
- Word поддерживает совместную работу, позволяя нескольким пользователям одновременно редактировать один и тот же документ. Это особенно полезно для командных проектов.
- Word интегрируется с облачными сервисами, такими как OneDrive, что позволяет пользователям сохранять документы в облаке и получать к ним доступ с различных устройств.
- Word поддерживает создание макросов с использованием Visual Basic for Applications (VBA), что позволяет автоматизировать повторяющиеся задачи.

Использование средств программы Microsoft Word для оформления электронного текста исторического исследования по требованиям Высшей аттестационной комиссии Республики Беларусь

Технические требования к оформлению текста ВАК включают такие параметры текста, как размеры полей, гарнитуру, размер и начертание шрифта, отступы, интервалы, автоматическое оглавление.

Здесь приводятся действия, выполняемые в Microsoft Word 2010.

Для установки требуемых размеров полей на ленте выбирается вкладка *Разметка страницы* в группе *Параметры страницы* выбирается кнопка открытия диалогового окна «Параметры страницы», в котором задаются значения левого, правого, верхнего и нижнего полей.

Для настройки шрифта во вкладке *Главная* в группе *Шрифт* либо в контекстном меню выбирается гарнитура, размер шрифта и начертание.

Для установки требуемых параметров абзаца во вкладке *Главная* в группе *Абзац* выбирается кнопка открытия диалогового окна «Абзац», в котором задаются левого и правого отступа, интервала перед и после абзаца, выбирается отступ первой строки и междустрочный интервал.

Для создания автоматического оглавления для каждого заголовка, подлежащего внесению в оглавление, в абзац заголовка помещается текстовый курсор и вызывается диалоговое окно «Абзац», в котором выбирается уровень заголовка в структуре документа. Далее, во вкладке *Ссылки* выбирается *Оглавление, Автособираемое оглавление 1*. При изменении, а также добавлении и удалении заголовков на оглавлении производится нажатие правой кнопкой мыши и в появившемся контекстном меню выбирается *Обновить поле* и далее «обновить целиком» При изменении текста документа при неизменных заголовках в этом случае выбирается «обновить только номера страниц».

ЛЕКЦИЯ 4. Базы данных в исторических исследованиях

1. Роль баз данных в исторических исследованиях.
2. Реляционные и нереляционные базы данных.
3. Обзор различных систем управления базами данных.
4. Базы данных в исторических исследованиях в Беларуси.
5. Основные подходы к проектированию структуры исторических баз данных и практика их создания.

1. Роль баз данных в исторических исследованиях

Коллекции машиночитаемых данных получили название баз данных. В широком смысле база данных – это массив данных, хранимый в вычислительной системе. Однако не всякий информационный массив является базой данных в строгом смысле этого понятия, поскольку согласно технологии баз

данных организация информации в базе данных должна быть подчинена определенным требованиям.

База данных – это совокупность структурированных взаимосвязанных данных при такой минимальной избыточности, которая допускает их использование для различных приложений в определенной предметной области.

Стандартные требования к организации базы данных:

– Интегрированность (централизованное хранение информации). Неинтегрированные базы данных по одной и той же проблеме (созданные, например, в разное время и с разными целями) почти неизбежно обладают избыточностью и не являются непротиворечивыми.

– Взаимосвязанность и структурированность, отражающие существенные свойства объектов реального мира.

– Независимость описания данных от прикладных программ (логическая и физическая независимость), т.е. изменения, касающиеся логической структуры данных, не должны влиять на их расположение в памяти системы.

В современной технологии баз данных эти задачи решаются централизованно с помощью систем управления базами данных (СУБД). Главная роль СУБД состоит в обеспечении пользователя необходимыми инструментальными средствами описания данных и средствами манипулирования данными как на логическом, так и на физическом уровне, а также в обеспечении защиты данных (от несанкционированного доступа, от разрушения при сбоях оборудования) и их целостности (непротиворечивости).

В исторической информатике одним из ведущих направлений является создание и анализ баз данных и информационных систем. Распространение этих технологий в 1980-х гг. стало наиболее ярким проявлением исторической информатики как новой междисциплинарной области.

Бадам данных предшествовали машиночитаемые источники, создававшиеся в 1960–1970-е гг. На их основе нельзя было проводить поиск информации, а это основная функция систем управления базами данных, но свою главную, т.е. аналитическую задачу они выполняли. В квантитативной истории, где преобладает т.н. проблемно-ориентированный подход к созданию электронных данных, несущественно, какую форму принимают с точки зрения технологии эти данные: имеют ли они вид обычных прямоугольных таблиц, как в эпоху больших ЭВМ, или это базы данных в узком смысле этого слова. Различие между базами данных и данными в других форматах есть не что иное, как различие в выборе систем подготовки данных, а не систем их анализа. Специфика электронных данных, создаваемых в процессе квантитативных исторических исследований, заключается в том, что для решения любой конкретно-исторической задачи не существует специального, идеального источника, который содержал бы всю необходимую- информацию, поэтому создаваемые электронные документы имеют, как правило, комплексный характер и аккумулируют материал из целого ряда источников, частично или

фрагментарно содержащих нужные сведения; исходные же данные при переводе источников в электронный вид обычно подвергаются ряду трансформаций: текстовая информация кодируется; информация, заданная на индивидуальном уровне, агрегируется; исходные группы объектов реструктурируются; исходные признаки преобразуются в относительные или сводные и т.д.

В результате такой источник не представляет собой «электронной копии» первичного источника в смысле полного переноса информации источника в электронный вид, а является по существу новым (вторичным) источником. Его даже можно назвать источником следующего поколения по отношению к предшествующим исходным документам: он может быть авторской редуцированной «электронной версией» исходного документа или мета-источником, интегрирующим сведения нескольких исходных документов и продуцированным в процессе исторического исследования.

Проблемно-ориентированный подход к созданию и анализу баз данных успешно применяется исследователями-квантификаторами. Если процесс создания вторичного источника достаточно подробно и тщательно документируется, то можно проследить происхождение всех элементов данных. Более того, такой источник можно дополнить, изменить, объединить с другими и т.п. При этом важно, что результаты компьютеризованного анализа, представленные в электронной форме, допускают вторичное использование. С наступлением эпохи микрокомпьютеров и появлением стандартного программного обеспечения популярность баз данных как стандарта подготовки электронных документов на основе массовых исторических источников резко возросла, и все чаще они стали создаваться в формате баз данных.

Весьма востребованы базы данных в исследованиях по исторической демографии, где особенно успешно анализируются такие массовые источники, как метрические книги и переписи. Успехи применения информационных технологий в исторической демографии привели к тому, что в этой предметной области появились крупные межрегиональные и межуниверситетские проекты по миграциям, генеалогии, истории семьи и др. На сегодняшний день практически синхронно работают несколько университетских рабочих групп по исторической демографии, целью которых являются вовлечение в научный оборот (создание значительных по объему баз данных по региональной истории) и последующий анализ значительных по объему массивов документов церковного, ревизского и административно-полицейского учета населения и других источников.

Другим примером успешной интеграции технологии баз данных в исследовательскую практику является социальная история. Так, в рамках проблематики «исторического профессиоведения» развиваются межрегиональные проекты, изучающие исторические аспекты возникновения и развития занятий населения, профессиональный состав и профессиональную

мобильность на различных территориях в различные исторические периоды. В этих работах для кодирования занятий населения, классификации профессий большое значение имеет источниковедческий анализ источниковой базы. Еще одним перспективным направлением создания и анализа баз данных в исторических исследованиях стала работа с текстовыми источниками. Здесь проявляется междисциплинарное взаимодействие с компьютерной лингвистикой.

2. Реляционные и нереляционные базы данных

Потребность в разработке модели, более независимой от аппаратных средств, привела к появлению в 1970 г. реляционного подхода. Самой популярной реляционной моделью базы данных стала в эпоху персональных компьютеров. Все данные в модели представляются в виде таблиц и только таблиц, в связи с чем это единственная модель, обеспечивающая единообразие представления данных. Простейшим графическим примером является классный журнал.

Достоинством реляционной модели является сравнительная простота инструментальных средств ее поддержки, недостатком жесткость структуры данных, невозможность, например, задания строк таблицы произвольной длины и зависимость скорости ее работы от размера базы данных.

В реляционной модели данные представляются в виде таблиц, связанных между собой по ключевому полю. Таблица описывает объект базы данных. Каждая строка таблицы – набор атрибутов (характеристик) данного объекта. Значение одного из атрибутов рассматривается как первичный ключ.

Термин «реляционный» происходит от английского relation – отношение.

Реляционная база данных – это система управления базами данных, предназначенная для хранения и управления структурированными данными с использованием схемы, которая определяет типы данных, отношения и ограничения между таблицами. Реляционные базы данных основаны на реляционной модели.

Реляционные базы данных предназначены для поддержания согласованности и целостности данных, а также обеспечения связей и ограничений между различными таблицами. Они полагаются на язык структурированных запросов (SQL) для запроса, манипулирования и организации данных. SQL – это мощный и широко распространенный язык запросов, позволяющий пользователям легко выполнять сложные операции с данными.

Реляционные базы данных, в отличие от нереляционных, соответствуют ACID – это требования к транзакционным системам. Соответствие им гарантирует сохранность данных и предсказуемость работы базы данных:

Atomicity, или атомарность – ни одна транзакция не будет зафиксирована в системе частично.

Consistency, или непротиворечивость – фиксируются только допустимые результаты транзакций.

Isolation, или изолированность – на результат транзакции не влияют транзакции, проходящие параллельно ей.

Durability, или долговечность – изменения в базе данных сохраняются несмотря на сбои или действия пользователей.

Некоторые популярные реляционные базы данных включают MySQL, PostgreSQL, Oracle и Microsoft SQL Server. Они были идеальным выбором для многих приложений, особенно с четко определенными структурами данных и связями, требующими согласованного и точного хранения данных.

Реляционные базы данных обладают рядом преимуществ, что делает их популярными для различных приложений. Некоторые из наиболее значительных преимуществ включают в себя:

- **Согласованность и целостность данных.** Путем обеспечения связей и ограничений между таблицами реляционные базы данных обеспечивают согласованность и надежность данных. Они поддерживают свойства ACID, гарантируя надежную обработку транзакций базы данных даже в случае системных сбоев или непредвиденных ошибок.

- **Гибкость при работе со сложными запросами.** SQL предлагает универсальный способ запроса данных в реляционных базах данных. Он позволяет пользователям анализировать данные и манипулировать ими с помощью различных операций, таких как фильтрация, сортировка, агрегирование и объединение. Это упрощает извлечение информации из нескольких таблиц и получение новых данных на основе сложных критериев.

- **Применение схемы.** В реляционных базах данных схема определяет структуру таблиц, их отношения и ограничения. Это гарантирует, что данные хранятся хорошо организованным и предсказуемым образом. Это также помогает обеспечить целостность данных и предотвращает несогласованность из-за неправильных типов данных или пропущенных значений.

- **Широкое внедрение и поддержка.** Реляционные базы данных на протяжении десятилетий были основой многих приложений, формируя большое и активное сообщество разработчиков, администраторов и экспертов. Доступно множество ресурсов, инструментов и библиотек, помогающих управлять, оптимизировать и разрабатывать приложения с использованием реляционных баз данных.

- **Совместимость с отраслевыми стандартами.** SQL – это стандартный язык запросов, упрощающий разработчикам работу с реляционными базами данных на разных платформах и системах. Это также обеспечивает лучшую совместимость и переносимость приложений, использующих реляционные базы данных.

Основными недостатками реляционной модели являются: отсутствие стандартных средств идентификации отдельных записей и сложность описания иерархических и сетевых связей; относительно низкая скорость

доступа и большой объем внешней памяти; трудность понимания структуры данных из-за появления большого количества таблиц в результате логического проектирования; далеко не всегда предметную область можно представить в виде совокупности таблиц.

Нереляционные базы данных, также известные как базы данных NoSQL (не только SQL), являются альтернативой традиционным реляционным базам данных. Эти базы данных предназначены для хранения данных в форматах, отличных от таблиц, и предоставляют более простое, гибкое и масштабируемое решение для хранения и управления неструктурированными или полуструктурированными данными. Нереляционные базы данных могут обрабатывать данные в различных форматах, таких как ключ-значение, семейство столбцов, документ и график.

Нереляционные базы данных имеют ряд преимуществ: Гибкость, масштабируемость, производительность, разнообразие типов баз данных.

3. Обзор различных систем управления базами данных (MySQL, PostgreSQL, Microsoft Access)

MySQL – это одна из самых популярных систем управления реляционными базами данных (СУБД), которая использует язык структурированных запросов (SQL) для управления данными. MySQL является открытым программным обеспечением и широко используется для разработки веб-приложений, а также в различных других областях, требующих надежного хранения и обработки данных.

Основные функции и возможности MySQL.

- Использует реляционную модель, что позволяет организовывать данные в таблицах, которые могут быть связаны друг с другом через ключи (первичные и внешние).
- Поддерживает стандартный язык SQL, который позволяет выполнять операции создания, чтения, обновления и удаления данных (CRUD). Это включает в себя сложные запросы, объединения таблиц, группировки и фильтрацию данных.
- Оптимизирован для высокой производительности и может обрабатывать большие объемы данных. Он поддерживает горизонтальное масштабирование, что позволяет распределять нагрузку между несколькими серверами.
- Предлагает различные механизмы безопасности, включая аутентификацию пользователей, управление правами доступа и шифрование данных, что помогает защитить информацию от несанкционированного доступа.
- Поддерживает транзакции, что позволяет гарантировать целостность данных. Это особенно важно для приложений, где требуется выполнение нескольких операций как единого целого (ACID-принципы).
- Поддерживает репликацию данных, что позволяет создавать резервные копии и обеспечивать высокую доступность. Это полезно для распределенных систем и для обеспечения отказоустойчивости.

- Предоставляет различные инструменты для администрирования и мониторинга баз данных, такие как MySQL Workbench, который позволяет визуально управлять базами данных и выполнять запросы.

- Является бесплатным и открытым программным обеспечением, что делает его доступным для широкого круга пользователей и разработчиков.

- Имеет большое сообщество пользователей и разработчиков, что обеспечивает доступ к обширной документации, форумам и ресурсам для решения проблем.

- Легко интегрируется с различными языками программирования и фреймворками, такими как PHP, Python, Java и другими, что делает его популярным выбором для веб-разработки.

Хотя MySQL поддерживает множество функций SQL, в некоторых случаях он может уступать другим СУБД, таким как PostgreSQL, в плане поддержки сложных запросов и аналитики.

При работе с очень большими объемами данных и сложными запросами производительность MySQL может снижаться, особенно если не оптимизированы индексы и структура данных.

PostgreSQL – это популярная свободная объектно-реляционная система управления базами данных. PostgreSQL базируется на языке SQL и поддерживает многочисленные возможности.

Преимущества PostgreSQL:

- поддержка баз данных неограниченного размера;
- поддерживает более широкий спектр современных типов данных, включая массивы, hstore (хранилище типа «ключ-значение») и JSONB (бинарный JSON), которые обеспечивают более гибкие и эффективные возможности хранения данных;

- поддерживает геопространственные данные, включая богатый набор типов данных, функций и операторов для работы с географическими данными;

- имеет более совершенную систему индексирования, чем MySQL, включая поддержку индексов B-tree, GiST (Generalized Search Tree. Обобщенное поисковое дерево) и GIN (Generalized Inverted Index. Обобщенный обратный индекс). Они предоставляют больше возможностей для оптимизации производительности запросов и поиска данных;

- поддерживает мульти-мастер репликацию;

- предлагает развитые возможности управления транзакциями, такие как уровни изолированности транзакций, атомарные транзакции и точки сохранения;

- поддерживает хранимые процедуры, написанные на различных языках, включая PL/pgSQL, PL/Tcl, PL/Perl и другие;

- расширяемая система встроенных языков программирования и поддержка загрузки C-совместимых модулей;

- реализует наследование таблиц;

- имеет надежную платформу расширений, которая позволяет разработчикам добавлять пользовательские функции и расширять основные возможности базы данных.

Microsoft Access – это система управления базами данных (СУБД), разработанная компанией Microsoft, которая позволяет пользователям создавать, управлять и анализировать базы данных. Access является частью пакета Microsoft Office и предоставляет мощные инструменты для работы с данными, что делает его популярным среди малых и средних предприятий, а также для личного использования.

Основные функции и возможности Access.

- Позволяет пользователям легко создавать новые базы данных с помощью шаблонов или с нуля. Пользователи могут определять таблицы, поля и связи между ними.

- Данные хранятся в таблицах, которые состоят из строк (записей) и столбцов (полей). Access поддерживает различные типы данных, включая текст, числа, даты и изображения.

- Предоставляет мощный инструмент для создания запросов, который позволяет пользователям извлекать, фильтровать и сортировать данные. Запросы могут быть созданы с помощью графического интерфейса или с использованием языка SQL.

- Пользователи могут создавать формы для ввода и редактирования данных. Формы позволяют упростить взаимодействие с базой данных и сделать его более удобным.

- Позволяет генерировать отчеты на основе данных из базы. Отчеты могут быть настроены для представления информации в удобном для анализа виде, включая графики и диаграммы.

- Поддерживает создание макросов и использование Visual Basic for Applications (VBA) для автоматизации задач и расширения функциональности базы данных.

- Позволяет импортировать данные из различных источников, таких как Excel, текстовые файлы и другие базы данных, а также экспортировать данные в различные форматы.

- Предлагает удобный графический интерфейс, который облегчает создание и управление базами данных даже для пользователей без глубоких технических знаний.

- Хорошо интегрируется с другими приложениями Microsoft Office, такими как Excel и Word, что упрощает обмен данными.

- Позволяет нескольким пользователям одновременно работать с одной базой данных, что делает его подходящим для командной работы.

- Лучше всего подходит для небольших и средних баз данных. При работе с большими объемами данных или высокой нагрузкой может возникнуть необходимость в более мощных СУБД, таких как SQL Server.

- При увеличении объема данных и количества пользователей производительность Access может снижаться.

4. Базы данных в исторических исследованиях в Беларуси

База данных «Лица, незаконно репрессированные». Основным источником необходимых сведений были материалы комиссий, которые пересматривали судебные и несудебные дела и принимали решение о реабилитации. Информация, которую они содержали и которая представляла интерес для потребителей, была однотипной: сведения, позволяющие идентифицировать личность; суть и обстоятельства обвинения; вид наказания и судьба человека после вынесения приговора; решение о реабилитации.

Проект Базы данных по церковным метрическим книгам и подушным переписям населения. Церковные метрические книги и подушные переписи населения (так называемые «ревизские сказки»). Сохранилось большое количество таких документов, которые на территории Беларуси охватывают преимущественно XVIII–XIX вв. Их ценность определяется в первую очередь тем, что они сохранили для нас сведения об именах, месте и времени жизни, главных жизненных событиях сотен тысяч людей всех сословий. Для большинства простых людей – это единственные письменные сведения, которые свидетельствуют об их существовании.

В 1995 г. Архивной службой Республики Беларусь была приобретена система проектирования баз данных «Конструктор».

С помощью системы «Конструктор» Белорусский научно-исследовательский институт документоведения и архивного дела совместно с Национальным историческим архивом предпринял попытку создания автоматизированной информационно-поисковой системы «Родовод», призванную обеспечить эффективный доступ к хранящимся в этом архиве генеалогическим документам. «Родовод» содержит информацию о жителях Беларуси за XVII–XIX вв.; БД состоит из четырех корней: Республика Беларусь, ревизские сказки, инвентари, метрические книги.

База данных по исторической географии Беларуси. База отражает сведения по исторической географии Беларуси, прежде всего истории населенных пунктов (X–XVI вв., до Люблинской унии). Фактически это электронный географический указатель, составленный на материалах разнообразных источников, которые имеются в архивах Беларуси, Литвы, Польши, России, опубликованных и неопубликованных. Пользователь базы сможет получить представление о том, когда, в каком источнике и в связи с чем упомянут тот или иной населенный пункт, о его административно-территориальной принадлежности и т.п.

Базы данных к 28-му тому Литовской Метрики. Две небольшие по объему базы данных к 28-му тому Литовской Метрики с условными названиями «Географический указатель» и «Именной указатель». Том содержит около 150 различных документов периода 1522–1552 гг. (преимущественно 1540–1542), здесь различного рода Листы – пожалования, на владение, потверженья, выроки и т. д.

Базы данных по материалам «Пописов Войска Великого княжества Литовского». Идет работа над Пописами Войска Великого княжества Литовского 1528, 1565, 1567 гг. (один из Пописов является документом Литовской Метрики), изданными в 1915 году в 33 томе Русской исторической библиотеки.

Базы данных по «Тарифам подымного налогообложения». Тарифы – это списки земельных владений и городских поселений с отметкой количества населенных хозяйств (дымов) и размера налогов в казну. За единицу учета принят населенный двор, который зовется дымом. Одним дымом считалась панская усадьба независимо от количества строений, усадьба шляхтича, который имел селян или обрабатывал землю лично. Дымом считался также крестьянский двор, в котором жил хозяин, имевший пахотную землю.

Созданы базы данных по Тарифу подымного налогообложения Оршанского повета 1667 г., «Реестру подымного налогообложения Минского воеводства (повета) 1667 г.», «Реестру подымного налогообложения Брестского воеводства (повета) 1667 г.» Созданы компьютерные карты населенных пунктов в среде ГИС «MapInfo».

Предпринята попытка создать базу данных по «Тарифу подымного налогообложения Оршанского повета 1747 г.» и «Тарифу подымного налогообложения Оршанского повета 1790 г.».

Полоцкая просопографическая база данных. Начата работа по составлению просопографической базы данных одного из первых высших учебных заведений на территории Беларуси – Полоцкой академии (1812–1820 гг.) и ее предшественника – Полоцкого иезуитского коллегіума (1581–1812 гг.). В Полоцкой академии училось более 600 студентов, на трех факультетах (геологическом, филологическом и свободных наук) работало 40 профессоров. Библиотека академии насчитывала около 40 тыс. томов. Только за период с 1797 по 1820 годы здесь было издано более 500 книг на русском, польском, латинском, немецком, итальянском, французском языках.

Базы данных «Музеи Беларуси в зеркале периодики 1985–1995 гг.» и «Архивы войны».

В основу создаваемой базы «Музеи Беларуси...» легли выходные данные статей из сборников, журналов, газет, справочников и т.д., включенные в библиографический указатель «Новая літаратура па культуры і мастацтву» под рубрикой «Музейное дело. Музееведение». Собран массив библиографической информации насчитывающий более 5 000 единиц. В качестве операционной среды реализации была избрана реляционная СУБД Paradox 2.

Основным источником разрабатываемой многоаспектной БД «Архивы войны» станут архивные документы (более 30 000 ед. хранения),

хранящиеся в Национальном архиве Республики Беларусь. Для управления данными файлов также использована СУБД Paradox.4.

Базы данных архивов и музеев РБ. Одиннадцать баз данных созданы в Белорусском государственном архиве кинофотофонодокументов. В архиве хранятся фотодокументы (около 170 тыс. ед. хран.), фонодокументы (около 9000 ед. хран.), документальные, научные, научно-популярные фильмы, а с 1994 г. художественные и мультипликационные фильмы, созданные на территории Беларуси (более 25 тыс. ед. хран.). Базы данных содержат информацию 1) по кино-, 2) фото-, 3) фонодокументам, 4) картотеку газетно-журнальных статей и изданий, использовавших фонд архива, 5) картотеку исследователей, запросов и отказов и 6) картотеку фондообразователей и юридических лиц. Для работы использован пакет прикладных программ CD/ISIS/M. Соответственно сформированы картотеки названия кинодокументов, названия сюжетов кинодокументов, географический указатель, алфавитный указатель персоналий, именной каталог режиссеров, операторов, сценаристов.

Кроме того, Национальный архив республики Беларусь создал 6 баз данных, Национальный исторический архив РБ – 7 баз данных.

С начала 90-х гг. началось использование компьютеров в музейном деле Беларуси. Первым приступил к этой работе Государственный историко-культурный заповедник «Заславль», сотрудниками которого была создана программа «Автоматизированное рабочее место хранителя фондов». В рамках информационно-поисковой системы предусмотрено решение следующих задач:

- создание, ведение и использование БД персоналий;
- формирование коллекционных и выставочных каталогов;
- учет движения музейных предметов;
- учет предметов обменного фонда;
- формирование и вывод на печать «унифицированного паспорта».

В середине 90-х гг. усовершенствованная программа «АРМ хранителя фонда» стала использоваться в Минском областном музее в г. Молодечно и Полоцком историко-культурном заповеднике, а также Музее медицины.

На основе опыта головных российских музеев была создана программа по подсистеме «Фонды». Она включает в себя два комплекса: создание и ведение базы актов; создание и ведение базы предметов. Используя эту разработку, приступили к заполнению баз данных литературных музеев М. Богдановича и Я. Коласа, Государственный музей истории белорусской литературы, а также Государственный музей Великой Отечественной войны и Национальный музей.

Недавно в заповеднике «Заславль» создана программа «Поисковая система научного архива музея», рассчитанная на ввод и обработку научной информации по истории Заславля и местечек Беларуси, а также по памятникам археологии и архитектуры заповедника.

Подводя итоги, отметим, что ведущее место в практике белорусских исследователей, использующих информационные технологии, занимает создание исторических баз данных. Базы данных используются:

- в научно-исследовательских целях;
- для археографических публикаций;
- для обеспечения деятельности архивов и музеев;
- для обучения в системе высшего образования.

Исходными данными для баз данных служат как архивные, так и опубликованные письменные исторические источники, относящиеся к территории современной Беларуси в хронологическом диапазоне от XVI века до наших дней. В качестве программного обеспечения используются:

- оригинальное программное обеспечение;
- объектно-ориентированные СУБД;
- коммерческие СУБД реляционного типа.

Наблюдается стремление историков включать в свои базы данных как полные тексты источников, так и оцифрованные копии их оригиналов.

5. Основные подходы к проектированию структуры исторических баз данных

Существуют два основных подхода к проектированию систем баз данных: нисходящий и восходящий. При восходящем подходе работа начинается с самого нижнего уровня атрибутов (т.е. свойств сущностей и связей), которые на основе анализа существующих между ними связей группируются в отношения, представляющие типы сущностей и связи между ними. Например, процесс нормализации представляет собой вариант восходящего подхода при проектировании баз данных. Нормализация предусматривает идентификацию требуемых атрибутов с последующим созданием из них нормализованных таблиц, основанных на функциональных зависимостях между этими атрибутами.

Восходящий подход в наибольшей степени приемлем для проектирования простых баз данных с относительно небольшим количеством атрибутов. Однако использование этого подхода существенно усложняется при проектировании баз данных с большим количеством атрибутов, установить среди которых все существующие функциональные зависимости довольно затруднительно. Поскольку концептуальная и логическая модели данных для сложных баз данных могут содержать от сотен до тысяч атрибутов, очень важно выбрать подход, который помог бы упростить этап проектирования. Кроме того, на начальных стадиях формулирования требований к данным в крупной базе данных может быть трудно установить все атрибуты, которые должны быть включены в модели данных.

Более подходящей стратегией проектирования сложных баз данных является использование нисходящего подхода. Начинается этот подход

с разработки моделей данных, которые содержат несколько высокоуровневых сущностей и связей, затем работа продолжается в виде серии нисходящих уточнений низкоуровневых сущностей, связей и относящихся к ним атрибутов. Нисходящий подход демонстрируется в концепции модели «сущность-связь». В этом случае работа начинается с выявления сущностей и связей между ними, интересующих данную организацию в наибольшей степени.

Кроме этих подходов для проектирования баз данных могут применяться другие подходы, например, подход «от общего к частному» или «смешанная стратегия проектирования». Подход «от общего к частному» напоминает восходящий подход, но отличается от него тем, что вначале выявляется набор основных сущностей с последующим расширением круга рассматриваемых сущностей, связей и атрибутов, которые взаимодействуют с первоначально определенными сущностями. В смешанной стратегии сначала используются восходящий и нисходящий подходы для создания разных частей модели, после чего все подготовленные фрагменты собираются в единое целое.

Практика создания исторических баз данных

До последнего времени историки пользовались для обработки исторических источников преимущественно стандартным (коммерческим) программным обеспечением – СУБД реляционного типа dBASE, FOXPRO, Paradox и др. Они с успехом используются для обработки источников формулярного типа, т.е. таких, которые уже имеют табличный вид (личные дела, анкеты, статистические таблицы).

Но существует большое количество нарративных, или текстовых, источников (тестаменты, описи имущества, документы юридического или административного происхождения), которые не имеют выразительной структуры. Проблемы, которые связаны с компьютерной обработкой подобных источников, вызвали в начале 80-х новое направление – источник-ориентированную обработку данных.

Наиболее последовательно идея источник-ориентированной обработки данных реализована в компьютерной программе KLEIO, которая предусматривает размещение информации в двух файлах – файле, который содержит данные, и в файле, который содержит модель данных, описывающей с помощью командного языка КЛИО структуру источника.

Естественной единицей измерения информации в КЛИО является документ. В документе, из которых складывается база данных КЛИО, может содержаться сколько угодно сведений. Каждый документ включает блоки сведений – информационные группы.

В основу КЛИО положена иерархическая модель данных, а не табличная, как в СУБД реляционного типа. Количество информационных групп

и элементарных информации определяется задачами исследования и структурой самого источника.

Широко известна среди историков программа ТАСТ, реализующая работу с электронным текстом. В настоящее время большинство систем обработки полнотекстовых баз данных в состоянии распознать такие элементы, как слова, строки и параграфы без указания исследователя. Ряд рутинных операций решается программным путем, например автоматическое составление списка всех слов текста с указанием частот их встречаемости в источнике, поиск слов и контекста, вывод на печать.

Работа с ТАСТом имеет ряд последовательных этапов. Первоначально исследователь должен преобразовать текстовый файл в текстовую базу данных. Далее, используя любой из трех предложенных программой описательных языков разметки документа, структурировать текст.

После образования базы из размеченного текста программа показывает список всех структурных элементов и словарный список текста. С помощью полнотекстовой базы данных можно анализировать полученный машиночитаемый источник.

В границах источник-ориентированной обработки данных существуют и иные подходы, которые соединяют возможности коммерческих программ с дополнительными процедурами. Среди них можно назвать систему SOCRATES – специализированную надстройку СУБД. SOCRATES предоставляет возможность работать с иным текстом, вводить полный текст данных и обрабатывать его.

Еще одна известная система CENSSYS разработана для обработки особого вида источников – переписей. CENSSYS соединяет черты базы данных, статистического пакета и системы обработки текстов. Она складывается из отдельных программ-модулей для хранения, поиска и анализа информации. WinCens – модификация программы для работы в среде ОС Windows.

ЛИТЕРАТУРА

Основная литература

1. Гарскова И.М. Базы и банки данных в исторических исследованиях. – Москва – Goettingen, 1994.
2. Змитрович А.И. Базы данных: учеб. пособие для вузов. – Минск: Университетское, 1991.
3. Информатика для гуманитариев: учеб. пособие / под ред. Л.И. Бородкина и И.М. Гарсковой – Москва; Саранск: Изд-во УРАО, 1997.
4. Историческая информатика / под ред. Л.И. Бородкина и И.М. Гарсковой – М.: Мосгорархив, 1996.
5. Историческая информатика: Информатика для исторических специальностей: учеб. пособие / под ред. В.Н. Сидорцова и Л.И. Бородкина. – Минск: ЗАО «Веды», 1998.
6. Лодыженский Г.М. Системы управления базами данных – кратко о главном. – М., 1995.
7. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: Финансы и статистика, 1998.
8. Шаров Ю. Введение в базы данных. – М., 1995.
9. Шафрин Ю. Основы компьютерной технологии: учеб. пособие. – М.: АБФ, 1997.

Дополнительная литература

1. Гарскова И.М. Историческая информатика: эволюция междисциплинарного направления. – СПб.: Алетейя, 2018. – 408 с.
2. Историческая информатика. учеб. пособие / под ред. Л.И. Бородкина и И.М. Гарсковой. – М.: Изд-во «Мосгорархив», 1996. – 400 с.
3. Бородкин, Л.И. Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив / Л.И. Бородкина. – СПб.: Алетейя, 2016. – 303 с.
4. Брезгунова, И.В. Технологии электронного обучения: учеб. пособие / И. В. Брезгунова, С.И. Максимов. – Минск: РИВШ, 2020. – 144 с.
5. Санникова, Н.И. Информационные технологии в исторических исследованиях и образовании: учеб.-метод. пособие / Н.И. Санникова. – Ханты-Мансийск: ЮГУ, 2018. – 114 с.
6. Информационный бюллетень Ассоциации «История и компьютер» (первоначально – Информационный бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при Отделении истории АН СССР) № 1 (1990) – 44 (2015).
7. Бородкин Л.И., Гарскова И.М. Историческая информатика: перезагрузка? // Вестник Пермского университета. Серия: История. – 2011. Вып. 2(16). – С. 5–11.

8. Бородкин Л. Digital history: применение цифровых медиа в сохранении историко-культурного наследия? // Историческая информатика. – 2012. – Т. 1, № 1. – С. 14–21.
9. Володин А.Ю. Цифровая история (digital history): виртуальная реальность или исследовательская практика? // Стены и мосты II: Междисциплинарные и полидисциплинарные исследования в истории. – М., 2014. – С. 140–147.
10. Историческая информатика: новые методы и технологии исторического исследования // Электронный научно-образовательный журнал «История». – 2015. – № 8(41). – URL: <http://history.jes.su/issue.2015.3.8.8-41>.
11. Астахов М.В., Филимонова Е.Н. О программе курса исторической информатики: поиск оптимальной модели // Информационный бюллетень Ассоциации «История и компьютер». – 1995. – № 15. – С. 46–57.
12. Белова Е.Б., Бородкин Л.И., Гарскова И.М., Измestьева Т.Ф., Лазарев В.В. Историческая информатика. – М., 1996.
13. Информатика для гуманитариев. Вводный курс / под ред. Л.И. Бородкина и И.М. Гарсковой. – М., 1997.
14. Сидорцов В.Н., Балыкина Е.Н., Комличенко В.Н., Липницкая О.Л., Носевич В.Л. Историческая информатика. – Минск, 1998.

Учебное издание

ИСТОРИЧЕСКАЯ ИНФОРМАТИКА

Курс лекций

Составитель

ДЕНИСЕНКО Татьяна Александровна

Технический редактор

Г.В. Разбоева

Компьютерный дизайн

Л.В. Рудницкая

Подписано в печать 13.11.2024. Формат 60x84¹/₁₆. Бумага офсетная.

Усл. печ. л. 2,73. Уч.-изд. л. 2,71. Тираж 45 экз. Заказ 160.

Издатель и полиграфическое исполнение – учреждение образования
«Витебский государственный университет имени П.М. Машерова».

Свидетельство о государственной регистрации в качестве издателя,
изготовителя, распространителя печатных изданий

№ 1/255 от 31.03.2014.

Отпечатано на ризографе учреждения образования
«Витебский государственный университет имени П.М. Машерова».

210038, г. Витебск, Московский проспект, 33.