

ИЗ ОПЫТА РАБОТЫ ПО СОЗДАНИЮ ВОСТОКОВЕДЧЕСКИХ ИНФОРМАЦИОННЫХ РЕСУРСОВ В ЦЕНТРЕ ВОСТОЧНЫХ РУКОПИСЕЙ И КСИЛОГРАФОВ ИНСТИТУТА МОНГОЛОВЕ- ДЕНИЯ, БУДДОЛОГИИ И ТИБЕТОЛОГИИ СО РАН

Развитие информационных технологий позволяет существенно пересмотреть роль и место архивных коллекций и фондов в современном обществе. Если раньше доступ к информации, хранящейся в них, был существенным образом ограничен, то теперь появляется возможность актуализировать огромные массивы информации, обеспечить к ним публичный доступ.

Особенно острым этот вопрос является для востоковедческих коллекций, которые традиционно находятся вне широкого научного и публичного интереса. Активное освоение информационных технологий позволяет им подняться с периферийного уровня, представить культурное наследие народов Востока во всей полноте.

Освоение духовного опыта народов Центральной Азии невозможно представить без открытого доступа к памятникам культурного наследия, сокрытым в музейных, архивных и научных фондах и часто доступным лишь узкому кругу исследователей. Для реконструкции духовных основ созданных в Центральной Азии самобытных, во многом уникальных культур особое значение имеет огромный сохранившийся корпус письменных источников, поэтому одна из главных задач современных исследователей – обеспечить общественный доступ к памятникам письменного наследия. На сегодня это означает широкое применение информационных технологий, телекоммуникаций, Интернета. Наряду с глобализацией, которая часто осмысливается как идейная и культурная унификация, ныне существуют реальные возможности интеграции культур, взаимного проникновения и влияния идей и традиций. Именно в условиях открытости и доступности информационных потоков опыт центрально-азиатских культур получает шанс быть востребованным далеко за пределами их традиционного бытования.

Территория этнической Бурятии исторически принадлежит монгольскому и, шире, центрально-азиатскому культурному ареалу. Этот фактор, наряду с многовековым бытованием тибетской традиции буддизма, послужил тому, что на протяжении нескольких столетий тибетский и монгольский литературные языки являлись языками культуры, образования, общественной деятельности. Тибетский, в первую очередь, был языком религиозного культа и связанных с буддизмом отраслей

знания («искусств») – философии, медицины, астрологии и т.д. Вместе с тем достаточно высокое место занимал старописьменный монгольский, который в течение XVIII–XIX вв. занял прочное положение как язык официального делопроизводства в органах самоуправления, светской литературы, письменного общения. С XVII в., со времени начала широкого распространения монгольской письменности и буддизма среди бурят, в Бурятии был создан колоссальный массив рукописных и печатных памятников на тибетском и монгольском языках. На протяжении всего этого времени, до оформления государственных границ между странами центрально-азиатского ареала и после этого, протекал процесс активного культурного взаимодействия между народами, обусловленный этнической, религиозной и социальной общностью. Книжная, письменная культура играла в этом процессе важную, определяющую роль.

Повсюду, где распространялся буддизм тибетского образца, при монастырях возникали центры переводческой и издательской деятельности. В стороне от этой традиции не осталась и Бурятия. В XVIII – начале XX в. было построено более 40 монастырей – дацанов. О масштабах издательской деятельности в бурятских дацанах свидетельствуют официальные документы ревизий, согласно которым, например, в 1911 г. печаталось 1 696 наименований книг. Большое число сочинений бытовало в рукописной форме.

В результате репрессивных действий советской власти по отношению к буддийской церкви, разрушения традиционных социальных институтов в 20–40-х гг. XX в. была прервана живая традиция бытования монгольского и тибетского литературных языков в Бурятии. Научные учреждения Академии наук стали, по существу, основными собирателями, хранителями и популяризаторами огромного пласта культурного наследия бурятского народа. Именно для этих целей – квалифицированного сбора, хранения и популяризации письменного наследия – с 1970 г. в структуре Бурятского института общественных наук СО РАН, ныне – Институт монголоведения, буддологии и тибетологии СО РАН (ИМБиТ) существует Центр восточных рукописей и ксилографов (ЦВРК, ранее – отдел памятников письменности).

Характеристика фондов

У истоков создания фондов ЦВРК стояли Б. Барадин, Ц. Жамцарано, Х. Намсараев, Г.-Д. Нацов, С.П. Балдаев, Л.Е. Элиасов. Усилиями этих людей и их последователей – Г.Н. Румянцева, Г.Ц. Цыбикова, Р.Е. Пубаева, П.Б. Балданжапова, Ц.А. Дугарнимаева, Б.Д. Дандарона,

К.М. Герасимовой – отдел памятников письменности (ОПП) стал одним из крупнейших в мире собраний памятников письменной культуры народов Центральной Азии (вторым в России после тибетского фонда Санкт-Петербургского филиала Института востоковедения РАН).

Значительной является коллекция тибетских памятников (до 70 тыс. единиц хранения). Она включает разные издания буддийского канона, собрания сочинений выдающихся религиозных деятелей Центральной Азии. Особенностью коллекции является большой массив разрозненных сочинений, так называемых торбу, объединяющий биографии политических и религиозных деятелей Центральной Азии, учебную литературу буддийских монастырей, обрядники, разнообразные компилятивные сочинения монгольских и бурятских авторов. Торбу в наибольшей степени характеризуют региональную специфику буддийской культуры Бурятии.

Монгольская коллекция, общим числом до 6 000 единиц хранения, является одной из крупнейших в мире. Здесь представлены все жанры монгольской и бурятской литературы, отражающие многообразную духовную жизнь народов Центральной Азии. Особое место занимают бурятские и монгольские исторические летописи и хроники.

Огромной научной и общественной значимостью обладают архивные фонды – личные архивы выдающихся исследователей и деятелей культуры, дневники экспедиций и т.д. Представляет исключительную ценность уникальный фольклорный аудиоархив, созданный на основе материалов фольклористических экспедиций бурятских ученых.

В 1996 г. инициирована деятельность по дигитализации фондов ОПП, которая ведется в рамках решения глобальной задачи – сохранения культурного наследия народов Бурятии как органичной части Южной и Восточной Сибири, Внутренней Азии. Эта деятельность нацелена, в первую очередь, на популяризацию культурного достояния народов Бурятии, ввод уникальных материалов в широкий научный и общественный оборот.

Начальные предпосылки

Основные организационно-методические проблемы, требовавшие безотлагательного решения на начальном этапе работ, были связаны с большим объемом обрабатываемого корпуса памятников письменности, их слабой предыдущей научно-технической обработкой. Огромное количество памятников, особенно в тибетском фонде, не было охвачено даже первичным учетом. Фонды и коллекции разнородны по содержанию, поскольку в основном формировались в «авральном»

режиме в условиях жестких антирелигиозных репрессий 20–30-х гг. XX столетия.

Научно-техническая обработка фондов и составление рукописных каталогов проводились на протяжении нескольких десятилетий разными исследователями, при этом отсутствовала единая концептуальная схема атрибутирования памятников. Схемы описаний существенно различаются от одной коллекции к другой. Количество параметров описания невелико, от 9 до 12, и охватывает в основном базовые физические параметры.

Таким образом, наиболее важной задачей на первом этапе было создание единой открытой, расширяемой методологической основы с учетом имеющегося международного опыта.

Этапы и инструменты

В проблеме информатизации письменных источников можно выделить несколько самостоятельных подзадач, которые в то же время можно рассматривать и как последовательные этапы решения глобальной проблемы сохранения культурного наследия. На каждом этапе необходимо решать специфический комплекс технических и технологических проблем. Существует также специфика информационных ресурсов, порождаемых на каждом из перечисленных ниже этапов:

- Инвентаризация и каталогизация существующего массива источников (СУБД).
- Публикация разрабатываемых информационных ресурсов – каталогов, научных описаний, библиографий и пр. (гипертекст, СУБД).
- Создание страховых фондов на электронных носителях – перенос сканированных изображений памятников на электронные носители в целях их долговременного хранения в качестве страховых копий, для полиграфического воспроизведения, библиотечного и научного обмена (обработка изображений, оптическое распознавание, СУБД).
- Полнотекстовый ввод источников в исходной графике с использованием ручного ввода, методов оптического распознавания символов и т.д. (оптическое распознавание, гипертекст, СУБД).

Необходимо с самого начала учитывать способ публикации создаваемых ресурсов при разработке структур данных, при выборе инструментов разработки, форматов хранения и т.д. При создании мультязычных ресурсов, предназначенных для публикации в Интернете, в первую очередь, необходимо определить способ представления текстовой информации. Например, для значительного числа восточных

языков, в том числе тибетского и монгольского, не существует утвердившегося стандарта на представление данных в глобальных сетях. Для компьютерной поддержки тибетского языка существует несколько конкурирующих между собой подходов, и потому для отображения тибетского текста необходимо либо поддерживать несколько его версий в различных кодировках, либо применять какие-то дополнительные программные средства. Поэтому на сегодня единственной реальной возможностью для языков с ограниченным кругом пользователей является ввод и передача транслитерированных с помощью латинской графики текстов.

Основные проблемы

Решение задачи информатизации монгольских и тибетских письменных источников сталкивается с рядом проблем, затрудняющих успешное развитие и расширение деятельности:

- отсутствие на момент начала работ методологии создания электронных востоковедческих коллекций;
- отсутствие фактических стандартов (наличие различных, несовместимых между собой систем кодировки символов и их компьютерного ввода, слабая поддержка UNICODE в операционных системах и Интернете, отсутствие в связи с этим унифицированной шрифтовой поддержки);
- необходимость адаптации существующего программного обеспечения и разработки собственного;
- недостаток финансирования вкупе с требованием лицензионной «чистоты» создаваемых информационных систем приводит к необходимости поиска альтернативных средств разработки, распространяемых в соответствии с моделью открытого программного обеспечения. В свою очередь, это затрудняет получение технической поддержки в ее традиционном виде. Поэтому возникает необходимость внедрения и освоения новых, «сетевых» подходов при разработке программного обеспечения и информационном наполнении ресурсов.

Основные требования и предлагаемые решения

С самого начала работ исполнители нацеливались на распространение результатов – электронных ресурсов – посредством глобальных коммуникаций. При этом наиболее важными требованиями к создаваемым ресурсам, а следовательно критериями отбора инструментальных средств, технологий доставки информации, являлись: 1) открытость, 2) мобильность, 3) масштабируемость.

Наличие таких требований обуславливает выбор подходов и технологий, создание методологии. Например, для обеспечения переносимости создаваемых ресурсов между разными платформами и операционными системами (мобильности) для электронного представления текстов применяются схемы латинской транслитерации для монгольского и тибетского языков. Другим фактором, обеспечивающим мобильность, является использование многоплатформенного программного обеспечения.

Открытость и масштабируемость обеспечивается за счет ориентации на дистрибуцию информации с помощью web-приложений, использования открытого программного обеспечения. Использование программного комплекса Apache (*web-сервер*) – MySQL (*СУБД*) – PHP (*язык разработки и среда выполнения web-приложений*) позволяет размещать создаваемые информационные ресурсы как в локальной сети, так и в Интернете.

Еще одним требованием является персонализация создаваемых информационных ресурсов. Каждая запись в базе данных, фиксирующая описание того или иного артефакта, авторизована. С одной стороны, это способствует точному учету объемов выполненных работ при грантовом финансировании, с другой – позволяет существенным образом повысить ответственность исследователя за качество своей работы и создает предпосылки для эффективного решения вопросов защиты авторского права.

Практический пример. Web-ресурс «Забытые страницы истории бурят-монголов: монголоязычные источники по истории интеграции бурят-монголов в Российское государство. XVII–XX вв.»

В рамках этого проекта осуществлены техническое и содержательное описание материалов коллекции МІ монгольского фонда ОПП ИМБит СО РАН, которая включает рукописные и печатные материалы на старомонгольской (вертикальной) письменности. Эта коллекция характерна обилием материалов, относящихся к истории развития отношений между Российским, а затем и советским государством и бурят-монгольским населением Прибайкалья и Забайкалья. Документы, большинство из которых являются уникальными, можно условно разделить на несколько категорий: 1) исторические хроники бурят; 2) документы, освещающие деятельность буддийских монастырей, буддийских объединений: путеводители по святым местам; 3) частное творчество отдельных авторов; 4) переводы официальных документов рос-

сийских и советских властей, пропагандистские материалы; 5) переводы китайских новелл, словари и разговорники, монгольские исторические хроники и т.д.

Описание памятников осуществляется по 34 параметрам, включая физические и содержательные характеристики. В частности, полностью введены колофоны при их наличии. Обработаны и введены в электронную базу данных 801 монгольский памятник. Обработка включала в себя следующие этапы: а) научно-техническое описание и обработка письменных памятников на старописьменном монгольском языке; б) отбор работ, подверженных угрозе «утухания» текста и физическому разрушению; в) отбор наиболее редких и ценных текстов, составление их списка; г) ввод описания по 34 параметрам в БД; д) сканирование оттисков печатей и штампов; е) подготовка к дальнейшему хранению (сортировка по формату, упаковка).

Особенностью подхода, используемого для ввода данных в БД, является применение системы транслитерации Л. Лигети для передачи графем вертикального монгольского письма (с помощью символов латинского алфавита и диакритических знаков на их основе). Такой подход, обеспечивая мобильность и воспроизводимость информации, не является препятствием для пользования БД подготовленными монголоведами и специалистами, знакомыми с латинским алфавитом.

Осуществлен перевод со старомонгольского литературного языка на русский 17 документов, наиболее ярко характеризующих тематическое разнообразие представляемой коллекции и представляющих особый интерес для общественности ввиду значимости раскрываемых с их помощью этапов истории бурятского народа, буддийской церкви.

Эти уникальные, впервые вводимые в широкий научный и общественный оборот материалы интегрированы в web-ресурсе, который размещен у внешнего хостинг-провайдера. Для него зарегистрирован домен <http://imbtarchive.ru>. Сайт создан с применением средств и технологий web-приложений на базе трехуровневой архитектуры «клиент – сервер приложений – сервер баз данных». Роль клиента выполняет Интернет-браузер на компьютере пользователя, роль сервера приложений – интерпретатор языка PHP, выполняемый в качестве службы web-сервера Apache, база данных реализована на основе СУБД MySQL.

Материалы ресурса могут послужить источником для дальнейших исследований в различных гуманитарных областях по проблемам истории и культуры Бурят-Монголии, развития письменности и письменной культуры, изучения культурных связей с другими регионами Центральной Азии. Ожидается, что к ним будет обращен интерес

историков, источниковедов, историков литературы, фольклористов, лингвистов, искусствоведов и специалистов других гуманитарных отраслей знаний как российских, так и зарубежных.

В ходе реализации проекта были получены существенные результаты, позволяющие говорить о создании новой методологии работ по электронной обработке архивных коллекций ОПП. Во-первых, была разработана реляционная модель для описания памятников на вертикальной монгольской графике, которая включает 34 параметра. Разработанная модель позволяет учитывать содержательные и физические параметры описываемых артефактов. Во-вторых, информационная модель была реализована в виде работающего web-приложения на базе современных свободно распространяемых программных средств, что позволяет существенно снизить расходы на разработку и поддержку web-ресурсов. В-третьих, была организована сетевая модель разработки информационного ресурса, дистанционного управления проектом с помощью электронной почты и Интернета. Получен ценный воспроизводимый опыт распределенной работы творческого коллектива.

Задачи и направления текущего этапа деятельности

Текущей задачей исследований в области информационных технологий в ОПП является дальнейшее развитие средств интеграции информационных web-ресурсов для комплексного представления предметной области – историко-культурного наследия Бурятии. Задача сопряжена с разрешением проблем разнородности и тематического разнообразия накопленных информационных ресурсов, разнородности средств их логической и системной организации.

Решение видится в развитии XML-ориентированных подходов, которые позволяют описывать объекты предметной области с неограниченной степенью детализации, представлять их взаимосвязи.

На базе XML-технологий создается расширяемая, удобная в использовании информационная среда, в которой все объекты историко-культурного наследия могут быть представлены во всей своей сложности и взаимосвязанности. Эта предметная область характеризуется слабой структурированностью объектов. Вместе с тем объекты различных типов обладают большим количеством взаимных связей. Так, например, можно ввести объект «персона». Экземпляр этого объекта с атрибутом «религиозный деятель» может быть связан с множеством объектов следующих типов: «письменный артефакт», «буддийский монастырь», «святое место», «географический объект», «историческое событие». Определяя связи между перечисленными объектами, можно

осуществлять сложные выборки информации из массива историко-культурных данных. Связи между объектами часто невозможно установить заранее, они определяются только в результате дополнительных изысканий. XML предоставляет удобные средства редактирования свойств и связей объекта без нарушения целостности его логической структуры.

Как видится нам, внедрение XML-технологий позволяет создать интерактивную, живую среду, своего рода «информационный континуум», способный интегрировать все существующие сегодня и создаваемые в ЦВРК электронные ресурсы.