

Научно-информационный портал «Материалы и исследования по монгольским языкам»: структура и контент*

The Science and Data Portal “Materials and Studies on the Mongolic Languages”: the Structure and Content

А. Т. Баянова (A. T. Bayanova)¹,

В. В. Куканова (V. V. Kukanova)²

¹ зав. Научной библиотекой им. П. Э. Алексеевой, Калмыцкий научный центр РАН (г. Элиста). E-mail: ale-bayanova@yandex.ru.

Head of the P. Alexeeva Scientific Library, Kalmyk Scientific Center of the RAS (Elista). E-mail: ale-bayanova@yandex.ru.

² кандидат филологических наук, директор, Калмыцкий научный центр Российской академии наук (г. Элиста). E-mail: vika.kukanova@gmail.com.

Ph. D. in Philology (Candidate of Philological Sciences), Director, Kalmyk Scientific Center of the RAS (Elista). E-mail: vika.kukanova@gmail.com.

Аннотация. В данной работе рассматриваются структура и контент научно-информационного портала «Материалы и исследования по монгольским языкам». Данный информационный ресурс позволяет аккумулировать исследования и материалы, в том числе лексикографические источники. Структуру портала условно можно разделить на три части: с одной стороны — материалы, в их числе рукописи, электронные ресурсы; с другой — словари, разговорники; с третьей — непосредственно исследования. Отдельной компонентой являются биографии лингвистов-монголоведов, где приводятся основные работы ученых.

Ключевые слова: научно-информационный портал, монгольские языки, исследования, рукописи, словари, книжные памятники.

Abstract. The paper examines the structure and content of the science and data portal “Materials and Studies on the Mongolic Languages”. The information resource allows to accumulate corresponding studies and materials, including lexicographic sources. The structure of the portal can be nominally divided as follows: materials, including manuscripts and digital resources; dictionaries and phrase books; the research studies proper. In a separate part are the biographies of Mongolic linguists alongside with the scholars’ main works.

Keywords: science and data portal, Mongolic languages, studies, manuscripts, dictionaries, book monuments.

* Публикация подготовлена при финансовой поддержке проекта РГНФ № 15-04-12038/в «Создание и развитие научно-информационного портала «Материалы и исследования по монгольским языкам».

Интернет-технологии как средства электронной коммуникации активно внедряются и в сферу науки. Одним из инновационных способов решения проблемы формирования единого научно-информационного пространства с целью максимального приближения научных источников к исследователю является создание научно-информационных порталов. Использование подобных Интернет-решений позволяет эффективно использовать физические и интеллектуальные возможности человечества и ускорить процесс работы с источниками и научной литературой по интересующей ученого проблематике.

Первое, с чем сталкивается исследователь в разработке новой темы, — с проблемой сбора информации для проведения исследования. Если раньше ученый посещал библиотеку в целях поиска материала, каталогизировал научную литературу, заказывал ее и только после этого начинал ее изучение, то в последнее время появилась возможность воспользоваться Интернет-ресурсами. Но не все исследователи прибегают к Интернету для поиска материала по историографии вопроса (проблемы). Во-первых, в ходе поиска даже через «продвинутые» поисковые системы возникает так называемая «зашумленность», в результате не всегда выдаются только научные изыскания, очень много «мусора» (новости, реклама и др.), никаким образом не связанного с научными исследованиями. Во-вторых, результаты разбросаны по различным источникам, которые сложно найти и обработать из-за разных форматов (html, diVu, pdf и др.). В-третьих, исследователь сталкивается с проблемой при поиске информации, когда поисковая система выдает предметно не ограниченные результаты. Так, например, если исследователю необходимо найти, предположим, материалы по фонетической системе монгольского языка, то поисковая система может выдать результаты, с одной стороны, не только по фонетике монгольского языка, но и, с другой, по любым исследованиям на материале монгольского языка.

Таким образом, назрела необходимость создания такого рода информационно-справочных систем, способных решить задачи обеспечения ученых материалом по историографии вопроса или проблемы в узкоспециализированном пространстве. На это нацелен

и проект «Материалы и исследования по монгольским языкам», поддержанный Российским гуманитарным научным фондом в 2015 г. Необходимость создания подобного портала очевидна и актуализирована возросшим в обществе интересом к восточным языкам, каковыми является группа монгольских языков. Переведенные в цифровой формат материалы и исследования по монгольским языкам, аккумулированные в одной системе и описанные по одному стандарту, предоставят широкие возможности для ученых-монголоведов, поскольку электронная библиотека — это качественно новый уровень организации и распространения информации. При этом информационно-справочная система сопровождается библиографическим описанием (ключевые слова, аннотация, выходные данные и др.), а также просмотром текстов в режиме он-лайн.

Целью данной статьи является описание технологии создания научно-информационного портала «Материалы и исследования по монгольским языкам», а также структуры и контента портала. Работа строилась в нескольких направлениях: 1) изучение проблемы защищенного просмотра текстов; 2) структурирование базы данных в соответствии со стандартами отечественного и зарубежного библиотечного дела; 3) проведение каталогизации источников, посвященных изучению монгольских языков; 4) оцифровка научных текстов на русском, калмыцком, бурятском, монгольском, английском языках. Первоначально был протестирован ряд технологических решений для создания каталога и электронной библиотеки (GreenStone, ResCarta, Либер и др.), каждое из которых обладало как преимуществами, так и недостатками [Загорулько и др. 2011].

Наиболее впечатляющим технологическим решением является LibeR, однако стоит данное программное обеспечение весьма дорого. В результате нами было принято решение остановиться на бесплатном решении ResCarta, которое обеспечивает защищенный просмотр файлов и не позволяет скачивать материалы на персональный компьютер.

Данный портал создаст условия для инкорпорации и интеграции научных ресурсов по монгольским языкам в единое информационное пространство и сделает возможным содержательный и свободный доступ к этим источникам [Федоров и др. 2007]. Науч-

ная значимость проекта несомненна, поскольку, с одной стороны, отсутствуют какие-либо репрезентативные электронные ресурсы, посвященные данной проблематике (если сравнить, например, с исследованиями по русскому или английскому языку), с другой — наличие электронной библиотеки поднимет на новый уровень исследования, посвященные проблемам монгольских языков.

Следующим этапом работы над проектом были оцифровка научных изданий для создания электронной библиотеки, а также их каталогизация. В первую очередь были отсканированы рукописные и литографические источники XVIII–XIX вв., редкие словари и грамматики, находящиеся в фондах Российской научной библиотеки, Научной библиотеки Восточного факультета СПбГУ и Научной библиотеки Калмыцкого научного центра РАН (КИГИ РАН). Зная, что многие современные электронные библиотеки обладают плохим качеством изображений, были выработаны критерии, направленные на обеспечение требуемого уровня сканирования, которых и придерживались при выполнении данного вида работы (300 dpi). Для оцифровки изданий использовалась программа FineReader Pro 11, приобретенная для реализации проекта РГНФ «Национальный корпус калмыцкого языка: создание и развитие», который выполнялся в 2012–2014 гг. Сканирование производилось на планшетном сканере в формате PDF.

Важным компонентом контента портала является библиографическое описание источников, которое основано на определенных общепринятых правилах, согласно ГОСТ. Каталогизация позволяет систематизировать и идентифицировать большой массив отсканированных источников, обеспечивает оптимальные возможности системы поиска информации на портале. Контент портала включает каталог и собственно библиотеку (полнотекстовые источники).

Портал условно можно разделить на три части: с одной стороны — материалы, в их числе рукописи, электронные ресурсы; с другой — словари, разговорники; с третьей — непосредственно исследования. Отдельной компонентой являются биографии лингвистов-монголоведов, где приводятся основные работы ученых.

Раздел «Материалы» состоит из: 1) ссылок на электронные ресурсы, в их числе базы данных, корпуса; 2) рукописей. Рукописи

были собраны из различных источников, которые были найдены в Интернет-пространстве, а также были оцифрованы в фондах Российской научной библиотеки, Научной библиотеки Восточного факультета СПбГУ, Научной библиотеки Калмыцкого научного центра РАН, Национального архива Республики Калмыкия. Большое количество рукописей хранится в архиве и в личных фондах Института восточных рукописей РАН, однако они доступны только для работы непосредственно в фондах Института.

Раздел «Словари» состоит из лексикографических источников, в основном из словарей XVIII–XIX вв. Большую ценность представляют словари на старописьменных графических системах. Роль ранних двуязычных и многоязычных рукописных и первых печатных словарей монгольских языков велика: они занимают особое место как в истории книжной культуры, так и в письменной традиции монгольских народов. Отметим, что словари начали появляться еще в XVIII в., однако период наиболее активного развития лексикографической работы пришелся на XIX в., когда усилились социально-экономические связи с Востоком, в том числе и с монгольским миром. В этот же период зарождается отечественное востоковедение, которое бурно развивается во второй половине XIX в. и начале XX в. Лексикографические источники по своему типу являются переводными, практически отсутствуют толковые, синонимические и другие виды словарей, что также обусловлено политическими и религиозными (христианизация) задачами общества XIX в. Собранные рукописные или печатные словари порой являются единственным источником для исследования истории лексической системы монгольских языков, наряду с рукописными памятниками и ксилографами. Объем словарей, как правило, небольшой, но имеются и довольно большие словари, входной язык либо монгольские языки, либо европейские языки (немецкий, французский, шведский, русский и др.). В некоторых словарях даются сведения о составителе, диалекте, о территории, где собран лексический материал для словаря.

В силу разных причин собранные словари оказались в течение долгого времени не известными, упоминания об их существовании не многочисленны, отрывочны, хотя лингвисты время от времени

привлекали их в качестве материала для исследования, тем самым вводили в научный оборот отдельные фрагменты.

Что касается структуры словарной статьи в словарях, имеющих на разрабатываемом портале, то следует сказать, что словарная статья может включать не только слово из оригинального языка и слова (эквивалента) из переводного языка, но также иметь расширенную структуру, включая пометы о диалектной принадлежности, перечень фразеологизмов, примеров и т. д.

Монгольская лексикография развивалась в основном по традициям русской лексикографии — алфавитное расположение заголовочных слов. Основой для словаря служил список лексем на том или ином языке, которым впоследствии осуществлялся поиск эквивалента. Конечно, словник включал наиболее стабильную, традиционную часть лексической системы монгольских языков. При наличии многовековых традиций и многочисленных лексикографических источников достаточно слабо изучена история монгольской лексикографии, хотя следует отметить работу Б. Д. Цыренова [2013].

Раздел «Исследования» состоит непосредственно из трудов, которые посвящены изучению монгольских языков. В данную категорию входят как монографии, коллективные труды, статьи, доклады, тезисы, так и авторефераты, диссертации. Все они отражают определенные этапы в развитии монголоведения как науки. Можно условно выделить несколько этапов в развитии монголоведения — «монгольского» в широком смысле языкознания. В XVIII в. происходило зарождение науки о монгольских языках, в XIX в. — становление, что проявилось в написании грамматик, содержащих общие сведения о грамматике того или иного языка монгольской группы, в подготовке словарей. В XX в., особенно во второй его половине, появляются работы, посвященные частным вопросам в каждой проблемной области. Однако важно отметить, что развитие монгольского языкознания соответствовало общим паттернам в лингвистике, парадигмам, устоявшимся в науке.

Сфера использования проекта достаточно широка: предлагаемый ресурс можно применять не только в собственно исследовательских целях, но и в процессе обучения в высших учебных заведениях, которые специализируются на восточных языках (например,

Санкт-Петербургский государственный университет, Калмыцкий государственный университет, Бурятский государственный университет, Институт языкознания РАН, Институт востоковедения РАН, Институт восточных рукописей РАН, Калмыцкий научный центр РАН, Институт монголоведения, буддологии и тибетологии РАН и др.). Создаваемый ресурс представляет собой базу работ, посвященных разным аспектам исследования монгольских языков. Разрабатываемая электронная библиотека размещена в Интернете.

Литература

Загорулько Ю. А., Боровикова О. И., Кононенко И. С. Обеспечение содержательного доступа к лингвистическим информационным ресурсам на основе технологии порталов знаний // Известия Томского политехнического университета. 2011. Т. 318. № 5. С. 99–104.

Федоров В. В., Антопольский А. Б. Стратегия развития электронных библиотек России [электронный ресурс] // URL: http://www.aselibrary.ru/press_center/journal/irr/2007/number_2/number_2_3/number_2_3549/ (дата обращения: 01.09.2016).

Цыренов Б. Д. Монголоязычно-русская лексикография (структура, принципы, семантизация). Улан-Удэ: ГБУ РЦ «Бэлиг», 2013. 208 с.