

Электронные коллекции Национальной библиотеки парламента Грузии

© Варамишвили Г.

Национальная библиотека парламента Грузии (НБПГ)

dlgroup@nplg.gov.ge

Аннотация

Эта статья описывает процесс и историю создания электронных коллекций Национальной библиотеки парламента Грузии (НБПГ) с помощью GSDL (Greenstone Digital Library). В работе также описаны основные характеристики GSDL.

1 Цель и история проекта

Центр электронных текстов (<http://www.nplg.gov.ge/ic/>) Национальной библиотеки парламента Грузии (НБПГ) (<http://www.nplg.gov.ge/>) в 2000 году начал создание электронной библиотеки грузинской художественной и научной литературы. Была создана электронная библиотека включающая поэму Шота Руставели «Витязь в тигровой шкуре», Новый завет и еще несколько произведений других грузинских авторов как классиков, так и современников. Тексты произведений вводились вручную, а электронные книги переводились сотрудниками центра в формат HTML. Внутри текста произведений использовались внутренние ссылки, а сами «книжки» присоединялись к библиотеке вручную. Неудобствами описанного метода создания коллекции были: отсутствие возможности поиска в коллекции и сложность добавления нового произведения в коллекцию.¹

В начале 2004 года, в рамках проекта Московского офиса ЮНЕСКО начался осуществление проекта «Антология Грузинской классической литературы». Проект включал в себе использование новой технологии для приготовления текста и разработки цифровой коллекции (НБПГ). Вместо ввода текстов вручную, была внедрена ПО для

распознавания грузинского печатного текста «Hawk-Eye», которое дало возможность перевести в текстовый формат отсканированные страницы книг, что в свою очередь помогло центру электронных текстов ввести большое количество текстов за меньший срок. После распознавания, текст проверялся корректорами центра.

Для создания цифровой коллекции была внедрена ПО для создания цифровой коллекции «Greenstone Digital Library Software».

2 Greenstone Digital Library Software (GSDL)

Greenstone - пакет программного обеспечения для формирования и распространения цифровых библиотечных коллекций. Работа с данной системой обеспечивает новый способ организации информации и ее публикации в сети Интернет или на CD-ROM. Greenstone создано в рамках Проекта Новозеландской Цифровой Библиотеки в Университете Вайкато, разработано и распространено при содействии ЮНЕСКО и Human Info NGO. Оно является свободно-распространяемым программным обеспечением, доступным по адресу <http://greenstone.org> в соответствии с Лицензией GNU о широком доступе. Цель программного обеспечения состоит в том, чтобы уполномочить пользователей, особенно в университетах, библиотеках, и других учреждениях обслуживания, формировать их собственные цифровые библиотеки. Цифровые библиотеки радикально преобразуют распространение и приобретение информации в рамках партнерства ЮНЕСКО с различными учреждениями в области образования, науки и культуры во всем мире, и

Труды 7^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL2005, Ярославль, Россия, 2005.

¹ Метод был сложным (неудобным), так как для каждого произведения приходилось создавать новый набор веб страниц связанных друг с другом, с оглавлением и с главным сайтом вручную, что в своё время требовало больше времени, чем использование Greenstone Digital Library Software, для создания коллекций.

особенно в развивающихся странах. Это программное обеспечение разработано и распространяется как продукт международного трехстороннего совместного сотрудничества, установленного в августе 2000.

Программа Greenstone - это исчерпывающая система для создания и просмотра коллекций, состоящих из многих миллионов документов, включающих тексты, рисунки, аудио- и видео-файлы. Типичная цифровая библиотека, созданная с помощью Greenstone, содержит в себе множество коллекций, организованных по отдельности, хотя они имеют много сходства друг с другом. Легко поддерживаемые, эти коллекции могут быть дополнены и перестроены автоматически. Существуют несколько способов поиска информации во множестве коллекций Greenstone. Например, можно искать определенные слова, которые встречаются в тексте или в определенной части документа. Можно просматривать документы по наименованию: простым нажатием на иконку книги для чтения. Также возможно просматривать документы по определенной тематике. При поиске программа Greenstone просматривает все документы, находящиеся в коллекции (называется "full text search" "поиск по всему тексту"). Во многих коллекциях существуют индексы, разбивающие документ на несколько частей, по которым пользователь также может осуществлять поиск.

GSDL (Greenstone Digital Library) наряду с другими стандартами поддерживает стандарт Dublin Core, что дает возможность создания полноценной электронной коллекции с помощью GSDL. Также, в среде Greenstone можно создать новый стандарт, редактировать уже присутствующие стандарты, добавить – модифицировать поля стандартов. GSDL можно установить, как web library что в свою очередь дает возможность без лишних проблем использовать его как веб ресурс с электронными коллекциями. Он поддерживает почти все распространенные платформы операционных систем, как windows, так и Unix.

3 Описание работ

Во время начала работ над проектом, пользовательский интерфейс GSDL не поддерживал грузинский язык. Было решено создать грузинский интерфейс для GSDL. В сотрудничестве с группой разработчиков GSDL через два месяца с начала работ над переводом Greenstone User Interface, грузинский язык был включен в список поддерживаемых языков и официально был введен в версию 2.51. Был осуществлен перевод всех текстов использованных в пользовательском интерфейсе GSDL, были сгенерированы графические кнопки пользовательского интерфейса на грузинском языке.

Надо отметить, что процесс «оцифровки» материала для цифровых коллекций, шел быстро за счет новой технологии, суть которой заключалась в сканировании книг и распознавания сканированного материала с помощью программы Hawk Eye.

Проектом ЮНЕСКО планировалось создать цифровую коллекцию антологии грузинской классической литературы, цифровой аналог 20 томного издания.

После подготовки документов для создания коллекции началась работа над их «оживлением» и включением в коллекцию документы были в формате Microsoft Office Word (.doc). для приписания метаданных к документам, использовалась программа GLI (Greenstone Librarian Interface). После приписания метаданных, определились классификаторы для сортировки документов по списку авторов и названию. использовался классификатор Generic List. Были выбраны индексы поиска: текст, автор, название.

В первом издании коллекции вся метainформация была транслитерирована (грузинские названия – на латинском алфавите). Это было сделано специально для популяризации грузинской классической литературы в Интернете. В коллекцию было добавлено 2056 произведений грузинских авторов. Коллекция строилась на версии GSDL 2.50, в которой еще не полностью был включен грузинский язык но, несмотря на это, интерфейс был на грузинском языке. В версии 2.50 было обнаружено несколько проблем и ошибок с расширенным поиском в коллекциях, но в сотрудничестве с разработчиками, в версии 2.51 эти проблемы были устранены. На втором этапе проекта должна была, создастся коллекция с метainформацией приписанной грузинским алфавитом. Так как, в версии 2.51 грузинский поддерживался полностью, решено было построить второе издание коллекции на версии 2.51. Во втором издании было добавлено порядка 50 новых произведений, были внесены коррективы в формат сортировки и вывода списка авторов и произведений и, кроме этого, был усовершенствован механизм поиска в коллекции и поиск по алфавитному указателю.

4 Текущее состояние проекта

На данный момент коллекция «Антология Грузинской классической литературы» содержит 2138 произведений. Эти произведения сортированы по алфавитному списку авторов и названию. Кроме того, в коллекции можно осуществить поиск по тексту, авторам и названиям. В коллекции также можно провести расширенный поиск по следующим параметрам: в разделе коллекции «параметры», можно установить большое окно для поиска, что дает

возможность, поиска целых абзацев по текстам коллекции. С помощью расширенной формы запроса, возможен логический поиск с операциями ! (НЕ), & (И), | (ИЛИ). Пользовательский интерфейс коллекции, кроме грузинского, поддерживает русский, английский, французский и испанские языки.

5 Планы на будущее

На будущее центр электронных текстов Национальной библиотеки парламента Грузии (НБПГ) планирует добавить в коллекцию около 500 произведений, постепенно в коллекцию будут добавлены все печатные материалы содержащие в хранилищах НБПГ. Также, планируется добавить классификатор даты для коллекции и выпуск CD/DVD дисков для распространения в школах и других образовательных учреждениях.

Литература

- [1] Greenstone digital library software Руководство пользователя, 2004.
<http://prdownloads.sourceforge.net/greenstone/User-en.pdf>
- [2] Greenstone digital library software Руководство разработчика, 2004.
<http://prdownloads.sourceforge.net/greenstone/Developer-en.pdf>

Digital collections of the National Parliamentary Library of Georgia (NPLG)

G. Varamishvili

This article describes process and history of building Digital collections of the National Parliamentary Library of Georgia (NPLG) using Greenstone Digital Library software. Also, in this work are described main points of the GSDL