

Программный комплекс для создания и работы с электронными каталогами с разнонаправленными графиками текста

Software Systems for Designing and Using Electronic Catalogs with Differently Directed Text Graphs

Програмний комплекс для створення роботи з електронними каталогами із різнонаправленими графіками тексту

С. В. Мошкин

Компания «Инфокомм», Санкт-Петербург, Россия

А. Л. Пашкова

Российская Национальная Библиотека, Санкт-Петербург, Россия

Sergey Moshkin

«InfoComm» Company, St.Petersburg, Russia

Alla Pashkova

Russian National Library, St.Petersburg, Russia

С. В. Мошкін

Компанія «Інфоком», Санкт-Петербург, Росія

А. Л. Пашкова

Російська національна бібліотека, Санкт-Петербург, Росія

В докладе описаны специфические особенности и возможности нового программного комплекса, предназначенного для работы с многоязыковыми электронными каталогами специализированных отделов РНБ.

The paper describes specifications and features of the new software system designed to support the RNL's multilanguage electronic catalogs.

В доповіді описані специфічні особливості та можливості нового програмного комплексу, призначеного для роботи з багатомовними електронними каталогами спеціалізованих відділів РНБ.

Создание электронных каталогов в крупных библиотеках имеет ряд специфических особенностей, затрудняющих использование для этих целей большинства распространенных АБИС. К подобным особенностям относятся:

– Требования полной совместимости создаваемых библиографических записей с национальным форматом RUSMARC (так, как эти записи являются потенциальными объектами заимствования), что требует от программного обеспечения наличия развитой системы формальных и семантических проверок корректности данных.

– Высокие требования к детальности и точности библиографических описаний, что подразумевает поддержку системой всех объектов и связей, предусмотренных RUSMARC.

– Использование при каталогизации системы авторитетных файлов, что требует от программного обеспечения наличия средств ведения этих файлов (с полномасштабной поддержкой RUSMARC) и специфических средств обеспечения логической целостности связей в таких файлах.

– Необходимость создания библиографических описаний документов на любых языках, в частности использующих диакритические символы или, например, левонаправленное письмо. Это требует от ПО полной поддержки юникода, как при создании записей, так и при обработке запросов к базе. Кроме того, необходимо обеспечение мультязыковой проверки орфографии, а также корректной обработки строк, содержащих одновременно символы разнонаправленных график.

– Существование в крупных библиотеках традиционных технологических цепочек обработки библиографических записей, что требует наличия в ПО дополнительной функциональности для

технологической поддержки (учет авторства и времени изменений, контроль прав, контроль прохождения, служба внутренних сообщений и т. д.)

Обсуждаемый программный комплекс был разработан в рамках проекта по созданию мини-системы автоматизации библиотечных процессов с многоязыковой поддержкой для специализированных отделов РНБ. В основу комплекса положено программное ядро системы «Каталогизатор 2», созданной в ООО «Инфокомм» в 2006 г. и предназначенной для промышленной ретроконверсии библиографических каталогов. При разработке комплекса учитывались все перечисленные выше требования.

Первая очередь ПО включает три программных модуля: «Администратор», «Каталогизатор» и «Авторитетные файлы».

Модуль «Администратор» предназначен для управления базами данных системы и настройки ее параметров, а том числе, и для создания конкретных рабочих мест технологической цепочки на основе конфигураций двух других модулей.

Модуль «Каталогизатор» предназначен для создания и редактирования библиографических записей, а модуль «Авторитетные файлы» – авторитетных записей, соответственно. Модуль «Авторитетные файлы» может вызываться как самостоятельная программа, или как встроенная функция модуля «Каталогизатор». Модули используют общее программное ядро, основанное на RUSMARC-ориентированной объектной модели. К важным для рассматриваемых задач особенностям ядра следует отнести следующее:

- Полная встроенная поддержка RUSMARC. Формализованное описание формата содержится в системной базе данных и является основой для автоматического создания правил проверок и автозаполнений, а так же, для генерации элементов интерфейса. Кроме формальных проверок применяются семантические проверки на соответствие связанных по смыслу элементов записи. Такие проверки, естественно, программируются вручную.

- Расширяемая система справочников для потенциально повторяющихся в каталоге элементов библиографической записи. В качестве справочников используются авторитетные файлы или, (для элементов, не попадающих в общепринятую систему авторитетных файлов, например – для издательств или типографий) собственные справочники системы. Подобный подход не только снижает трудозатраты при создании записи, но и значительно уменьшает количество ошибок, а так же, упрощает проверку записей и исправление обнаруженных опечаток.

- Модуль построения поисковых запросов к базам данных с набором интуитивно понятных пользовательских интерфейсов. Модуль позволяет строить запросы любой сложности с использованием в качестве поисковых признаков любых элементов RUSMARC.

- Полная поддержка юникода и работы с многими языками. Любые символы юникода могут быть использованы как при создании записи, так и при поиске в базе данных. Возможен поиск, как с точным совпадением символов, так и в режиме «снятия диакритики». Для ввода предлагается несколько систем ввода символов, в том числе, виртуальные клавиатуры для всех поддерживаемых ОС языков, настраиваемые таблицы символов и, наконец, система автоматической простановки диакритических символов на основе многоязычных орфографических словарей. Поддерживается проверка орфографии вводимого текста для большинства европейских языков и иврита. Система проверки орфографии может быть расширена, путем добавления словарей, совместимых со стандартом Microsoft.

- Поддержка записей, содержащих различные графики текста, в том числе и разнонаправленные. В частности, возможно автоматическое определение графики поля RUSMARC, определение связанных по альтернативной графике полей и проставление необходимых подполей связей. Кроме того решено большинство проблем, связанных с некорректным отображением в ОС Windows с правонаправленной локализацией текстов левосторонней графики, содержащих неалфавитные символы.

В настоящее время, комплекс проходит опытную эксплуатацию РНБ в отделе Литературы на языках стран Азии и Африки для создания каталога изданий на идиш. В каталоге принята система одновременного использования трех график – иврит, латинской и кириллической.