

Электронная энциклопедическая коллекция Рубрикон

© Попов Игорь Васильевич, Фролкина Нина Александровна

ООО «Твинком бизнес» генеральный подрядчик, исполнитель работ по проекту
Рубрикон

content@rubricon.com

Аннотация

В сообщении представлена электронная коллекция справочно-энциклопедических материалов, размещенная в интернете на сайте Рубрикон (www.rubricon.com) и общий обзор информационной системы, разработанной для данного интернет-проекта и технологических решений, лежащих в ее основе.

1. Состав и структура Рубрикона

Сайт Рубрикон представляет собой электронную библиотеку, которая содержит информацию различных типов – энциклопедии, словари, справочники, учебники, монографии, статьи из журналов и др.

Но прежде всего проект Рубрикон – крупнейший словарно-энциклопедический интернет-ресурс. Основой энциклопедической коллекции является полная электронная версия 30-томной Большой советской энциклопедии (3-е изд., 1969-78). На 1 августа 2003 г. на сайте Рубрикон опубликованы интернет-версии **58** энциклопедий, словарей и справочников (**572665** статей и **79819** иллюстраций и карт). Это беспрецедентная по объему электронная справочная библиотека в Рунете, которая базируется на высококачественных изданиях издательства «Большая российская энциклопедия» и других авторитетных источниках. Уникальная коллекция Рубрикона включает как классические справочно-энциклопедические издания прошлого, составляющие часть культурного наследия страны, так и современные издания от энциклопедий до бизнес-справочников универсального, регионального и тематического характера.

Кроме основного – энциклопедического – блока, Рубрикон содержит разделы:

- «Бизнес-образование» – полнотекстовые электронные версии книг по бизнесу и менеджменту в рамках масштабного проекта электронной библиотеки.
- Календарь событий - ежедневно обновляемый список значимых событий, связанный

гиперссылками с текстами энциклопедических статей и ресурсами интернета.

- Кругозор – ежедневная тематическая подборка публикаций из научных и научно-популярных журналов по определенным для каждого дня темам.
- Каталог информативных ресурсов интернета, представленных в едином рубрикаторе параллельно с энциклопедическими статьями.
- Флаги мира (качественные изображения государственных флагов со ссылками на статьи о государствах).
- Магазин справочно-энциклопедической литературы (по партнерской программе с [Озоном](#)).
- Интерактивный информационный блок: «статья дня», «сейчас ищут», «сейчас читают», отражающий запросы пользователей в режиме реального времени.
- Раздел «О Рубриконе», включающий информацию о системе оплаты доступа.

Далее в данном сообщении будет рассматриваться в основном энциклопедическая, системообразующая и наиболее проработанная составляющая проекта Рубрикон.

2. Особенности публикации электронных версий изданий на Рубриконе

Публикацию словарно-энциклопедических изданий в рамках проекта Рубрикон отличают следующие характеристики:

- Тщательная корректура оцифрованных текстов.
- Сохранение структуры печатного издания (наличие более чем одного словника, приложений, указателей и др.).
- Сохранение основных особенностей графического оформления текста печатного оригинала (жирный шрифт, курсив, соотношение кеглей, разрядка).
- Воспроизведение таблиц, графиков, диаграмм.
- Воспроизведение специальных символов (букв с диакритическими знаками разных языков, а также математических, физических и других символов и формул).

Труды 5^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL2003, Санкт-Петербург, Россия, 2003.

- Высокое качество воспроизведения иллюстраций, в отдельных случаях лучшее, чем в печатном издании.
- Приведение некоторых типов данных к единообразному виду (напр. века, даты и т.п.).
- Корректное отображение сносок в статьях и возможность удобной навигации по ним.
- Преобразование внутрискриптовых ссылок в гипертекстовые.

Но Рубрикон с самого начала ориентировался не только на сохранение аутентичности электронных версий печатным оригиналам (для этого достаточно было бы выложить книги в одном из графических форматов), но на предоставление пользователю дополнительных преимуществ обращения к той же информации в электронном виде (в частности, «возможность проведения разнообразных и качественных информационных поисков, наличие простых и эффективных средств навигации в доступном пользователю информационном пространстве» [1])

С этой целью на Рубриконе применяется:

- Организация и хранение информации в виде баз данных.
- Хранение каждой словарно-энциклопедической статьи как единицы БД со своими атрибутами.
- Раздельное хранение текстов и иллюстраций, что позволяет организовывать множественные связи между ними и работать с иллюстрациями как с отдельным информационным массивом.
- Подписи к иллюстрациям хранятся и индексируются как особые виды текстов, что позволяет редактировать их и организовать поиск иллюстраций.
- При подготовке к публикации во многих изданиях раскрыты значимые сокращения для повышения релевантности полнотекстового поиска.
- Через имеющиеся инструменты работы с данными могут быть проставлены дополнительные гипертекстовые ссылки, в том числе и между разными изданиями (напр., такими перекрестными ссылками связаны между собой многие статьи в киноэнциклопедиях Рубрикона, что позволяет получить более полную и целостную картину кинематографического процесса – при том, что в печатных изданиях о кино издательства «Материк» отсутствуют даже внутренние ссылки).
- Информация может обрабатываться пользователем с помощью стандартных средств веб-браузера (выделение нужных фрагментов, копирование, экспорт в текстовые форматы и др.).

Конечно, все вышеперечисленное значительно увеличивает трудоемкость, и, следовательно, совокупную стоимость работ для разработчиков – но, безусловно, не только экономит время и усилия

пользователей, но и дает более объемную и многомерную информационную картину.

3. Сервисы Рубрикона

Пользователи Рубрикона имеют в своем распоряжении различные средства доступа и сервисы:

- поисковая система: поиск в одном, нескольких или всех изданиях Рубрикона с учетом морфологии русского языка - в текстах статей или только в их заглавиях; поиск слов и словосочетаний на английском языке; поиск иллюстраций по подписям к ним; поиск в интернете (в основном на страницах ресурсов, отраженных в БД каталога);
- группировка результатов поиска по рубрикам;
- навигация по словарям энциклопедий и разделам книг;
- навигация по иерархическому дереву рубрикатора, по которому распределены и энциклопедические статьи и описания интернет-ресурсов;
- гипертекстовые ссылки внутри и между энциклопедиями;
- ссылки типа «см. также» на одноименные статьи и приставочные ссылки на ресурсы интернета близкой тематики в энциклопедических статьях;
- формы для обратной связи с пользователями для внесения исправлений и дополнений в каждую статью;
- вспомогательные инструменты для организации работы со справочной информацией;
- форма поиска в энциклопедиях для держателей интернет-сайтов;
- виртуальная клавиатура для зарубежных пользователей.

При перемещении по каталогу, словарям и просмотре результатов поиска пользователю доступны дополнительные настройки. Текущее положение пользователя на сайте отображается в специальной навигационной строке в виде пути от главной страницы. По адресу www.rubr.ru находится «облегченная» версия сайта, содержащая только поисковую форму и элементы управления параметрами поиска.

4. Информационная система Рубрикона в целом

Функционирование ресурса обеспечивается информационной системой Рубрикона – комплексом программно-аппаратных средств на основе собственных разработок (Рис. 1).

Хранилище данных Рубрикона организовано на базе Microsoft SQL Server 2000. В нем размещены

энциклопедии, книги, рубрикаторы, описания сайтов и различная вспомогательная информация, задействованная в процессах обработки и отображения данных. Концептуально оно состоит из двух основных частей – **редактируемой базы данных**, предназначенной для хранения и редактирования информации, и **публикуемой базы данных**, оптимизированной для быстрого доступа к данным через сайт www.rubricom.com. Редактируемая база данных отличается от публикуемой как составом хранимых данных (например, информация о правах доступа редакторов, журнал учета изменений), так и более высокой степенью нормализации.

Программное обеспечение информационной системы сайта представляет собой совокупность программ и сценариев, использующих технологии HTML, JavaScript, SSI, CSS, XML, XSLT, ASP, COM+, SQL и т.д.

При публикации каждого нового издания наряду с применением стандартных процедур обработки информации может потребоваться решение специфических задач обработки и отображения информации, связанных с характером данного издания: представление частей издания в виде отдельных информационных блоков, обработка табличных и цифровых данных, выделение отдельных иллюстраций, опубликованных в книге

на листовых вкладках и привязка каждой к соответствующей статье, отображение различных видов приложений, указателей, глоссариев и др.

4.1. Подготовка данных.

Под подготовкой данных в Рубриконе понимается согласованный перевод текстовой и графической информации из разнородных источников в заданные цифровые форматы. Перерабатываемые данные включают:

- Тексты энциклопедических статей, главы книг, статьи из журналов, сопроводительные тексты.
- Иллюстрации и подписи к ним.
- Связи иллюстраций с текстами.
- Сноски в теле статей.
- Ссылки между статьями.

Источниками информации могут являться: бумажные носители (книги, журналы); иллюстративные материалы (фотографии, слайды, оригиналы рисунков); файлы издательских систем (Quark Xpress, Adobe PageMaker, Corel Ventura и т.д.); электронные публикации (компакт-диски, Web, PDF); текстовые файлы, базы данных и другие источники.

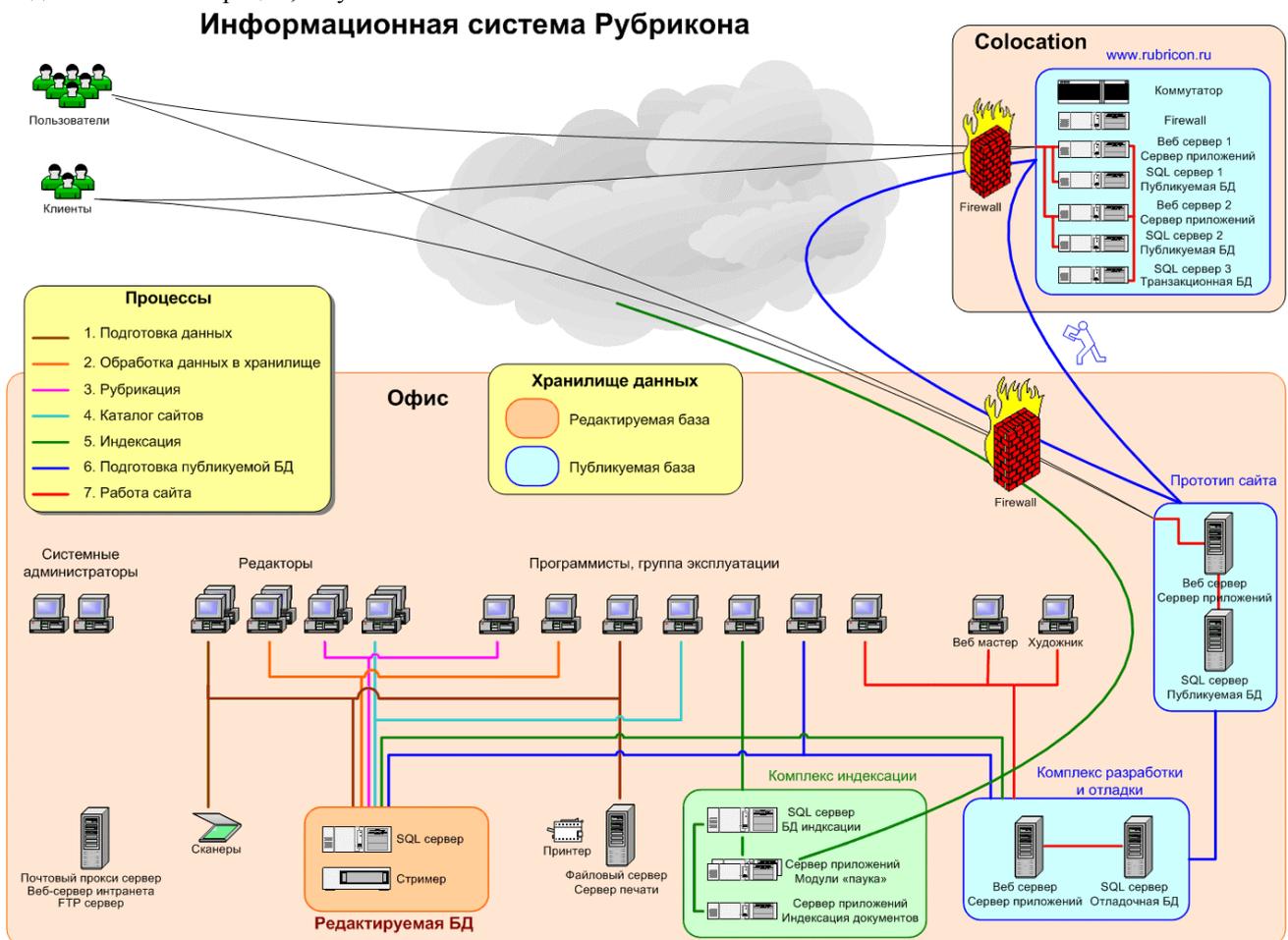


Рис. 1. Схема информационной системы Рубрикона

Полученные исходные материалы перерабатываются в соответствии с едиными нормативными требованиями и с соблюдением стандартов системы СИБИД [4], архивируются, паспортизируются и размещаются в хранилищах данных.

Все текстовые материалы проходят посимвольную обработку для обеспечения полнотекстового поиска. При этом возникают проблемы отображения текстов на языках с экзотической графикой, нестандартных символов, формул и т.п., которые обрабатываются специальным образом и хранятся в виде графических изображений, связанных с текстом данной статьи (т.н. in-line-графика), а отображаются непосредственно в «теле» текста статьи.

Подготовка данных состоит из нескольких процессов, выполняющихся как вручную, так и с использованием средств автоматизации, и полностью автоматических. Общее представление об этих процессах – см. Рис. 2 и Табл. 1.

4.2. Хранилище данных Рубрикона

Доступ к энциклопедиям и книгам в хранилище данных Рубрикона организован в виде специального рабочего места, называемого *Автоматизированное рабочее место редактора энциклопедий* (далее – АРМРЭ). Используется для проверки целостности импортированных в хранилище данных, их

тестирования и исправления ошибок, обнаруженных как при подготовке, так и после публикации. Обеспечивает все необходимые возможности редактирования размещенных в хранилище Рубрикона изданий и защиту от несанкционированного доступа к данным.

Объекты, с которыми позволяет работать АРМРЭ:

- статьи энциклопедий и словарей с их атрибутами;
- иллюстрации с их атрибутами;
- аннотированные ссылки на интернет-ресурсы;
- рубрикаторы (классификаторы).

Количество статей, иллюстраций, ссылок, количество рубрикаторов и узлов в каждом из них, а также количество одновременно обрабатываемых и хранящихся изданий – не ограничено.

АРМРЭ реализовано в архитектуре клиент-сервер. Хранилище данных организовано на базе реляционной СУБД Microsoft SQL Server 2000. Собственно АРМРЭ – это клиентское приложение, выполняющееся в среде Microsoft Windows 9x, 2000, XP. Его отличает дружественный эргономичный интерфейс, широко используемые средства повышения эффективности работы редактора (контекстные меню, технология drag&drop, расширенные операции копирования и вставки и др.).

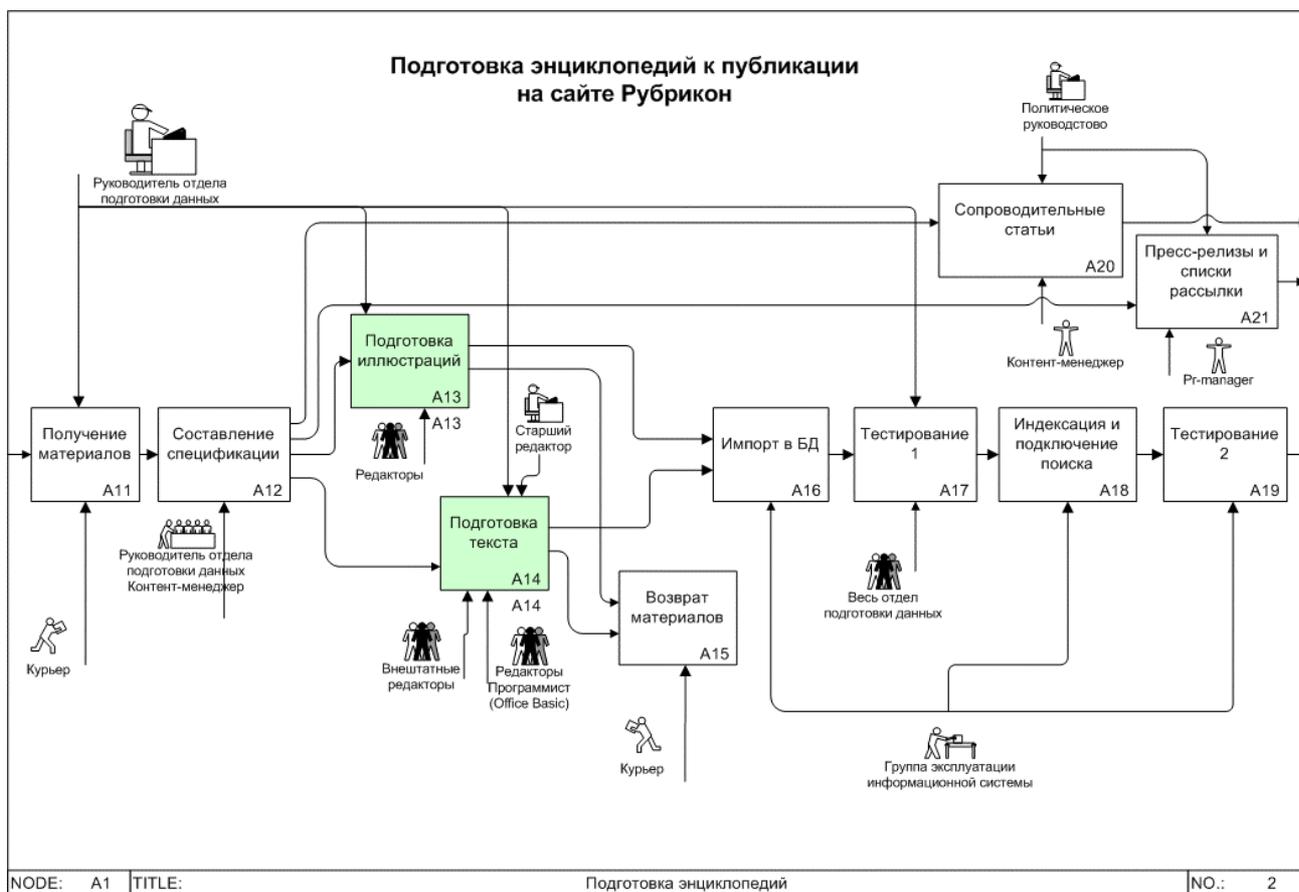


Рис. 2. Схема организации процессов подготовки изданий к публикации на Рубриконе

<i>Процесс</i>	<i>Описание</i>
Прием и паспортизация изданий	Получение исходных данных от правообладателей, определение и документирование особенностей каждого издания.
Перевод текстовых данных в документы Microsoft Word	Сканирование/распознавание текста, конвертация из файлов других форматов.
Вычитка и корректура	Вычитка и корректура текста издания, включая подписи к иллюстрациям, после оцифровки и распознавания.
Приведение текстовых данных к стандарту для помещения в базу данных	Набор процедур ручной, полуавтоматической и автоматической обработки документов (разбиение текста на словарные статьи, выделение словарных слов (заголовков статей), выделение ударных гласных, разметка для полуавтоматической простановки гипертекстовых ссылок, обработка таблиц и др.).
Перевод иллюстраций в стандартные цифровые форматы	Сканирование, ретуширование, конвертация в формат JPEG, GIF, автоматизированная подготовка пиктограмм.
Паспортизация иллюстраций	Сопровождение иллюстраций текстовыми описаниями и ссылками на связанные статьи.
Архивирование данных	Запись графических файлов и документов Microsoft Word на CD-R.
Импорт данных в хранилище	Конвертирование данных в форматы хранения и их размещение в базах SQL сервера.
Учет и планирование трудозатрат	Стандартизованная отчетность исполнителей работ.

Табл. 1. Описание процессов подготовки изданий к публикации

Основным средством доступа к статьям является упорядоченный по алфавиту список словарных слов. Статьи, представленные в списке, могут быть отобраны путем применения до 10 одновременно действующих фильтров (наименования энциклопедий, словарное слово или диапазон символов, дата последней редакции и ее авторы, наличие внутренних и внешних ссылок, иллюстраций и связей с выбранными узлами рубрикатора)

Если статья содержит ссылки на другие статьи, возможен переход по этим ссылкам, а также в любую из просмотренных и/или отредактированных в текущем сеансе работы статей. Сформированный и отсортированный список статей может быть экспортирован в документ Microsoft Excel.

Текст отображается в виде форматированного документа. Допустимо редактирование в двух режимах: в виде форматированного документа (режим WYSIWYG) и в виде HTML кода.

Гиперссылки на другие статьи отображаются и могут редактироваться непосредственно в тексте. Кроме того, выводится два списка связей с другими статьями: полный список статей, на которые ссылается данная статья и полный список статей, которые ссылаются на данную статью. Возможен переход на любую статью из этих списков.

Иллюстрации рассматриваются как самостоятельные объекты. Их атрибутами являются: собственно иллюстрация, ее

уменьшенная копия (пиктограмма), подписи (собственная подпись к иллюстрации, как к объекту в хранилище данных, и подпись к иллюстрации в данной статье могут отличаться). Каждая иллюстрация может быть привязана к произвольному количеству статей, относящихся как к одному, так и к разным изданиям, а каждая связь иллюстрации с новой статьей может иметь отдельную подпись. Подсистема работы с иллюстрациями позволяет производить операции просмотра и редактирования, осуществлять поиск иллюстраций по подписям к ним.

Форматы данных, публикуемых на Рубриконе, могут быть расширены и включать также аудио- и видео файлы, связанные с энциклопедическими статьями.

После завершения всех необходимых этапов работы в редактируемой базе данных производится подготовка *публикуемой базы данных*, которая состоит в передаче некоторого подмножества данных из одной группы таблиц с одним набором атрибутов в другую группу таблиц с другим набором атрибутов. Публикация данных на сайте представляет собой последовательный набор промежуточных публикаций, позволяющий обеспечить согласованное изменение данных и средств их обработки и отображения.

Разработанное в Рубриконе АРМ редактора может быть с успехом применено для создания на его основе эффективной издательской системы для

энциклопедического или многопрофильного издательства. Переход на хранение данных из файловой системы в формат базы данных с удобным интерфейсом и необходимыми средствами мониторинга, администрирования и защиты данных сродни переходу от печатной машинки к текстовому процессору.

4.3. Структурирование данных

Огромный информационный массив, аккумулированный Рубриконом, предъявляет повышенные требования к структурированию информации. В Рубриконе используется собственная классификационная система, разработанная с привлечением специалистов из РГБ.

С узлами рубрикаторов связываются статьи энциклопедий, «Кругозора» и интернет-сайты, книги. Информационная система Рубрикона может работать одновременно с произвольным количеством рубрикаторов без ограничения глубины вложенности и количества узлов в каждом из них. В настоящее время система фасетных рубрикаторов включает рубрикатеры тематический (более 9,5 тыс. рубрик, до 8 уровней иерархии), географический, этнический, специальных и общих типовых делений. Некоторые энциклопедии имеют также собственные рубрикатеры. Для описаний сайтов дополнительно применяются одноуровневые рубрикатеры и набор полей, отчасти коррелирующих с набором метаданных Дублинского ядра.

Рубрикатеры размещаются в хранилище данных Рубрикона, организованном на базе реляционной СУБД Microsoft SQL Server 2000. Для доступа к ним в архитектуре клиент-сервер разработано специальное *автоматизированное рабочее место редактора рубрикаторов*. АРМ редактора рубрикаторов – это инструмент для создания и полного редактирования древовидных классификационных структур, оно обладает встроенными средствами контроля уникальности данных и контроля доступа.

В режиме работы в АРМРЭ применяется ручная рубрикация для систематизации ресурсов интернета, а для быстрого рубрицирования энциклопедических статей разработана система автоматизированной рубрикации, реализованная в виде специального *рабочего места редактора лингвиста*.

В системе автоматизированной рубрикации применяются оригинальные эвристики, учитывающие специфику энциклопедических массивов информации и особенностей используемой системы рубрикаторов.

Система автоматизированной рубрикации была представлена на Международном семинаре по компьютерной лингвистике и ее приложениям Диалог-2002 [3].

На основе уже отрубрицированного массива энциклопедических статей создана первая версия

АРМ «энциклопедический конструктор информационно-поискового тезауруса», построенного в соответствии с ГОСТ 7.25-2001 [2].

В Рубриконе разработан также АРМ-конвертор для связывания рубрикаторов. Оно позволяет работать одновременно с двумя рубрикаторами, организовать связи между их рубриками типа «многие к одному» и транслировать связи документов от исходного рубрикатера к целевому.

Специальное АРМ предназначено для поддержания обратной связи с пользователями. Все исправления, дополнения и ссылки на профильные интернет-сайты, которые пользователи Рубрикона отсылают через формы обратной связи в статьях, аккумулируются в отдельной БД, связанной с хранилищем данных, что позволяет организовать в режиме реального времени процессы редактирования материалов Рубрикона и отображения результатов на сайте.

4.4. Индексирование и поиск данных

Рубрикон использует поисковый движок собственной разработки. Для его функционирования осуществляется предварительное индексирование документов, т.е. подготовка специальных данных (так называемых инвертированных списков или индексов), используемых для обеспечения быстрого поиска.

Системы индексации Рубрикона состоят из пакетов клиент-серверных приложений, работающих с общими очередями обрабатываемых документов. Очереди организованы на SQL сервере, модули индексации выполняются на серверах приложений.

Системы индексации реализованы в последовательно-параллельной архитектуре (модули обрабатывают данные последовательно, каждый модуль может быть загружен и выполняться в одном или нескольких экземплярах на одном или нескольких серверах приложений). Этим достигается масштабируемость и надежность систем.

Завершает индексацию процедура подключения вновь сформированных дополнений словарей и индексов к основной базе данных. Рубриконом реализован также модуль сбора и индексации HTML-страниц из интернета (т.н. «паук»).

Система поиска обнаруживает релевантные запросу документы и обеспечивает сортировку результатов, учитывая количество исковых слов в документе, а также их нахождение в особых зонах документа и совпадение словоформ. Кроме того, учитывается расстояние между словоформами в документе, их принадлежность одному предложению или абзацу. Поисковый запрос дополняется фильтрами, определяемыми пользователем, либо вытекающими из его текущего положения на сайте (поиск в отдельном издании, в группе изданий, поиск внутри тематической рубрики, поиск только в названиях статей). Также

имеется специальная форма поиска с группировкой результатов по тематическим рубрикам.

Модули, осуществляющие поиск, реализованы в виде COM+ объектов, с использованием языка структурированных запросов и хранимых процедур СУБД Microsoft SQL Server 2000. Постоянно ведутся работы по повышению производительности системы поиска и её эффективности.

5. Клиринг клиент — Рубрикон — правообладатель

В настоящее время доступ к подавляющему большинству ресурсов Рубрикона предоставляется пользователям по подписке. В проекте применяется система контроля доступа к ресурсам и клиринговая система собственной разработки, призванная обеспечить взаиморасчеты между клиентами Рубрикона, Рубриконом и правообладателями платных ресурсов. Основной задачей при выполнении клиринговых операций на сайте Рубрикон является предоставление клиентам и правообладателям достоверной и полной информации, на основании которой бухгалтерией Рубрикона осуществляются взаиморасчеты по предоставленным платным услугам. Для этого правообладатели регистрируются на сайте как клиенты Рубрикона (в специальной группе «Правообладатели») с предоставлением персонального счета. Взаиморасчеты происходят в соответствии с договорами с каждым правообладателем. Через свою персональную страницу правообладатель получает доступ к странице с детальной информацией о поступлении оплаты за услуги и начислении роялти за выбранный период.

6. Небольшой экскурс в историю проекта Рубрикон

В середине 2000 г. были начаты работы по проекту Рубрикон. Были заключены лицензионные договоры с издательствами «Большая российская энциклопедия», «Медицинская энциклопедия», «Русский язык» и начата переработка данных для публикации их в интернете. Перед разработчиками были поставлены довольно амбициозные организационные и технические задачи:

- За короткий срок опубликовать в интернете критическую массу словарно-энциклопедических материалов, чтобы занять ведущее положение среди энциклопедических ресурсов Рунета. Именно этим определялись предпочтения при выборе тех или иных технологических средств.
- Разработать информационную систему, которая включала бы технологические решения, обеспечивающие:
 - эффективную обработку оцифрованной информации (т.н. «фабрику контента»);
 - одновременный доступ к энциклопедическим материалам и ресурсам интернета;

- качественный полнотекстовый поиск по разнородной информации;
- систематизацию энциклопедических материалов и ресурсов интернета;
- организацию обратной связи с пользователями;
- эффективные механизмы работы с данными, помещенными в хранилище и мн. др.

При разработке интерфейса сайта команда Рубрикона старалась использовать и развивать опыт таких уже существовавших интернет-проектов, как энциклопедия «Британника», энциклопедия Кирилла и Мефодия, словари на сайте «Академик». Интерфейсы перечисленных сайтов имеют сходные «родовые» черты, что обуславливается сходством выполняемых ими функций – обеспечения удобного доступа и комфортной работы со словарно-энциклопедической информацией. Вместе с тем Рубрикон имеет некоторые особенности, присущие только ему (напр., настройки отображения результатов поиска, ссылки «см. также» к одноименным статьям других энциклопедий, приставленные ссылки на интернет-ресурсы, группировка результатов поиска по рубрикам).

7. Итоги и перспективы

По прошествии трех лет можно сделать вывод, что технологии и программное обеспечение, разработанные для проекта Рубрикон, а также построенные на их основе процессы, позволили за короткий срок подготовить к публикации и опубликовать в интернете крупные массивы энциклопедической и иной информации. Практически сразу после официального открытия сайта www.rubricon.com 4 апреля 2001 года Рубрикон занял лидирующее положение в русскоязычном интернете среди справочно-энциклопедических и образовательных ресурсов. В течение 2001 и 2002 года это было подтверждено наградами наиболее престижных конкурсов Рунета. То есть основная задача разработчиками успешно решена.

Вместе с тем на решение некоторых задач не хватило ресурсов, а другие были не вполне адекватно поставлены или требуют пересмотра с учетом стремительных изменений интернет-среды.

В частности, специалистами РГБ был разработан рубрикатор фасетного типа для классифицирования энциклопедических статей и ресурсов интернет. Но в Рунете уже существовали обширные и признанные каталоги интернет-ресурсов, «догонять и перегонять» которые не имело смысла. Трудозатраты на поддержку и развитие каталога оказались неоправданно высоки. Сейчас пользователь видит на экране только тематическую часть рубрикатора, а рубрики географические, этнические, типовые (общие и специальные) пока доступны только в рабочем месте редактора. Но ситуация может измениться, а программный модуль редактирования рубрикатора может быть использован как инструмент и для других задач.

Много усилий было затрачено также на создание работа-паука для обхода и сбора HTML-страниц интернета. Но его производительность оказалась очень высока, что привело к быстрому заполнению дискового пространства на серверах и генерации большого входящего трафика, поэтому пришлось отказаться от его эксплуатации на полную мощность. Возможна его настройка для решения более узких и специализированных задач.

Ставившаяся изначально задача одновременного доступа к лучшим энциклопедиям и лучшим сайтам интернета оказалась в целом не очень востребованной пользователями, потому что поиск в интернете и поиск словарно-энциклопедической информации обуславливаются, как правило, разными целями и интересами пользователей и происходят не одновременно. Подобный сервис предоставляется в партнерском проекте «Яндекс-энциклопедии», основу которого заложили энциклопедические БД Рубрикона.

Поэтому в настоящий момент Рубрикон сосредотачивает усилия на создании и развитии энциклопедической и учебной интернет-библиотеки, непрерывно совершенствуя технологии обработки оцифрованной информации и предоставляя пользователям более качественные сервисы.

В ближайшей перспективе редакция Рубрикона предполагает:

- Опубликовать новую версию дизайна, который позволит «облегчить» главную страницу, более четко структурировать разделы сайта и сделать навигацию по нему более «прозрачной».
- Переработать и дополнить биллинговую систему, которая даст возможность оплачивать услуги не только в «пакетном» режиме, но и «заполняя» свою индивидуальную корзину, вводить разные тарифные планы, а также оплачивать доступ к отдельным документам.
- Расширить использование XML-технологий и XSLT-шаблонов в динамических элементах сайта с целью повышения его функциональности.
- Подключить язык запросов.
- Добавить к рабочему месту редактора модули, позволяющие комфортно работать с новыми типами и элементами информационных ресурсов.
- Разработать и подключить систему классификации иллюстраций как самостоятельных объектов БД.
- Доработать и апробировать в промышленном режиме АРМ «конструктор энциклопедического тезауруса» и АРМ визуализации связей объектов.
- Найти дополнительные источники финансирования, которые позволят решить все вышеперечисленные задачи и воплотить другие интересные идеи.

Имеющиеся в Рубриконе технологии и программное обеспечение, а также накопленный опыт разработки и управления процессами могут быть использованы при создании масштабируемых комплексов обработки и поиска данных как для общедоступных интернет-проектов, особенно в сфере образования, науки и культуры, так и для корпоративных информационных систем. Энциклопедический контент может быть предоставлен другим разработчиками лингвистического обеспечения информационных систем.

Литература

- [1] Горный Е., Вигурский К. [Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы](#) (статья из сборника «Интернет и российское общество») 2002.
- [2] ГОСТ 7.25-2001. Система стандартов по информации, библиотечному и издательскому делу. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав и форма представления. Введ. с 01.07.2002 г.
- [3] Попов И. В., Руссова Н. В., Скотников А. П., Фролкина Н. А.. Автоматизированная рубрикация энциклопедических статей http://www.dialog-21.ru/archive_article.asp?param=7640&y=2002&vol=6078). 2002
- [4] СИБИБД. [Система стандартов по информации, библиотечному и издательскому делу.](#)

The electronic encyclopedic collection Rubricon

© Popov I. V., Frolkina N. A.

«Twincom Business» (provides development and maintenance of the site www.rubricon.com)

content@rubricon.com

This work presents electronic collection of Russian encyclopedias, dictionaries and other reference titles RUBRICON (www.rubricon.com). Paper contains the general overview of RUBRICON information system and used technologies.