

ОРГАНИЗАЦИЯ ИНФОРМАЦИОННОЙ РАБОТЫ

УДК 002:[004.738.5:574]

В. Г. Шамаев

Особенности информационного обеспечения и программно-технологический комплекс подготовки интернет-энциклопедии по экологии

Описывается технология создания интернет-энциклопедии по экологии. Программно-технологический комплекс Энциклопедии состоит из двух основных функциональных блоков — блока подготовки базы данных Энциклопедии к публикации в Интернете и блока, обеспечивающего доступ пользователя к материалам Энциклопедии, размещенным на веб-сайте. Блок подготовки обеспечивает ввод, накопление и хранение статей Энциклопедии, их редактирование, снабжение иллюстрациями, перекрестными ссылками а также данными для реализации механизмов поиска и вспомогательной информацией. Второй блок обеспечивает сервисные возможности при пользовании Энциклопедией в Интернете.

Параллельно с реализацией этого проекта интернет-энциклопедии начата подготовка к изданию "Российской экологической энциклопедии" в печатном виде.

ВВЕДЕНИЕ

На одном из сайтов в Интернете сообщается, что "эпоха многотомных энциклопедий давно канула в Лету — этот факт известен даже школьникам". Сомнительное на сегодняшний день утверждение — даже школьники видят в книжных магазинах десятки энциклопедических изданий по всем отраслям естественных, технических, гуманитарных, экономических и других наук. Их выпускают как традиционные издательства "Большая российская энциклопедия" (ранее — "Советская энциклопедия"), "Мысль", "Просвещение", "Педагогика" и др., так и новые такие, как "Терра", "Эксмо", "АСТ", "Инфра-М" и др. Таким образом, мы еще находимся внутри "Галактики Гутенберга" по определению Г. М. Маклюэна*, но буквально в ближайшие одно-два десятилетия ситуация изменится.

Даже сейчас, когда ученики младших классов спрашивают учителя, что такое энциклопедия, он показывает и толстые тома какой-нибудь энциклопедии, что у него под рукой, и CD-ROMы, на которых записаны не только энциклопедические статьи по разным областям науки, техники и т. д., но и огромное количество иллюстраций и видео-фрагментов. А поскольку сведения, находящиеся на CD-ROMе, показываются на компьютере, то естественен и переход на сайты в Интернете, где можно также получить необходимые сведения — например, энциклопедии на

сервере Yandex (<http://encycl.yandex.ru>); энциклопедии на сайте Российской национальной библиотеки (http://www.nlr.ru/res/inv/ic_www/index.php); словари и энциклопедии on-line (<http://dic.academic.ru>); энциклопедический ресурс Интернета "Рубрикон" (<http://www.rubricon.com>); энциклопедия Microsoft Encarta (<http://encarta.msn.com>); Britannica online (<http://www.britannica.com>), Wikipedia (<http://www.wikipedia.org>) и др. Наиболее продвинутые виртуальные энциклопедии изобилуют большим количеством копий исторических документов, аудио- и видеофрагментами различных событий, а также подробными географическими атласами, звездными картами и т. д.

Виртуальные энциклопедии в своем большинстве как бы живут во времени, дополняясь новыми статьями, сведениями, модернизируясь в оформлении и, таким образом, превосходя в информативности свои оффлайн-аналоги, если они существуют.

И то, что находится в Интернете, им, 7-9-летним школьникам, оказывается более понятным. Они, в отличие от нас, уж точно находятся в XXI веке — веке информации и с детства знакомы с компьютером. Толстые тома энциклопедий, словарей и справочников для них менее удобны и менее понятны — ведь нужные сведения всегда можно найти в Интернете.

Отметим, что создание электронных и интернет-энциклопедий по различным областям науки

* Г. М. Маклюэн (1911-1980) — канадский ученый и философ. Получил большую известность благодаря исследованию влияния средств массовой коммуникации на развитие общества.

и техники является одним из основных направлений развития современных информационных технологий. Возможность поиска статей, удобный интерфейс, использование гипертекстовых ссылок облегчают работу пользователей с информацией, содержащейся в электронном издании. Электронное представление позволяет оперативно дополнять и изменять наполнение энциклопедии, а визуальные возможности многоцветного изображения иллюстрируют статьи энциклопедии не только черно-белыми рисунками и схемами, но и цветными фотографиями, схемами, диаграммами и мультимедийными данными.

КОНЦЕПЦИЯ ЭКОЛОГИЧЕСКОЙ ЭНЦИКЛОПЕДИИ

Этапной работой для современной экологии было создание “Экологического энциклопедического словаря” [1] на основе привлечения большой группы составителей с включением объемных статей. Идея словаря возникла в Ассоциации “Экология и мир” в 1992 г. и принадлежала академику С. П. Залыгину. Она появилась на волне экологического движения конца 1980-х и начала 1990-х гг. в борьбе с идеями поворота сибирских рек и в борьбе за спасение озера Байкал и Аральского моря. В 1993 г. проект словаря был представлен в федеральную программу “Экологическая безопасность России”. К работе над созданием словаря были привлечены около 50 ученых из 14 научных институтов и учебных заведений и четырех общественных научных организаций России. Словарь вышел в свет в начале 1999 г.

К десятилетию со дня окончания работы над “Экологическим энциклопедическим словарем” руководителем предыдущего проекта профессором К. С. Лосевым был подготовлен крупный массив, порядка 12 тыс., терминов по экологии, в котором представлен практически весь спектр экологических проблем, включая охрану природы и экологию человека. Это позволило поставить вопрос о создании экологической энциклопедии. Анализ словаря показал, что это может быть издание ориентировочно в шести томах объемом около 50 авторских листов каждый. Еще одним стимулом для создания энциклопедии были новые научные результаты и научные направления, появившиеся в сфере экологии и сопредельных отраслях знания, что требовало обновления содержания ряда терминов и включения новых терминов. В связи с этим была разработана концепция “Экологической энциклопедии”:

1) включение максимально возможного количества терминов по экологии и смежным областям знания;

2) приведение в статьях в максимально возможном объеме примеров, табличных данных, иллюстраций;

3) учет новой терминологии и новых научных достижений;

4) помещении в последнем томе англо-русского и русско-английского словаря терминов, расширенного по сравнению с числом терминов в Энциклопедии.

Создание энциклопедии — это огромный объем работы, который требует финансирования, но ее создание — это и фундаментальная работа, которая сохраняет заложенную в нее информацию на многие десятки лет как фундаментальная научная информационная система. В связи с этим был подготовлен конкурсный проект на соискание гранта РФФИ для подготовки в первую очередь электронной версии экологической энциклопедии. Проект был одобрен экспертами, и с 2004 г. начата работа по созданию Энциклопедии. Очевидно, что электронная энциклопедия — это содержательная и программная части и эта вторая часть и обсуждается в данной работе.

Технология создания интернет-энциклопедии включает несколько этапов [2]. Вопросы подготовки исходного материала — текстов статей, иллюстраций и таблиц — отражены в ранее опубликованной статье [3], и мы на этом останавливаться не будем. Здесь же рассмотрим проблемы, касающиеся создания программной оболочки электронной энциклопедии. Во-первых, это — формирование предметной области и связей между объектами, объединяющих отдельные статьи в единое целое: формирование списка ключевых слов для поиска, составление оглавления, привязка иллюстраций, установление перекрестных ссылок между статьями и т. п. Во-вторых, выбор платформы для информационного обеспечения и разработка реляционной схемы базы данных (БД). В-третьих, разработка механизмов наполнения базы данных. В-четвертых, создание программного обеспечения для функционирования электронной энциклопедии. При публикации в сети Интернет существенными являются также вопросы защиты данных от несанкционированного изменения и удаления.

Предметная область БД Энциклопедии включает объекты двух типов — статьи и иллюстрации. Статьи являются основным (образующим) объектом БД. Статья идентифицируется по заголовку, являющемуся термином Энциклопедии. Данные, по которым проводится обработка (сортировка, формирование ссылок и алфавитного указателя, получение ключевых слов), вынесены в отдельные поля. Таким образом, одним из основных отношений нашей реляционной модели является отношение “Статья” со схемой, включающей следующие атрибуты:

- заголовок статьи энциклопедии — термин;
- перевод термина на английский язык;
- синонимы термина;
- сокращение термина, например, “АЭС” для словосочетания “атомная электростанция”;
- полный вариант для сокращения, например, “биологически допустимый выброс” для “БДВ”;
- статьи-ссылки, т. е. заглавия статей, на которые ссылается данная статья с использованием конструкций “см.” или “см. также” (например, статья “Внеземная жизнь” имеет ссылку “см. также” на статью “Возникновение жизни”);
- дополнение к заголовку статьи (примером могут служить случаи, когда в заголовках статей, посвященных ученым, указаны годы жизни. Так, в статье с заголовком “Базилевич Наталья Ивановна” — дополнение — “1910–1997 гг.”);

- текст статьи, представляющий собой фрагмент html-страницы;
- ключевые слова для статьи.

Статья как информационный модуль собирается из этих атрибутов по определенному шаблону, принятому для Энциклопедии. В интернет-энциклопедии активно используется механизм гипертекстовых ссылок. HTML-ссылки реализованы не только для случая, когда из одной статьи дается ссылка на другую с использованием конструкций “см.” или “см. также” с указанием полного заглавия статьи-ссылки, но и для случая, когда упоминание термина встречается в середине текста статьи, причем не обязательно в той форме, как он (термин) представлен в заглавии соответствующей статьи. В этом случае в качестве чувствительной области гипертекстовой ссылки может быть выбран любой фрагмент текста, в том числе лишь косвенно связанный с заглавием статьи-ссылки. Вторым объектом предметной области БД являются иллюстрации к статьям. В качестве иллюстраций могут выступать не только рисунки в форматах JPEG, BMP и GIF, но и дополнительный текст сложной структуры (как правило, таблицы большого размера). При просмотре статьи интернет-энциклопедии ее иллюстрации визуализируются в отдельном окне веб-браузера. Это позволяет “разгрузить” основной текст статьи, сделать его более наглядным и удобным для просмотра. Вместе с тем небольшие рисунки могут быть размещены непосредственно в тексте статьи. Таким образом, другим ключевым отношением нашей реляционной модели является отношение “Иллюстрация” со схемой, включающей:

- имя файла с иллюстрацией;
- графическое изображение;
- подпись под иллюстрацией;
- размеры изображения;
- заставку, появляющуюся на месте ненайденного или незагруженного изображения;
- дополнительный текст — html-страницу.

Накопление данных для наполнения БД Энциклопедии (загрузка статей и иллюстраций к ним, установление связей между ними, формирование списка ключевых слов, внесение правки и пр.) выполняется на внутреннем сервере ВИНИТИ РАН, не доступном извне. Для этого под СУБД Microsoft SQL Server 2000 создана база данных, имеющая структуру, аналогичную структуре БД интернет-энциклопедии, но включающая дополнительные элементы (служебные атрибуты в отношениях, ограничения, триггеры и пр.), используемые для обеспечения ссылочной целостности данных, поддержки бизнес-правил*, отслеживания внесения изменений и технологического состояния записей. Для ведения этой БД создан специальный программный модуль [4]. Статьи Энциклопедии загружаются из файлов, подготовленных в Microsoft Word, html-файлов, а также оригинал-макета в издательской системе Ventura 2.0, в которой был подготовлен для печати “Энциклопедический экологический словарь”. Также предусмотрена передача из БД текстов статей для публикации в виде печатного издания, которая реализована как выгрузка

в QuarkXPress. Эта подсистема подробно описана в [5]. Для передачи данных электронной энциклопедии из базы данных на СУБД Microsoft SQL Server 2000 в базу данных на СУБД MySQL, использующуюся для хранения Энциклопедии на веб-сайте, разработаны соответствующие механизмы, включающие выгрузку данных из исходной БД с определением состава и объема выгружаемой информации и загрузку данных в БД MySQL. Для передачи статей может использоваться как CSV-файл, так и пакетный файл SQL-команд. Как показала опытная эксплуатация, использование SQL-скриптов предпочтительнее. Все необходимые изменения выполняются в исходной БД, а затем измененный фрагмент данных переносится в БД интернет-энциклопедии. Для этого реализован механизм передачи выбранных статей, дополненных необходимой информацией из подчиненных таблиц. Хранение исходной БД электронной энциклопедии на внутреннем сервере ВИНИТИ, не доступном извне, обеспечивает режим сохранности исходной информации.

При создании интернет-энциклопедии мы ориентировались на наиболее полное использование возможностей электронного представления. В результате был выбран вариант публикации с минимальным числом статических HTML-страниц — почти все html-страницы формируются с динамически PHP-скриптами на основе данных из БД. Статической является только стартовая страница, содержащая эмблему интернет-энциклопедии, форму для поиска по ключевым словам, а также русский и латинский алфавиты для доступа к статьям энциклопедии по оглавлению. Все остальные страницы, включая листы оглавления, формируются динамически. С помощью PHP-скриптов реализованы следующие функции: поиск статей энциклопедии по заданному фрагменту ключевого слова или словосочетания; визуализация списка статей, отобранных при поиске; формирование html-страницы — статьи энциклопедии и страниц просмотра иллюстраций к статье по данным из БД; формирование оглавления. При проектировании БД, содержащей исходные данные и с использованием которой проводится основная работа по формированию электронной энциклопедии, была задействована большая часть возможной СУБД Microsoft SQL Server, что позволило сократить объем прикладного программирования и унифицировать обращение к БД. Разработанная система, включающая базу данных и программный комплекс, может использоваться также для создания интернет-энциклопедий в других областях знания, для представления информации в которых достаточно средств HTML, т. е. с минимальным количеством сложных математических формул, либо без необходимости поиска по ним [6].

ПРОГРАММНО-ТЕХНОЛОГИЧЕСКИЙ КОМПЛЕКС

Программно-технологический комплекс интернет-энциклопедии состоит из двух основных функциональных блоков — блока, обеспечивающего доступ пользователя к материалам Энциклопедии,

*Бизнес-правила — это процедуры управления, которые указывают, как клиент получает доступ к данным на сервере.

размещенным на веб-сайте, и блока подготовки базы данных энциклопедии к публикации в Интернете.

Доступ пользователя к материалам Энциклопедии, размещенным на веб-сайте, разработан в двух вариантах — статическом и динамическом.

В статическом варианте исполнения материалы Энциклопедии на веб-сайте представлены в виде совокупности статических html-страниц, связанных гипертекстовыми ссылками. Этот вариант реализации размещения Энциклопедии в Интернете предъявляет минимальные требования к программному обеспечению веб-сайта.

На основной странице сайта Энциклопедии размещается алфавитный указатель статей Энциклопедии. С основной страницы осуществляется переход на страницу списка статей, содержащую ссылку на нужную статью.

В динамическом варианте исполнения материалы Энциклопедии на веб-сайте представлены в виде совокупности данных, загруженных в СУБД MySQL (рис. 1).

Поиск нужной статьи Энциклопедии и вывод ее на экран осуществляется с помощью скриптов, написанных на PHP 4.3.0. Такой вариант размеще-

ния Энциклопедии в Интернете предъявляет более высокие требования к программному обеспечению веб-сайта, однако пользовательский интерфейс в этом случае гораздо удобнее и информативнее, что, учитывая когнитивные аспекты при разработке информационных продуктов, надеемся, помогает пользователям в поиске нужных статей. Ниже описан вариант реализации демонстрационной версии Энциклопедии на научном портале ВИНТИ РАН.

Кроме алфавитного указателя статей предусмотрен поиск статей Энциклопедии по ключевым словам. В качестве ключевого слова может использоваться название статьи энциклопедии (термин), перевод названия на английский язык (или синоним данного термина). Реализованы несколько режимов поиска — поиск совпадения термина целиком, поиск совпадения введенной строки с началом термина, или поиск совпадения введенной строки с любой частью термина.

Результаты поиска выдаются на экран в виде списка статей или статьи, если термин задан исчерпывающе (рис. 2). Пользователю сообщается также общее число найденных статей, в соответствии с заданным режимом поиска.

Всероссийский Институт Научной и Технической Информации

Автоматика География Лингвистика Металлургия Физика Экономика
 Астрономия Геология Математика Механика Химия Электротехника
 Биология Информатика Машиностроение Транспорт Экология Энергетика
 Проекты

| ГСНТИ | e-Земля | Безопасность | Живые системы | Открытые базы | Диалог России-ЕС | Нанотехнологии |

Информационные ресурсы Организации Персоналии Публикации Контакты Библиография	Экологический словарь Энциклопедический словарь по экологии
	Предлагаем Вам почитать демонстрационную версию энциклопедии по экологии. Энциклопедия включает более 5 тысяч терминов.
	Оглавление
	Index

Поиск по ключевому слову

Введите термин (или его часть...):

Режим поиска:
 Поиск на совпадение...
 термина целиком
 с началом термина
 заданной строки с любой частью термина

Всероссийский институт Научной и Технической Информации
 Все права защищены и охраняются законом. © 2004-2006

Рис. 1. Вид начальной страницы Энциклопедии в динамическом варианте



[Автоматика](#) [География](#) [Лингвистика](#) [Металлургия](#) [Физика](#) [Экономика](#)
[Астрономия](#) [Геология](#) [Математика](#) [Механика](#) [Химия](#) [Электротехника](#)
[Биология](#) [Информатика](#) [Машиностроение](#) [Транспорт](#) [Экология](#) [Энергетика](#)

Поиск термина 🔍

Поиск по ключевому слову АВТОМАТИЧЕСКАЯ КЛИМАТИЧЕСКАЯ СТАНЦИЯ (по полному совпадению)

Результат поиска по ключевому слову АВТОМАТИЧЕСКАЯ КЛИМАТИЧЕСКАЯ СТАНЦИЯ

Найдена 1 статья

Номер	Статья
1	АВТОМАТИЧЕСКАЯ КЛИМАТИЧЕСКАЯ СТАНЦИЯ

Всероссийский Институт Научной и Технической Информации
 Все права защищены и охраняются законом. © 2004-2006

Рис. 2. Вид страницы выдачи по результатам поиска (список найденных статей)

Всероссийский Институт Научной и Технической Информации



[Автоматика](#) [География](#) [Лингвистика](#) [Металлургия](#) [Физика](#) [Экономика](#)
[Астрономия](#) [Геология](#) [Математика](#) [Механика](#) [Химия](#) [Электротехника](#)
[Биология](#) [Информатика](#) [Машиностроение](#) [Транспорт](#) [Экология](#) [Энергетика](#)

Статья 🔍

АВТОМАТИЧЕСКАЯ КЛИМАТИЧЕСКАЯ СТАНЦИЯ

АВТОМАТИЧЕСКАЯ КЛИМАТИЧЕСКАЯ СТАНЦИЯ – automation climatic station – станция, на которой автоматически наблюдают и регистрируют метеорологические явления по специальной программе, необходимой для решения климатических задач и передачи полученных данных.

Ключевые слова: AUTOMATION CLIMATIC STATION, АВТОМАТИЧЕСКАЯ КЛИМАТИЧЕСКАЯ СТАНЦИЯ

Всероссийский Институт Научной и Технической Информации
 Все права защищены и охраняются законом. © 2004-2006

Рис. 3 Вид страницы выдачи текста статьи

Через механизм гиперссылок из списка найденных статей можно открыть текст соответствующей статьи (рис. 3).

Сам текст статьи также может содержать гипертекстовые ссылки на другие статьи энциклопедии. Кроме того, текст статьи может содержать также ссылки на иллюстрации, которые обычно открываются в отдельном окне (как уже сообщалось выше, небольшие иллюстрации могут помещаться в тексте статьи)

РЕЛЯЦИОННАЯ МОДЕЛЬ БАЗЫ ДАННЫХ ДЛЯ ПОДГОТОВКИ ЭКОЛОГИЧЕСКОЙ ЭНЦИКЛОПЕДИИ

База данных подсистемы подготовки Экологической энциклопедии основана на СУБД Microsoft SQL Server 2000 и состоит из основного отношения, содержащего статьи Энциклопедии, отношения с

иллюстрациями к статьям, отношения связей между статьями и иллюстрациями или между двумя статьями, отношения ключевых слов, отношения связи между статьей и рисунком, а также отношения для обеспечения быстрого поиска по ключевым словам. Для формирования поисковых отношений созданы 3 хранимые процедуры.

Для контроля изменений все отношения снабжены атрибутами, отражающими время последнего изменения записи и имя пользователя с административными функциями, высшего изменения. Эти атрибуты поддерживаются автоматически при создании и изменении записи.

Основное отношение — статьи энциклопедии

Статьи энциклопедии хранятся в отношении ARTICLES. Полный список атрибутов отношения приведен в табл. 1.

Атрибуты и их описание в отношении
ARTICLES

Атрибуты	Описание
id_art	Идентификатор статьи
hat_rus	Название статьи
hat_eng	Перевод названия
article	Текст статьи
rubr_code	Рубрикационный шифр статьи
keywords	Ключевые слова
len_article	Длина текста статьи (вычисляемое поле)
upd_date	Дата и время внесения последнего изменения
upd_user	Пользователь, внесший изменение
synonyms	Синоним
ref_hat	Название статьи, на которую имеется ссылка
status_rec	Технологический статус записи
n_pr	Количество абзацев, подсчитанное при загрузке
comment	Комментарий
sort_code	Порядковый номер статьи
short_hat	Сокращение термина (названия статьи)
full_hat	Полное название термина для сокращения
cut_from_hat	Фрагмент исходного названия статьи, вырезанный для приведения названия к принятому виду. Добавляется к названию при визуализации
id_file	Идентификатор файла, из которого загружена запись

Таблица 2

Атрибуты и их описание в отношении
PICTURES

Атрибуты	Описание
img_name	Идентификатор изображения
picture	Изображение (BLOB-поле)
size_picture	Размер изображения в байтах — вычисляемое поле
pict_h	Размер изображения в пикселях (по вертикали)
pict_w	Размер изображения в пикселях (по горизонтали)
img_format	Формат изображения
descript_pict	Описание изображения. Используется как подпись под рисунком
upd_date	Дата и время внесения последнего изменения
upd_user	Пользователь, внесший изменение
hat_pict	Название рисунка. Используется как заголовок ссылки в html-формате и как заглавие в соответствующем файле, а также выводится вместо рисунка, когда изображение не подгружается
alter_text	Дополнительный текст. Используется как подпись (пояснение) к рисунку, если оно не помещается в поле descript_pict, а также может использоваться отдельно (без рисунка)
len_alt_text	Размер дополнительного текста в байтах — вычисляемое поле

*BLOB-поля отличаются от полей другого типа. Реально BLOB-поле имеет несколько подтипов (sub-type). Знание подтипа BLOB-поля существенно при создании приложения для работы с базами данных, которые включают в себя BLOB-поля. BLOB-поля могут быть трех подтипов: подтип 0, подтип 1 (два встроенных подтипа) и пользовательский подтип. Подтип 0 BLOB-поля создается при выполнении команды CREATE, когда подтип не определен. Для ясности, в синтаксисе SQL все же рекомендуется явно указывать, что BLOB-поле относится к подтипу 0. Данный подтип BLOB-поля используется для хранения бинарных данных. СУБД не проводит никакого анализа хранимых данных, она просто сохраняет данные в BLOB-поле байт-за-байтом. Наиболее частое применение BLOB-полей в приложениях Windows — хранение двоичных данных изображения. Для этой цели подходит или BLOB-поле подтипа 0, или BLOB-поле пользовательского подтипа.

Отношение иллюстраций

Рисунки, иллюстрирующие статьи, хранятся в отношении PICTURES. Сами иллюстрации хранятся, как правило, в JPEG-формате, но в базу заносятся в битовом виде (BLOB-поле*). Полный список атрибутов отношения приведен в табл. 2.

Отношение связей

Отношение связей REF_ARTICLE, схема которого приведена в табл. 3, содержит связи между статьей и рисунком, использующимся в ней в качестве иллюстрации, или между двумя статьями, в одной из которых дается ссылка на другую. В качестве ссылок берутся идентификаторы записей.

Таблица 3

Атрибуты и их описание в отношении
REF_ARTICLE

Атрибуты	Описание
id_art	Идентификатор статьи
img_name	Идентификатор изображения, иллюстрирующего статью id_art
id_ref	Идентификатор статьи, на которую ссылается статья id_art
upd_date	Дата и время внесения последнего изменения
upd_user	Пользователь, внесший последнее изменение

Отношение ключевых слов

Поиск статей по ключевым терминам осуществляется следующим образом. Термин ищется в списке ключевых слов. По найденному коду (одному или нескольким, если поиск осуществляется не по полному совпадению) в инверсном списке отбираются идентификаторы статей. А затем по списку идентификаторов статей (если такие статьи были найдены) выдается информация о статьях в требуемом виде. Список ключевых терминов содержится в отношении KEY_WORDS, схема которого приведена в табл. 4.

Таблица 4

Атрибуты и их описание в отношении
KEY_WORDS

Атрибуты	Описание
d_key_word	Код ключевого слова (термина)
key_word	Ключевое слово (термин)
upd_date	Дата и время внесения последнего изменения
upd_user	Пользователь, внесший последнее изменение

Инвертированный список ключевых слов (терминов) для статей

В отношении INV_ARTICLE содержатся связи между статьями из отношения ARTICLES и ключевыми словами из списка ключевых слов отношения KEY_WORDS. Таким образом, отношение INV_ARTICLE представляет собой инвертированный список ключевых слов для статей. Возможно использование этого отношения как для получения списка статей, для которых выбраный термин является ключевым словом, так и для получения списка ключевых слов для выбранной статьи. Отношение спроектировано с возможностью поддержки полнотекстового поиска и поиска по расстоянию между вхождениями двух лексем, однако в настоящее время эта возможность не реализована на уровне программного обеспечения. Список ключевых терминов содержится в отношении INV_ARTICLE, схема которого приведена в табл. 5.

Таблица 5

Атрибуты и их описание в отношении INV_ARTICLE

Атрибуты	Описание
id_art	Идентификатор статьи
id_key_word	Код ключевого слова (термина)
n_par	Номер абзаца, в котором встретилось слово
n_phr	Номер предложения в абзаце
n_wrd	Номер слова в предложении
upd_date	Дата и время внесения последнего изменения
upd_user	Пользователь, внесший последнее изменение

РЕЛЯЦИОННАЯ МОДЕЛЬ БАЗЫ ДАННЫХ ДЛЯ ХРАНЕНИЯ МАТЕРИАЛОВ ЭКОЛОГИЧЕСКОЙ ЭНЦИКЛОПЕДИИ НА ВЕБ-САЙТЕ

Хранилище данных для веб-публикации

В качестве хранилища публикуемых данных используется база данных СУБД MySQL. Выбор обусловлен выбором платформы (ОС семейства Unix и веб-сервер Apache), а также декларируемой скоростью работы при относительно небольших объемах. Сохранность информации и строгость контроля бизнес-правил внутри СУБД в данном случае не играют существенной роли, так как исходные данные хранятся в БД под СУБД Microsoft SQL Server 2000, обеспечивающей достаточную защиту данных и необходимый уровень контроля целостности данных. В БД под MySQL не предполагается вносить каких-либо изменений. Данные вносятся и корректируются в исходной базе данных, которая описана в предыдущем разделе. Затем данные переносятся в БД под СУБД MySQL. При необходимости предполагается проводить повторную загрузку, полную или частичную. Изображения, иллюстрирующие статьи Энциклопедии, в БД под MySQL не вносятся, а хранятся в виде отдельных файлов.

Структура базы данных

База данных, используемая для хранения Энциклопедии на веб-сайте, основана на СУБД MySQL

и состоит также из пяти отношений. Состав отношений в БД под СУБД MySQL в основном совпадает с набором отношений БД под СУБД Microsoft SQL Server. Структура отношений также имеет много общего в двух БД под разными СУБД. Однако есть и отличия: в БД, основанной на СУБД MySQL, отсутствуют технологические атрибуты, такие, как пользователь и дата последнего изменения записи, вспомогательные атрибуты, например, количество абзацев в статье, атрибут для записи комментария, размер текста статьи, а также резервные атрибуты, которые были созданы в исходном отношении, но пока не используются. В табл. 6 приведены все отношения базы данных, используемой для веб-публикации статей Энциклопедии.

Таблица 6

Список отношений БД

Атрибуты	Описание
ARTICLES KEY_WORDS INV_ARTICLE	Статьи энциклопедии Список ключевых слов Инвертированный список -- соответствие ключевых слов статьям
ARTICLE_REF_HAT_VIEW	Таблица ссылок от статьи на статьи (связь "один-ко-многим" в отношении ARTICLES)
PICTURES	Иллюстрации статей

Процедура передачи данных Экологической энциклопедии из БД системы подготовки в БД, используемую для хранения Энциклопедии на веб-сайте

Передача данных из одной СУБД в другую осуществляется посредством текстовых файлов, формируемых программно после того, как завершен очередной этап подготовки статей Энциклопедии к публикации. Передача данных происходит в два этапа: выгрузка данных в текстовые файлы и загрузка данных из текстовых файлов в базу данных под СУБД MySQL. Кроме того, рисунки, иллюстрирующие статьи Энциклопедии, выгружаются из базы данных в файлы соответствующих графических форматов для помещения на сервер.

ЗАКЛЮЧЕНИЕ

В заключение отметим, что содержательную часть энциклопедии создают более 70 человек, представляющих более 20 институтов РАН и других научных учреждений, а также ряда университетов России. В настоящей статье дано краткое описание программно-технологического комплекса интернет-энциклопедии, его моделирования и реализации. Расширенное описание см. в работах [2, 4, 5, 7].

И, наконец, вернемся к Введению. Мы уже отмечали [8], что в 1990-х гг. расплодившиеся многочисленные издательства выпускали тиражами в десятки и даже сотни тысяч экземпляров книги, журналы, газеты не только без редактирования, но даже и без корректуры. По наивности, думалось, что это безобразие в том десятилетии и осталось, период взросления и внутривидового отбора прошел. Однако издательство "Эксмо", выпустившее в 2007 г. тиражом в 4000 экз. "Большую