

# Управление контентом в крупных научно-технических Internet-библиотеках \*

© Е.Е. Сальникова,

ЦИТФорум  
elev@citforum.ru

С.А. Сальников,

ЦИТФорум  
serg@citforum.ru

С.Д. Кузнецов

ИСП РАН  
kuzloc@ispras.ru

## Аннотация

Существующие системы управления Web-контентом не ориентированы на поддержку научно-технических интернет-библиотек, в которых текстовый контент может обладать произвольно большим размером. На основе многолетнего опыта поддержки и развития Internet-библиотеки ЦИТФорум формулируется набор функциональных требований, которым должна отвечать желательная система управления Web-контентом. Обсуждаются основные черты существующих систем, и отмечается их неудовлетворительность для использования в требуемых целях. Кратко рассматриваются готовые компоненты с открытым кодом, которые можно использовать при создании новой системы. Приводятся соображения относительно способов миграции контента существующих Internet-библиотек в новую среду. В заключение отмечаются другие области возможного применения разрабатываемой системы.

## 1 Введение

Полнотекстовые библиотеки научно-технической литературы являются популярными ресурсами российской части Internet. Во многих случаях эти библиотеки начали формироваться на заре Рунет, когда для публикации использовались подготовленные обычным редактором статические файлы HTML. По мере разрастания библиотек, повышения темпов их пополнения и увеличения уровня требований к качеству текстов эта унаследованная технология становится обременительной.

Желателен переход к современной технологии динамического формирования HTML-страниц на основе контента, сохраняемого в базе данных, обеспечения средств редактирования контента,

выполнения требуемого администрирования и т.д. Однако, как будет показано в разд. 3, существующие системы управления Web-контентом (Web Content Management System, WCMS) не ориентированы на поддержку научно-технических интернет-библиотек, в которых текстовый контент может обладать произвольно большим размером. Требуется выполнение исследований, проектирования и разработки новой системы управления контентом.

Кроме того, сложной и трудоемкой задачей является перенос накопленного контента из унаследованной в новую WCMS. Обычно разработчики WCMS считают, что задача миграции контента должна решаться силами пользователей новой системы и не предлагают каких-либо инструментальных средств, хотя бы частично облегчающих ее решение. На Западе существуют отдельные компании, специализирующиеся на оказании подобных услуг (см., например, [15]). Естественно, в России такие услуги никем не предоставляются, и временами это приводит к результатам, печальным для пользователей соответствующих Internet-ресурсов.

В качестве наиболее близкого авторам примера можно привести Internet-библиотеку опубликованных материалов издательства «Открытые системы» [31]. С 1995 по 1998 гг. это издательство публиковало первый и до сих пор единственный в России журнал «СУБД», целиком посвященный тематике баз данных. Все выпуски журнала одновременно публиковались в бумажной форме и на сайте издательства. Опубликованные статьи были весьма востребованы читателями. Однако несколько лет назад издательство решило обновить свою систему управления контентом и, чтобы облегчить задачу миграции контента, оставило материалы журнала «СУБД» на старом сайте (несколько раз поменяв его имя). В результате интереснейшие статьи стали практически недоступны читателям, их невозможно найти через поисковые машины.

Чтобы спасти хотя бы наиболее ценные материалы (переводы ряда классических статей) в 2009 г. с согласия руководства «Открытых систем» мы заново отредактировали их и опубликовали на сайте ЦИТФорум [30], который сам остро нуждается в переходе на новую WCMS. Этот пример, во-

---

Труды 11<sup>й</sup> Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2009, Петрозаводск, Россия, 2009.

первых, показывает важность надежно поддерживаемых Internet-изданий в современном мире, а во-вторых, демонстрирует нетривиальность процесса обновления их технологии.

Дальнейший материал доклада организован следующим образом. В разд. 2 формулируются функциональные требования к WCMS, пригодной для поддержки полнотекстовых HTML-ориентированных Internet-библиотек. В разд. 3. обсуждается состояние дел в области систем управления Web-контентом, и обосновывается их, как минимум, неполная пригодность для поддержки Internet-библиотек данного класса. В разд. 4 кратко описываются подходы, которые планируют использовать авторы для создания новой WCMS, удовлетворяющей поставленным требованиям. Разд. 5 посвящен обсуждению вопросов миграции контента на новую платформу. Наконец, в заключительном разд. 6 рассматриваются другие области потенциального применения новой системы.

## **2 Функциональные требования к системе управления контентом Internet-библиотеки**

С 1996 г. авторы развивают и поддерживают Internet-библиотеку информационных технологий CITForum ([www.citforum.ru](http://www.citforum.ru)). Исторически сложилось так, что хранимые в библиотеке документы (крупные и мелкие статьи, книги) сохраняются на сервере в файлах в формате HTML. Тщательно отслеживается, чтобы эти документы были доступны для просмотра в любом браузере.

Контент библиотеки покрывает большинство тем, относящихся к информационным технологиям: инженерия программного обеспечения, методы и средства программирования, операционные системы, базы данных, сетевые и Internet-технологии, информационная безопасность, аппаратные средства. Все разделы сайта постоянно пополняются новыми материалами (с обязательным сохранением ранее опубликованных материалов). В библиотеке публикуются материалы конференций, проводимых компанией ЦИТФорум, обзоры ведущих зарубежных журналов и т.д. На протяжении многих лет поддерживается ряд форумов: для системных администраторов, программистов и Web-мастеров и т.д.

Сайт пользуется устойчивой популярностью. В среднем каждый день его посещает 15-20 тыс. человек, в месяц – более 300 тысяч из России, Украины, Белоруссии, Казахстана, Литвы, Израиля, США, Канады и других стран ближнего и дальнего зарубежья (все материалы представлены на русском языке). На рассылку новостей библиотеки подписано 48 тыс. человек. Материалы библиотеки активно используются преподавателями и студентами университетов и ВУЗ'ов, о чем, в частности, свидетельствует значительный рост

числа ее посетителей во время экзаменационных сессий.

Контент библиотеки в основном вторичен, т.е. включает документы, ранее опубликованные в печатных изданиях и/или в Internet и перепечатанные с разрешения владельцев авторских прав. Централизованное надежное хранение вторичного контента оказалось целесообразным и оправданным: к настоящему времени многие не утратившие своей ценности документы остаются доступными только в библиотеке CITForum (и на ее зеркалах). Однако нетривиальной задачей является импорт материалов, которые могут быть представлены в самых разнообразных форматах: HTML с разнообразными вариантами верстки, XML с таблицами стилей, TeX, Word различных версий, PDF и т.д. Другой очевидной (и решаемой, в лучшем случае, абсолютно вручную) проблемой в этом случае является обеспечение согласованности копии с потенциально изменяющимся оригиналом.

С другой стороны, в библиотеке CITForum размещается много оригинальных материалов, написанных или переведенных специально для публикации на этом сайте. Эти материалы обычно приходится тщательно редактировать (в ряде случаев в удаленном контакте с авторами). Без наличия средств коллективного редактирования делать это очень неудобно, громоздко и ненадежно. Кроме того, в соответствии с политикой редакции CITForum, любой материал, опубликованный в этой библиотеке, может быть перепечатан любым образом с разрешения редакции. Особенностью Internet-публикаций является потенциальная изменчивость текстов по инициативе авторов и/или редакторов. Очевидно, что в этом случае требуется поддержка версий опубликованных документов и истории их изменений.

В настоящее время программные средства, используемые для поддержки библиотеки CITForum (и большинства известных авторам других научно-технических Internet-библиотек, например, упомянутой ранее библиотеки издательства «Открытые системы» [31], насыщенной, но плохо организованной библиотеки компании «Интерфейс» [32]), не поддерживают эти и другие жизненно-важные функции. В результате, по мере расширения библиотеки,

- все труднее обеспечивать ее должное качество;
- велика трудоемкость подготовки первичных и вторичных материалов к публикации;
- трудно менять рубрикации материалов, вводить новые разделы и т.д.;
- нелегко обеспечивать абсолютную гарантию сохранности и доступности ранее опубликованных материалов;
- все более сложной становится задача общего администрирования библиотеки.

Многолетний опыт авторов по развитию и поддержке библиотеки CITForum позволяет выдвинуть к желаемой системе управления контентом следующие функциональные требования.

## 2.1 Поддержка версий и истории изменений

Для любого хранимого в системе документа система должна:

- поддерживать требуемое число его версий,
- предоставлять данные о числе версий,
- обеспечивать доступ к каждой версии и
- выдавать подробный и понятный список изменений, обеспечивавших переход от предыдущей версии к последующей.

Другими словами, в WCMS должна иметься встроенная современная система управления версиями.

## 2.2 Удобные средства для работы с крупными документами

В полнотекстовых Internet-библиотеках предельная длина HTML-документа должна быть практически неограниченной (даже сейчас максимальная длина HTML-документа в библиотеке CitForum составляет 11 мегабайт). Для редактирования документа сейчас требуется его полная загрузка из серверного хранилища, а разбиение на страницы делается почти вручную.

В желаемой системе управления контентом:

- должна иметься возможность выполнять мелкие (локальные) изменения в хранимых документах без их полной загрузки;
- разбиение больших документов на страницы должно быть полностью автоматизировано;
- должна предоставляться возможность изменять разбиение документа на страницы, сохраняя корректность ссылок на внутренние точки документа.

## 2.3 Развитые средства форматирования

В целом требуемые средства форматирования должны отвечать общепринятым представлениям о системах управления контентом:

- в обязательном порядке должна поддерживаться вставка в отображаемый текст отдельно хранимых мультимедийных объектов (изображений, звука, видео);
- должны обеспечиваться удобные и мощные средства вставки математических формул (скорее всего, по образу основанного на TeX механизма Википедии [28]);
- менее важны, хотя и желательны, поддержка упрощенного синтаксиса разметки и наличие средства редактирования типа WYSIWYG (это менее важно, поскольку предполагается, что в данном случае с системой управления контентом будет работать ограниченное число редакторов и авторов с достаточной технической подготовкой).

## 2.4 Преобразование внешних источников к внутреннему представлению системы управления контентом

Система должна обеспечивать преобразование к своему внутреннему формату внешних документов, представленных в различных иных форматах (в частности, HTML, TeX, Word и т.д.). В системе должна обеспечиваться связь импортированного документа с первоисточником и автоматическая синхронизация при изменении первоисточника. Естественно, полностью автоматизировать преобразование из произвольного внешнего формата невозможно, поэтому следует распознавать случаи, когда изменения первоисточника требуют вмешательства редактора и изменения правил преобразования.

Подсистема импорта должна быть, прежде всего, опробована при переносе в новую систему текущего контента библиотеки CitForum, произведенного и поддерживаемого почти «вручную»..

## 2.5 Средства поиска

Поскольку речь идет о системах управления полнотекстовым контентом, в первую очередь, должны поддерживаться качественные средства полнотекстового поиска, обеспечивающие высокий уровень релевантности результатов. Кроме того, поскольку в системе управления контентом для каждого документа будут сохраняться различные метаданные, полнотекстовый поиск должен быть интегрирован с поиском документов по их метаданным.

## 2.6 Проверка распространенных ошибок

Система должна проверять разнообразные часто встречающиеся ошибки:

- синтаксические ошибки;
- опечатки; вероятно, в систему должен быть встроен спеллер;
- ошибки отображения,
  - в частности, вылезание части текста за пределы экрана на распространенных устройствах доступа;
  - соответствующее средство должно обновляться по мере распространения новых браузеров и операционных систем и регулярно запускаться по мере обновления как документов, так и самого средства;
- неработающие («битые») ссылки;
  - это средство также должно запускаться на регулярной основе.

## 2.7 Поддержка метаданных, полезных пользователям

Наряду со служебными метаданными, используемыми самой системой, должны поддерживаться метаданные с информацией, полезной пользователям:

- история документа (даты публикации, перевода, перепечатки и т.д.),
- авторы, их координаты и т.д.

По этим метаданным также должен поддерживаться поиск.

## 2.8 Поддержка тэгов как средства пользовательской навигации

В системе должны поддерживаться определяемые пользователями тэги («ярлыки») наподобие тех, которые используются на популярных сайтах типа livejournal [29], slashdot [22], gmail [5] и т.д. Наличие таких тэгов облегчает навигацию пользователя по материалам библиотеки.

## 2.9 Представления документов

Система должна обеспечивать различные внешние представления сохраняемых документов, например, в виде «одной страницы», в формате pdf, в форматах, предназначенных для разных размеров экрана и каналов в интернет (подстановка уменьшенных копий вместо картинок) и т.д.

При соблюдении перечисленных требований система управления контентом должна быть пригодна для эффективной поддержки Internet-библиотек любого размера – от библиотек масштаба CitForum до «сайтов одной книги».

## 3 Состояние дел в области систем оперативной подготовки и сопровождения Web-контента

Системы категории WCMS начали разрабатываться с середины 1990-х гг. Основное назначение этих систем состояло в том, чтобы обеспечить простые средства создания Web-контента (без потребностей в хорошем знании языков разметки и/или программирования), его редактирования, управления и сопровождения. К настоящему времени количество реально используемых WCMS составляет несколько десятков. В их число входят как коммерческие системы известных поставщиков (IBM Lotus Web Content Management [13], Microsoft Office SharePoint [20], EMC Documentum [10] и т.д.), так и многочисленные WCMS, произведенные в сообществе open source (к наиболее интересным системам можно отнести Drupal [8], Alfresco [1], Magnolia [18] и др.).

Следует отметить, что коммерческие WCMS в большой степени ориентированы на поддержку корпоративных Web-сайтов, Web-ориентированных систем e-бизнеса и т.д. Они интегрируются с системами управления документооборотом и/или бизнес-процессов, с системами управления связями с заказчиками и т.д. В коммерческих системах имеется много функциональных возможностей, совершенно не требуемых при управлении контентом Internet-библиотек, а с другой стороны, в

них, как правило, не хватает ряда средств, упомянутых в разд. 2: возможности поддержки документов большого размера, коллективного редактирования и т.д.

Многие из WCMS категории open-source также в основном предназначены для поддержки корпоративных Web-ориентированных информационных систем и систем поддержки e-бизнеса. В частности, к таким системам относятся Alfresco и Magnolia. Ряд других систем этой категории, в частности, Drupal, направлен на содействие коллективному Web-творчеству. В них, безусловно, поддерживается коллективное редактирование, хотя бы примитивное управление версиями и т.д.

WCMS, поддерживающие онлайн-подготовку и коллективное редактирование контента, обычно называются «Wiki» [24]. Такие системы более близки к тому, что требуется для поддержки и развития Internet-библиотек. Возможно, при создании новой системы управления контентом стоит частично использовать код существующих Wiki – это еще не оценивалось, анализировались только возможности с точки зрения пользователя.

Тем не менее, использовать Wiki-подобную WCMS для целей поддержки полнотекстовой HTML-ориентированной Internet-библиотеки оказывается невозможно. В частности, нам неизвестны Wiki- WCMS, которые бы хорошо подходили для следующих сценариев.

### 3.1 Длинный документ (книга)

Длинным считается такой документ, который становится неудобно держать на одной странице. В единственной известной нам библиотеке Wikibooks [25], основанной на такой Wiki-подобной системе управления контентом, используется инструмент MediaWiki [12], предназначенный для поддержки энциклопедии – набора небольших статей, ссылающихся одна на другую. Насколько мы можем судить, редакторы собственно Википедии сталкиваются с заметными неудобствами, когда статья постепенно вырастает настолько, что ее части пора выделять в отдельные страницы.

В Drupal также имеется средство для подготовки, редактирования и публикации книг. Однако под книгой здесь понимается набор страниц, связанных в некоторую иерархию. Такая структура хорошо подходит для технических руководств, ответов на часто задаваемые вопросы и т.д., но совершенно неочевидно, что она будет пригодна для удобного размещения в Web произвольной книги, которая писалась в расчете на бумажное издание.

### 3.2 Документы, которые создаются в формате, не совпадающем с внутренним форматом

Практически во всех современных WCMS поддерживаются средства импорта/экспорта для

переноса контента из одной системы в другую. Однако нам не удалось обнаружить ни одной коммерческой или свободно распространяемой системы, в которую были бы интегрированы средства импорта документов, созданных вне этой системы в каком-либо внешнем формате. По всей видимости, такие средства просто не требуются в сценариях, для которых предназначены типичные WCMS. Как отмечалось выше, в WCMS, используемой для поддержки полнотекстовых HTML-ориентированных Internet-библиотек, возможность импорта внешних источников просто необходима.

В программном обеспечении Википедии имеется ряд конверторов документов, представленных во внешних форматах, в формат Wiki (в частности, из форматов HTML, Word, Open Office, LaTeX) [26]. Безусловно, при разработке новой WCMS следует их оценить и, возможно, использовать. Однако нам не встречались упоминания о каких-либо инструментах, позволяющих отслеживать изменения в таких документах и помогающих соответствующим образом изменять документы, хранимые во внутреннем формате WCMS.

### 3.3 Проверка ошибок

В известных нам WCMS проверке ошибок в создаваемых или редактируемых документах уделяется незначительное внимание. И это понятно, поскольку инструменты проверки ошибок тем более актуальны, чем больше имеется возможностей поменять большое количество страниц, не посмотрев внимательно на результат. Поскольку Wiki-подобные WCMS предназначены для правки отдельных страниц, они таких инструментов и не содержат.

В онлайн-овых Internet-библиотеках авторами и редакторами могут правиться очень крупные документы, например, книги. Правки могут быть массовыми, затрагивающими несколько страниц. В этих условиях наличие развитых средств проверки различного рода ошибок становится обязательным.

## 4 Подход к созданию новой WCMS

Из всего сказанного выше легко заметить, что заметная часть нашей задачи ранее уже решалась. Перечислим крупные компоненты создаваемой системы, для которых кажется неестественным «изобретать велосипед», а можно воспользоваться готовыми решениями с открытыми кодами.

### 4.1 Системы управления версиями

«Индустриальным стандартом» современных систем управления версиями в настоящее время считается Subversion [34]. Эта система пришла на смену традиционной и широко распространенной системе CVS [4]. Наиболее привлекательные черты Subversion с точки зрения WCMS состоят в следующем:

- отслеживание изменений каталогов – под управление версиями попадают и файлы, и каталоги;
- атомарная фиксация изменений – каждый набор изменений либо попадает в хранилище целиком, либо не попадает туда вовсе;
- метаданные с версиями – у каждого файла и каталога имеется собственный набор метаданных, которые также находятся под управлением версиями;
- автоматическое (или полуавтоматическое) слияние параллельно выполняемых изменений;
- затраты ресурсов пропорциональны размеру изменений, а не размеру данных, затронутых изменениями;
- библиотеки для языков PHP, Python, Perl, Java позволяют встроить функциональность клиента Subversion в программы, написанные на этих языках.

Возможно, будут полезны эксперименты с более новыми распределенными системами Bazaar [2], Darcs [6], Git [11], Mercurial [19] и Monotone [21].

### 4.2 Библиотеки для работы со сложными форматами (языки разметки, форматы изображений и т.п.)

Очень большое количество модулей для работы со сложными данными содержится в библиотеке CPAN [3]. Эта библиотека предназначена для поддержки разработки приложений на языке Perl, хорошо поддерживается, легко доступна. Наличие в CPAN модулей, облегчающих разработку пользовательских интерфейсов, поддерживающих использование программ, написанных на других языках, обеспечивающих простой доступ из среды Perl к готовым коммерческим и свободно доступным системам стимулирует применение Perl как основного языка разработки новой WCMS.

### 4.3 Средства создания развитых Web-интерфейсов на языке JavaScript

Для разработки Web-интерфейсов в настоящее время принято использовать смесь технологий XML и JavaScript, получившую название AJAX. Имеется большой выбор реализаций AJAX с открытым кодом (см., например, [9]). Кроме того, доступен ряд редакторов rich text [7], которые позволяют в онлайн-овом режиме в стиле WYSIWYG редактировать контент Web-сайтов.

### 4.4 Верстка формул

Признанным средством верстки формул является TeX (см., например, [16]). TeX широко используется для написания математических и физических статей, и средства верстки формул годами обрабатывались с учетом требований ученых.

## 4.5 Wiki

Как отмечалось ранее, код существующих Wiki нами еще не изучался, но имеется надежда, что удастся использовать, по крайней мере, часть кода

- UseModWiki [23] – системы управления контентом, которая раньше использовалась в Википедии; в ее плагинах реализованы полезные функции, например, вставки формул и/или
- IkiWiki [14] – здесь привлекает представление разработчиков о хранении и публикации: документы хранятся в "настоящей" системе контроля версий – по умолчанию в Subversion, и компилируются в статический HTML.

## 5 Как будет производиться миграция на новую систему управления контентом и что это даст

Перенос отдельных документов стоит пытаться автоматизировать целиком или почти целиком – кроме отдельных документов, сверстанных совсем плохо.

Очевидно, нужно сделать и массовый инструмент, автоматизирующий обработку большинства документов с существующего сайта. Насколько это большинство будет подавляющим – зависит от того, насколько последовательно поддерживался исходный сайт. Для библиотеки CITForum, с учетом некоторого опыта автоматической обработки, можно предположить, что более 90% контента удастся перенести именно полностью автоматически. При этом нужно стремиться делать этот обработчик достаточно общим, чтобы можно было его применять и к другим сайтам.

Очевидно, что после завершения процесса миграции будет существовать заметный период доработки и тестирования (до нескольких месяцев), когда исходная библиотека будет жить в существующем формате и своей жизнью, а новая система управления контентом должна будет учитывать происходящие там изменения и без повторяющейся ручной работы соответствовать текущей версии.

Переход на новую систему (внутри и снаружи) позволит значительно сократить трудозатраты на публикацию документов. Возможно, станет целесообразно публиковать материалы, за которые раньше не брались из-за их чрезмерной сложности. (Полу)автоматическая синхронизация даст возможность во многих случаях избежать устаревания документов – в библиотеке CITForum в настоящее время это причиняет довольно серьезные неприятности. Упрощение и систематизация разметки, автоматическая проверка ошибок улучшит качество сайта. Поиск и тэги, альтернативные внешние представления – это просто новые возможности для читателей.

## 6 Заключение. Кому и как может быть полезна новая система

Прежде всего, как отмечалось выше, создаваемая система должна быть удобна не только для сайта с вторичным контентом, но и как один из вариантов системы, в которой контент создается. Здесь нам определенно пригодится опыт, полученный во время публикации книг [27] и [33].

Хорошее решение задачи будет полезно не только для сайтов, но и вообще для работы над документами, в том числе, в случае, когда публикация собственно в формате HTML не очень интересна.

Например, думается, что создаваемое решение пригодится для подготовки документов в тех областях, для которых сейчас принято использовать LaTeX с обычным текстовым редактором (научные работы по математике и т.п.). Существующему решению на основе TeX, очевидно, не хватает следующих возможностей, приблизительно в порядке важности:

- удобства совместной работы над документом;
- удобных средств централизованного хранения готовых документов и проектов в доступном для членов команды (или публично, в зависимости от конкретного применения) месте, включая поиск по этим документам;
- простоты установки и использования: при работе с TeX возникают разнообразно тонкие проблемы, особенно если переносить исходный текст на другую систему – в такой ситуации просто напрашивается серверное решение;
- функций типа WYSIWYG, особенно вставки изображений<sup>1</sup>.

Полностью открытый характер нашего проекта и использование лицензии open source, безусловно, помогут найти и другие области полезного применения создаваемой системы управления контентом.

## Литература

- [1] Alfresco. <http://www.alfresco.com/>
- [2] Bazaar. <http://bazaar-vcs.org/>
- [3] CPAN: Comprehensive Perl Archive Network. <http://www.cpan.org/>
- [4] CVS – Concurrent Versions System. <http://www.nongnu.org/cvs/>
- [5] Справка – Gmail – Ярлыки. <http://mail.google.com/support/bin/topic.py?hl=ru&topic=12881>
- [6] Darcs. Distributed. Interactive. Smart. <http://darcs.net/>
- [7] Don Albrecht. 5 Open Source Rich Text Editors, 2007. <http://www.ajaxbestiary.com/2007/08/14/5-open-source-rich-text-editors/>
- [8] Drupal community. <http://drupal.org>
- [9] Edmon Begoli. An Open Source AJAX Comparison Matrix, 2006.

- <http://www.devx.com/AJAXRoundup/Article/33209>
- [10] EMC Documentum. <http://www.emc.com/products/family/documentum-family.htm>
- [11] Git – the fast version control system. <http://git-scm.com/>
- [12] How does MediaWiki work? [http://www.mediawiki.org/wiki/How\\_does\\_MediaWiki\\_work%3F](http://www.mediawiki.org/wiki/How_does_MediaWiki_work%3F)
- [13] IBM Lotus Web Content Management. <http://www-01.ibm.com/software/lotus/products/webcontentmanagement/>
- [14] IkiWiki. <http://ikiwiki.info/>
- [15] Kyle Short. No Small Task: Migrating Content to a New CMS, 2008. <http://www.cmswire.com/cms/web-publishing/no-small-task-migrating-content-to-a-new-cms-002437.php>
- [16] LaTeX – A document preparation system. <http://www.latex-project.org/>
- [17] LyX – The Document Processor. <http://www.lyx.org/>
- [18] Magnolia. <http://www.magnolia-cms.com/>
- [19] Mercurial. <http://www.selenic.com/mercurial/wiki/>
- [20] Microsoft SharePoint. <http://sharepoint.microsoft.com/Pages/Default.aspx>
- [21] Monotone. <http://www.monotone.ca/>
- [22] Slashdot FAQ. <http://slashdot.org/faq/>
- [23] UseModWiki. <http://www.usemod.com/cgi-bin/wiki.pl>
- [24] What Is Wiki. <http://wiki.org/wiki.cgi?WhatIsWiki>
- [25] Wikibooks. <http://wikibooks.org/>
- [26] Wikipedia:Tools/Editing tools. [http://en.wikipedia.org/wiki/Wikipedia:Tools/Editing\\_tools](http://en.wikipedia.org/wiki/Wikipedia:Tools/Editing_tools)
- [27] Алексей Федорчук. Zenwalk. Приобщение к Linux. 2008. <http://citkit.ru/articles/892/>
- [28] Википедия:Формулы. <http://ru.wikipedia.org/wiki/Википедия:Формулы>
- [29] ЖЖ: самое важное. [http://www.livejournal.com/tour\\_rus/about\\_lj.bml](http://www.livejournal.com/tour_rus/about_lj.bml)
- [30] Переводы классических статей по тематике баз данных, новая редакция, 2009. <http://citforum.ru/database/classics/index.shtml>
- [31] Сайт издательства «Открытые системы». [www.osp.ru](http://www.osp.ru)
- [32] Сайт компании «Интерфейс». <http://www.interface.ru/home.asp>
- [33] Сергей Кузнецов. Базы данных. Вводный курс. 2008. [http://www.citforum.ru/database/advanced\\_intro/](http://www.citforum.ru/database/advanced_intro/)
- [34] Управление версиями в Subversion. <http://svnbook.red-bean.com/>

## Content Management in Large Technological Internet Libraries

E.E.Salnikova, S.A.Salnikov, S.D.Kuznetsov

Existing Web content management systems (WCMSs) are not designed for technological Internet libraries containing documents of unlimited size. Based on many years of experience in development and maintenance of the CifForum Internet library the authors define a set of functional requirements for a Web content management system appropriate for this task. Key features of existing systems are discussed, and their inadequacy for this kind of usage is justified. Then, open source components which can be used to build a new system are briefly considered. Finally, some considerations on possible ways to migrate content of old Internet libraries to the new environment and on other possible applications of the proposed system are presented.

\* Работа поддерживается грантом РФФИ № 09-07-00282.

<sup>1</sup> Заметим, что здесь некоторое решение предлагает, например, LyX [17] – и, возможно, стоит учесть этот опыт. Впрочем, эта функция представляется ценной далеко не всегда, и, что более важно, предыдущие пункты все равно намекают на онлайн-овое решение, а таковые нам неизвестны.