

О создании открытой научной электронной библиотеки периодических изданий НАНУ

© В.А. Резниченко, Г.Ю. Проскудина, К.А. Кудим

Институт программных систем НАН Украины, г. Киев
reznich@isofts.kiev.ua, gupros@isofts.kiev.ua, kuzma@isofts.kiev.ua

Аннотация

В работе рассматриваются вопросы создания открытой научной электронной библиотеки периодических изданий НАН Украины. Сделан обзор некоторых известных проектов порталов периодических изданий. А также особенности функциональных возможностей программного обеспечения DSpace применительно к решаемой задаче.

1 Введение

Работы по данному направлению выполняются в рамках проекта «Разработка и реализация проектных решений автоматизированного библиотечного сервиса в интересах организаций НАН Украины». Проект предполагает выполнение нескольких заданий, одно из которых – «Разработка технологий, методов и средств создания научных электронных библиотек НАН Украины». Согласно этому заданию проводились работы:

- по созданию научных электронных библиотек в научно-исследовательских организациях НАНУ;
- а также выработка методов и средств создания интегрированных электронных библиотек в Академии наук в целом.

Доклад представляет собой обобщение нашего опыта по созданию интегрированной электронной библиотеки периодических изданий – одного из основных видов информационных ресурсов научно-исследовательских академических организаций на основе современных тенденций и решений.

Существующие в настоящее время и разрабатываемые новые системы электронных библиотек (ЭБ) характеризуются следующими тремя ключевыми особенностями:

- прежде всего это информационные системы, то есть системы сбора, хранения, обработки и выдачи информации;
- во-вторых, это системы, поддерживающие

именно электронные документы (ресурсы) различного типа;

- и, наконец, важным моментом в современном понятии ЭБ является то, что такие системы работают в глобальных информационных сетях передачи данных.

Следует отметить, что в мире на протяжении последних 20 лет интенсивно проводятся исследования и разработки по созданию ЭБ. Осознание того факта, что ЭБ являются одной из важнейших составляющих информатизации общества, привело к тому, что во многих ведущих странах мира были приняты национальные программы по созданию ЭБ, которые поддерживаются правительствами и которые имеют внушительную финансовую поддержку. Вот далеко не полный перечень таких программ:

- Национальная электронная библиотека США (Библиотека конгресса США).
- «Электронная библиотека XXI столетия» – Япония.
- Национальная программа Великобритании eLib по созданию ЭБ.
- Программа ЕС «Создание Европейского библиотечного пространства».
- Программа «Электронные библиотеки России».
- Программа «Национальная электронная библиотека Украины».
- Национальные НЭБ США, Франции, Великобритании, Канады, России.

В результате этих усилий в настоящее время в сети Интернет появилось множество ЭБ различной направленности, в том числе и научных электронных библиотек (НЭБ). Они в основном создаются в НИИ, учебных заведениях, международных ассоциациях. Среди них имеются и НЭБ открытого доступа. Наиболее известны два международных каталога НЭБ открытого доступа:

- OpenDOAR – каталог репозиториев открытого доступа, насчитывающий 1172 НЭБ, <http://www.opendoar.org>;
- ROAR – реестр репозиториев открытого доступа, насчитывающий 1101 НЭБ, <http://roar.eprints.org>.

Следовательно, в настоящее время в открытом доступе уже более 1100 НЭБ, тогда как до середины прошлого года не было ни одной НЭБ от Украины.

Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL2008, Дубна, Россия, 2008.

Однако уже к концу года в этих реестрах появилось 7 украинских НЭБ, в том числе 2 от Академии наук, 4 от ВУЗов и 1 от общественной организации при поддержке фонда «Возрождение». Причем 4 из них были созданы при нашем непосредственном участии. Эти библиотеки создавались на базе научных и учебных заведений, специфика которых определила характер и состав содержащихся в них научных информационных ресурсов.

Основная тема нашего доклада – создание НЭБ периодических изданий открытого доступа (предполагается публикация приблизительно 100 журналов). Здесь также сделано достаточно много в мире. Международный каталог журналов открытого доступа DOAJ в настоящее время содержит 3245 таких журналов и среди них 10 из Украины.

Анализ специализированных журналов Украины и НАНУ, в частности [1], показал, что 80% периодических изданий физико-математических наук, 87% химических, 95% геологических, 95% географических, и 97% технических наук практически остаются невидимыми.

Следует отметить, в Национальной библиотеке Украины им. Вернадского (НБУВ) с 2006 года ведется Общеакадемический портал научной периодики, куда включено приблизительно 500 научных периодических изданий Украины. В нем предпринята первая в Украине успешная попытка объединить научные информационные ресурсы периодических изданий. Наши предложения и решения являются не альтернативой тому, что уже создано в НБУВ, а его дальнейшим развитием с учетом новых, современных концепций. В частности, предполагается основное внимание уделить следующим нерешенным вопросам, которые на наш взгляд являются узким местом:

- поиск по картотеке статей (электронный каталог), нет полнотекстового поиска;
- нет средств пополнения библиотеки с рабочих мест в институтах;
- нет средств интеграции в международные каталоги, реестры, репозитории.

Далее в работе рассматриваются: примеры известных проектов порталов научной периодики открытого доступа (раздел 2); формулировка целей и концептуальных положений открытой НЭБ периодических изданий НАНУ (раздел 3); обоснование выбора платформы ПО для решения этой задачи (раздел 4); краткая информация о ПО (раздел 5). Вопросы масштабируемости системы DSpace рассмотрены в разделе 6; некоторые важные функциональные возможности системы DSpace рассмотрены в разделах 7–10 и 12; процедура внесения статей в систему описана в разделе 11.

2 Обзор известных проектов порталов научной периодики открытого доступа

Рассмотрим более детально особенности построения и реализации подобных систем на примере

довольно крупных и функционально приемлемых решений.

Хорватская система периодических изданий <http://hrcak.srce.hr/?lang=en>, на момент написания статьи содержала 163 наименований журналов по различным направлениям научной деятельности (естественные, медицинские, технические и гуманитарные науки), 2087 выпусков, 24,609 статей, из которых 22,287 полнотекстовых. Здесь предоставляется как простой, так и расширенный поиск, который включает булеву (операторы И, ИЛИ, И-НЕ) комбинацию до четырех полей метаданных (автор, название, аннотация, издательство, ключевое слово, название журнала) и полного текста, а также запрос можно уточнить дополнительными значениями области исследования, года выпуска статьи, языка и словом в тексте статьи. Так, например, на запрос с указанием только лишь русского языка, система выдала 71 полнотекстовую статью.

Бразильская интегрированная система публикации периодических изданий SciELO (Scientific Electronic Library Online, (<http://www.scielo.org/php/index.php>)) содержит:

- 550 журналов из Аргентины, Бразилии, Чили, Кубы, Испании, Португалии, Колумбии и Венесуэлы;
- 11,941 выпусков;
- 180,187 полнотекстовых статей;
- 3,530,025 ссылок цитирования.

Помимо предоставления сервисов просмотра и поиска (в том числе полнотекстового), вычисляются интегрированные статистические показатели популярности (использования) и влияния (импакт-факторы) периодических изданий, научных публикаций.

Наиболее мощной, на наш взгляд, является система HighWire Press (<http://highwire.stanford.edu/>), подразделение библиотеки Стэнфордского университета, которая публикует онлайн версии журналов с высоким импакт фактором (high-impact) и уровнем рецензирования (peer-reviewed) и другой научный (учебный) контент. HighWire Press награждена премией ALPSP (Association of Learned and Professional Society Publishers) за предоставляемые «Сервисы для некоммерческой публикации». HighWire в партнерстве с влиятельными академическими сообществами, университетскими издательствами, создает прекрасные онлайн-коллекции полностью доступной для поиска исследовательской и клинической литературы. Совместно эти партнеры выпускают 71 из 200 наиболее часто цитируемых научных журнальных публикаций.

HighWire Press представляет собой наиболее крупный в мире репозиторий полнотекстовых статей наук о жизни (биология, медицина, антропология, социология и т. п.), где (на момент написания этого доклада) содержится 1,904,070 статей, которые находятся в открытом доступе. Всего же ЭБ содержит 1157 журналов и 4,806,738 полнотекстовых статей из более чем 140 научных издательств.

3 Цели и концептуальные положения

Создание открытой НЭБ периодических изданий НАН Украины преследует следующие цели:

- Достижение качественно нового уровня, полноты и оперативности удовлетворения информационных потребностей научных сотрудников НАНУ.
- Повышение эффективности использования имеющихся информационных ресурсов НАНУ.
- Оперативное и полное информирование международной научной общественности о научных достижениях и разработках институтов НАНУ.

Концептуальными моментами создания этой НЭБ являются следующие:

- Единый портал для описания журналов и хранения статей.
- Централизованное ведение репозитория, администрирование и предоставление услуг.
- Предоставление возможности по пополнению НЭБ электронными статьями с рабочих мест институтов или редакций журналов.
- Интеграция с другими международными каталогами, реестрами, репозиториями.

Схема функционирования НЭБ представлена на рис. 1.



Рис. 1. Общая схема функционирования НЭБ

При таком решении централизованно осуществляются следующие функции:

- администрирование;
- ведение библиотеки;
- предоставление услуг (сервисов);
- сопровождение программного обеспечения.

На местах, где редакции институтов, как правило, не имеют соответствующих специалистов для установки и поддержки библиотечной системы, осуществляются только лишь функции, связанные с наполнением контентом данной системы, а именно:

- оформление статей своих периодических изданий в электронном виде;
- загрузка или передача статей в систему, используя Интернет-браузер.

В итоге мы получаем продукт сотрудничества IT-специалистов и прикладных специалистов информационного профиля различных предметных областей.

4 Выбор платформы системы ПО

С начала 2005 года нами было изучено и опробовано несколько платформ ПО для построения такого рода систем. Результаты этой работы освещались в [2–5].

Для реализации НЭБ периодических изданий открытого доступа была выбрана ПО DSpace [6].

Из особенностей и преимуществ применения данной системы можно выделить следующие:

- В подавляющем большинстве приложений ПО DSpace используется для создания репозитория одной отдельной организации. В нашем приложении DSpace устанавливается и администрируется централизованно, но применяется многими организациями-партнерами, редакционными коллегиями периодических изданий институтов НАНУ. Каждая такая организация имеет свой раздел (разделы) и отвечает за его своевременное наполнение.
- Не осуществляется самоархивирования своих материалов самими авторами статей (важная функциональная особенность данной системы). В данном случае система выступает в качестве провайдера сервисов. Контент наполняется исключительно сотрудниками редакций периодических изданий.
- Возможность предоставления функций просмотра и поиска по основным атрибутам (и логическим комбинациям таких атрибутов) публикации (названию, автору, году издания, ключевым словам), полнотекстового поиска и синдикации RSS о новых поступлениях в библиотеку [7].
- Благодаря хранению описаний публикаций (метаданных) в соответствии с международными стандартами (в данном случае – расширенное Дублинское ядро) и поддержке протокола OAI-PMH [8] существует возможность интеграции с международными каталогами, например, OAIster, OpenDOAR, ROAR и др.
- Индексация Google.

5 DSpace – общие сведения

DSpace – программное обеспечение (ПО) системы управления цифровыми ресурсами или активами (DAMS – Digital Asset Management System), предоставляющее услуги по доступу, наполнению, управлению и повторному использованию цифровых активов, преимущественно образовательных и исследовательских материалов. В настоящее время система DSpace установлена в более чем 220

университетах мира, с суммарным объемом хранения более чем один миллион цифровых ресурсов, представляющих исследовательские статьи, технические отчеты, наборы данных, изображений и видео. Кроме того, некоторые другие организации используют DSpace, для хранения, организации и сбережения своих цифровых активов, например, HP Labs. В настоящее время, DSpace – система с открытым исходным кодом, размещаемая на SourceForge и распространяемая по лицензии BSD [9], авторские права принадлежат HP и MIT [6]. Для поддержки и защиты поставщиков информации на платформе DSpace в 2007 была создана некоммерческая организация, которая призвана вовлечь многие организации (являющиеся пользователями DSpace) в управление ПО DSpace.

Система DSpace развивается также в рамках программы Google Summer of Code™ (GSoC) под руководством Роберта Тенли (программиста первоначальной версии системы), где в настоящее время на платформе DSpace выполняются проекты [10]: сервисы целостности контента, статистика, поддержка версий, визуализация.

В настоящее время интенсивно обсуждается совместный проект сотрудничества DSpace и Fedora [11].

Таким образом, разработка системы DSpace находится в постоянном развитии и при непосредственном участии многочисленных (нескольких сотен) пользователей, что гарантирует долговременное использование, своевременное обновление, а также необходимую помощь и поддержку в случае возникновения непредвиденных проблем (например, утечки собственных кадров, IT-специалистов).

6 Масштабируемость

Перспективы развития системы также освещены в [12] на совещании групп пользователей системы DSpace, где особое внимание было уделено вопросам масштабируемости, способности системы адаптироваться к расширению предъявляемых требований и возрастанию объемов расширяемых задач. Проблемы масштабируемости рассматриваются в 3-х аспектах [13]:

- емкость, или вместимость, т.е. как DSpace будет справляться с большими объемами данных?
- пропускная способность, как быстро DSpace будет поглощать (импортировать) контент?
- параллелизм, как DSpace работает под нагрузкой, при одновременном доступе многих пользователей к репозиторию?

Исходя из этого были сформулированы требования масштабируемости для DSpace:

- DSpace может содержать 10 млн. записей;
- DSpace может обрабатывать файлы произвольных размеров;
- Время занесения элемента (статьи) в 10 млн. репозиторий не превышает 1 сек;

- DSpace поддерживает параллельную работу 10 пользователей на внесение информации и 100 пользователей на чтение, при наличии достаточной полосы пропускания, приемлемого процессора и емкости памяти;
- стратегия реализации DSpace – сбалансированные по загрузке кластеры серверов, управляющие одним репозиторием.

Наряду с обнадеживающими перспективами масштабируемости системы DSpace, анализ функциональности ЭБ [14] позволил выделить три важнейшие функциональные возможности, благодаря которым был сделан выбор в пользу данной системы:

- поддержка иерархической модели коллекций документов;
- достаточно хорошо поддающийся настройке пользовательский интерфейс;
- гибкая настройка прав доступа к коллекциям, отлаженные механизмы безопасности (аутентификации, авторизации и т.п.).

Остановимся несколько подробнее на перечисленных особенностях и наших проектных решениях.

7 Архитектура

Способ организации данных в DSpace [4] отражает структуру организации, использующей данную систему. Разделы и подразделы образуют структурную иерархию, т.е. подразделы могут также содержать свои подразделы, внизу иерархии находятся коллекции. В данном случае раздел соответствует конкретному периодическому изданию, периодические издания в свою очередь структурированы в иерархию согласно направлениям деятельности и существующим отделением (структурным подразделениям НАНУ, см. рис. 2).

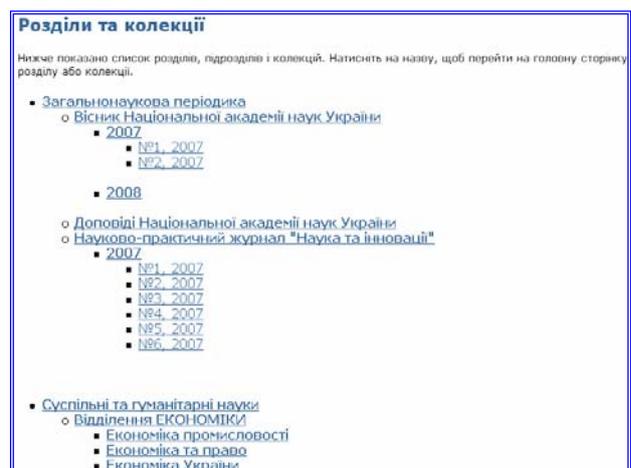


Рис. 2. Фрагмент иерархии разделов, подразделов и коллекций

КОЛЛЕКЦИЯ соответствует конкретному номеру журнала. Например, коллекция №1, 2007 журнала «Вестник Национальной академии наук Украины» (название журнала соответствует ПОДРАЗДЕЛУ, год выпуска журнала, например, 2007 – также соответствует ПОДРАЗДЕЛУ).

В результате имеем следующую структуру: раздел (community) – направление деятельности НАНУ; подраздел (sub-community) – отделение НАНУ и/или периодическое издание (журнал); подраздел (sub-community) – год или том издания; коллекция (collection) – выпуск или номер журнала.

Поскольку создаваемая библиотека не имеет предметной специализации, с большой вероятностью можно предположить, что каждому периодическому изданию или группе периодических изданий потребуется своя, специфическая процедура ввода и набор описательных атрибутов (обязательные, необязательные или факультативные, управляемые на момент внесения, свои тематические словари) при внесении своих ресурсов (в данном случае статей). Для реализации индивидуальных требований участников внесения контента в системе предусмотрена настройка как процедуры внесения, так и форм внесения, содержащих полный набор необходимых описательных атрибутов, форматов полных текстов статей, принятия лицензии по авторскому праву, дополнительные процедуры проверки своего контента и т.д. Установление разных параметров внесения для каждой коллекции (раздела) позволяет администратору системы устанавливать разный контент внесения, а также параметры проверки и одобрения (в случае необходимости) для каждой из коллекций и/или групп пользователей, которые определены в НЭБ.

Здесь же следует отметить важную возможность создания домашней странички для каждого раздела, подраздела, коллекции (рис. 3).



Рис. 3. Главная страница коллекции “№1, 2007” журнала “Вісник Національної академії наук України”

8 Настройка шагов внесения статей в коллекцию

Стандартная последовательность шагов внесения или процедура передачи документа в DSpace следующая:

Описание→Описание→Описание→Загрузка→Проверка→Лицензия→Завершение.

Однако при необходимости такая последовательность внесения может быть изменена, например, следующим образом:

Лицензия→Описание→Описание→Загрузка→Проверка→ Завершение.

Существует возможность переупорядочивания шагов внесения, либо некоторые шаги могут быть вообще пропущены, причем каждой коллекции может быть назначена своя процедура внесения. Настройка и конфигурирование пользовательского интерфейса DSpace внесения контента подробно описана в [15].

9 Настройка пользовательских экранов внесения

В DSpace возможна настройка форм для разных коллекций. Это делается с помощью конфигурационного файла input-forms.xml. В [16] представлено подробное описание его формата. Там можно указать какие именно поля Дублинского ядра (ДЯ) нужно заполнять на этапе внесения ресурса (статьи) в данную коллекцию, их последовательность и текстовые обозначения (метки) этих полей, как можно разбить ввод метаописаний на несколько страниц. Также через веб-интерфейс для определенной коллекции можно ввести начальные значения для полей или необходимые списки выбора значений.

Описание того, как создавать свой собственный набор пользовательских форм также приведено в Приложении А.

10 Гибкая настройка прав доступа к коллекциям

DSpace обладает гибкой системой прав пользователя, которая тесно связана с применяемой моделью данных. Выделены следующие группы пользователей: администраторы, редакторы, участники процесса внесения, подписчики и пользователи с разрешением просматривать необщедоступные элементы (документы).

Гибкая система прав доступа в DSpace позволяет ограничивать доступ к различным частям репозитория. Каждому разделу можно назначить группу пользователей, которым разрешается доступ к данному разделу. Каждой коллекции назначается множество из отдельных пользователей и групп, которые могут вносить контент для этой коллекции, будут иметь доступ к содержимому, выполнять функции редакторов, или, наконец, администрировать коллекцию.

Пользователь может быть ассоциирован с несколькими группами одновременно. Каждый пользователь группы наделяется соответствующими правами.

Управление группами и отдельными пользователями, а также редактирование прав доступа к разделу или коллекции осуществляется через веб-интерфейс. Доступ к этим функциям имеет пользователь с правами администратора [4].

11 Процедура внесения статей в НЭБ

Для внесения статей в систему нами разработана «Инструкция пользователя. Внесение статей», согласно которой эти действия могут производить только зарегистрированные пользователи системы, наделенные правами вносить и изменять содержимое только строго определенного раздела (или разделов), а также, возможно, правами создания коллекции внутри этого раздела (если администратор системы сочтет это действие возможным). Такие пользователи могут вносить статьи только в существующие в НЭБ номера (выпуски) своего журнала. Каждый выпуск журнала имеет в НЭБ статус КОЛЛЕКЦИИ.

Когда новый пользователь регистрируется в системе, ему выделяется отдельная рабочая область, куда он может загружать свои документы. Существует две возможности инициирования внесения статей в журнал:

1. Выбрать опцию МОЯ БИБЛИОТЕКА и НАЧАТЬ НОВОЕ ВНЕСЕНИЕ, и затем выбрать ту коллекцию, куда будет произведено внесение.

2. Выбрать опцию просмотра РАЗДЕЛЫ И КОЛЛЕКЦИИ на главной странице системы (рис. 2), затем выбрать разрешенный для внесения раздел, коллекцию.

Далее нужно выполнить последовательность шагов в соответствии с конфигурацией внесения выбранной коллекции. Если такая конфигурация специально не определена, то следует процедура по умолчанию:

- выбрать опции, влияющие на множество полей, доступных для ввода на последующих шагах (рис. 4);

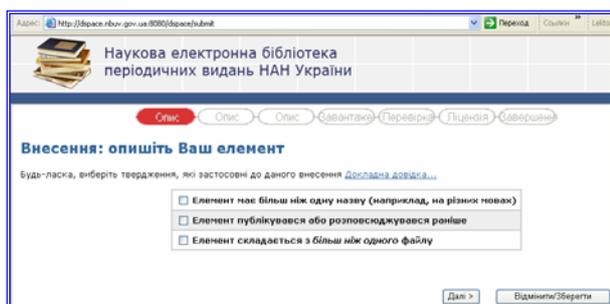


Рис. 4. Процедура внесения: Начало. Описание общих характеристик элемента (статьи)

- ввести значения основных полей метаданных;
- ввести ключевые слова, классифицирующие тему элемента, а также ввести дополнительные поля метаданных;
- загрузить файлы;
- проверить, все ли в порядке с загруженными файлами и их форматом и отредактировать в случае наличия ошибок;
- проверить, все ли в порядке во всех данных, введенных ранее; отредактировать в случае наличия ошибок;
- принять лицензионное соглашение.

Прерывание процесса внесения безопасно с точки зрения сохранности введенных данных.

12 Статистика в DSpace

Для обеспечения эффективного контроля функционирования НЭБ система DSpace включает модуль статистики, позволяющий собирать достаточно подробный статистический анализ использования публикуемой информации:

- количество уникальных посетителей;
- количество визитов в день/месяц;
- количество просмотренных страниц в день/месяц;
- количество визитов страницы, раздела, коллекции, документа;
- поисковые ключевые фразы и слова;
- наиболее популярные страницы и прочее.

По этой информации можно создать различные представления (разные статистические отчеты).

Также для статистического анализа работы системы может быть подключен внешний модуль статистики [10].

13 Заключение

ПО DSpace с открытым исходным кодом применено к специфичному контенту, а именно периодическим изданиям. В настоящее время осуществляется открытый доступ пока только к нескольким научным журналам, наполняемых редакционными коллегами разных организаций НАНУ: <http://dspace.nbu.gov.ua:8080/dspace>.

Архитектура системы выбрана таким образом, что каждое периодическое издание может иметь свою настройку как домашней странички издания, так и полей заполнения метаданных своих ресурсов (статей), а также, возможно, форм вывода (просмотра). Такой подход может быть использован для создания различных многодисциплинарных репозитив (а также тематических), например, научных отчетов многих организаций со 100% открытым доступом, а также 100% предоставлением полных текстов статей.

Что надо сделать?

- Выработка и поддержание политики функционирования НЭБ (регламент подачи статей, со-

- блюдение авторских прав, положение о НЭБ, распоряжение по НАНУ, отслеживание соблюдения политики функционирования).
- Эксплуатация НЭБ (техническое оснащение, обеспечение бесперебойной круглосуточной работы, конфигурирование и настройка НЭБ, отслеживание соблюдения политики пополнения НЭБ).
- Программно-методическое сопровождение (сопровождение программного обеспечения, инструкции, методики, руководства и другие материалы, обучение и консультирование).
- Бесперебойное пополнение НЭБ (решение организационно-технических вопросов, обучение персонала, представление документов в электронном виде, пополнение ЭБ новыми выпусками журналов, соблюдение политики функционирования ЭБ).

Литература

- [1] Мриглод І., Мриглод О. Наука України у світовому інформаційному просторі // Вісник Національної Академії наук України. — 2007 — №10. — С. 3–18.
<http://dspace.nbuv.gov.ua:8080/dspace/bitstream/123456789/562/1/a1-10.pdf>
- [2] Резниченко В.А., Проскудина Г.Ю., Овдий О.М. Создание цифровой библиотеки коллекций периодических изданий на основе Greenstone // Электронные библиотеки. — 2005. — 8. Вып. 6.
<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>
- [3] Новицкий А.В., Кудим К.А., Резниченко В.А., Проскудина Г.Ю. Создание научных архивов с помощью системы EPrints // Проблемы программирования. — 2007. — № 1. — С. 46–60.
<http://eprints.isofts.kiev.ua/157/>
- [4] Кудим К.А., Проскудина Г.Ю., Резниченко В.А. Создание научных электронных библиотек с помощью системы DSpace // Проблемы программирования. — 2007. — № 3. — С. 49–60.
<http://eprints.isofts.kiev.ua/233/>
- [5] Кудим К.А., Проскудина Г.Ю., Резниченко В.А. Сравнительный анализ функциональных возможностей систем электронных библиотек // Проблемы программирования. — 2007. — № 4. — С. 32–48. <http://eprints.isofts.kiev.ua/320/>
- [6] DSpace Federation Web site <http://dspace.org>
- [7] <http://ru.wikipedia.org/wiki/RSS>
- [8] The Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14.
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- [9] Open Source BSD License. Available at <http://www.opensource.org/licenses/bsd-license.php>
- [10] Robert Tansley Google Summer of Code // In. Proc. DSpace User Group Meeting. — Rome, Italy. — Oct 2007.
<http://www.aepic.it/conf/viewabstract.php?id=338&cf=11>
- [11] Fedora – Flexible Extensible Digital Object Repository Architecture. <http://www.fedora-commons.org/>
- [12] Robert Tansley DSpace Architecture: an update// In. Proc. DSpace User Group Meeting. — Rome, Italy. — Oct 2007. 2007
<http://www.aepic.it/conf/viewpaper.php?id=328&cf=11>
- [13] Toward the next generation: Recommendations for the next DSpace Architecture. Written by the DSpace architecture review group John Mark Ockerbloom, chair. — Jan. 2007.
http://wiki.dspace.org/static_files/0/0e/DSpace-recs.pdf
- [14] Резниченко В.А., Проскудина Г.Ю., Овдий О.М. Функциональные возможности современных систем электронных библиотек // Проблемы программирования. Специальный выпуск №2–3. — Материалы конференции УкрПрог '2008. — С. 525–532. <http://eprints.isofts.kiev.ua/420/>
- [15] Customizing and Configuring Submission User Interface.
<http://researchspace.auckland.ac.nz/docs/submission.html>
- [16] Custom Metadata-entry Pages for Submission http://www.dspace.org/index.php?option=com_content&task=view&id=155
- [17] Handle http://www.dspace.org/index.php?option=com_content&task=view&id=149#handles
- [18] Configuring Controlled Vocabularies http://www.dspace.org/index.php?option=com_content&task=view&id=147#controlledvocabulary

On creation the open access scientific digital library for periodical of NASU

V.A. Reznichenko, G.Yu. Proskudina, K.A. Kudim

The basic issues for creation of the open access scientific digital libraries for periodical of NAS Ukraine are considered in the work. The review of some known projects of portals of periodicals is presented. And also features of functionalities of software DSpace with respect to solved task are discussed.

ПРИЛОЖЕНИЕ А

Описание пользовательских форм метаданных

Описание набора страниц, на которых пользователь вводит свои метаданные, называются формой (или набором форм). Форма идентифицируется уникальным символьным именем. В XML-структуре форма разбивается на несколько страниц, каждая из которых представляет собой отдельную веб-страницу для сбора элементов метаданных. Чтобы настроить какую-либо из пользовательских коллекций в DSpace формами ввода, следует войти в карту формы (form-map). Фактически она представляет собой таблицу, связывающую коллекцию с соответствующей формой, соединяя идентификатор Handle коллекции с именем формы. Коллекции идентифицируются при помощи Handle [17], т.к. их имена могут изменяться и они не обязательно уникальны, тогда как идентификаторы Handle уникальны и постоянны. Специальная строка таблицы предназначена для набора форм по умолчанию, которые вызываются в случае, если специально не указан идентификатор Handle коллекции. Они применяются ко всем коллекциям, которые специально не заданы в этой таблице. В XML-примере набор форм назван как традиционной (traditional, для традиционного пользовательского интерфейса системы DSpace), но его можно назвать и иначе.

Структура input-forms.xml

Конфигурационный XML-файл имеет один высокоуровневый элемент <input-forms>, содержащий в определенном порядке три элемента.

```
<input-forms>
  <!-- Соответствие коллекций наборам форм -->
  <form-map>
    <name-map collection-handle="default" form-name="traditional" />
    ...
  </form-map>
  <!-- Определения набора форм -->
  <form-definitions>
    <form name="traditional">
    ...
  </form-definitions>
  <!-- Пары Имя/Значение, используемые в элементах множественного выбора -->
  <form-value-pairs>
    <value-pairs value-pairs-name="common_iso_languages" dc-term="language_iso">
    ...
  </form-value-pairs>
</input-forms>
```

Добавление карты коллекции

Каждый элемент <name-map> в пределах тэга <form-map> связывает коллекцию с именем набора форм. Его атрибут collection-handle определяет Handle-адрес коллекции и атрибут form-name определяет имя набора форм. Например, следующий фрагмент показывает, как коллекция с Handle-идентификатором равным "12345.6789/42" присоединена к набору форм "TechRpt":

```
<form-map>
  <name-map collection-handle="12345.6789/42"
  form-name="TechRpt" />
  ...
</form-map>
<form-definitions>
  <form name="TechRpt">
  ...
</form-definitions>
```

Здесь, в примере input-forms.xml, заложена хорошая идея – хранить определение name-map по умолчанию. В случае, когда для некоторых коллекций такие формы специально не определены, всегда можно получить набор форм по умолчанию.

Получение идентификатора Handle коллекции

Для того чтобы назначить пользовательский набор форм нужно знать идентификатор Handle коллекции. Его можно получить, выбрав просмотр «Разделы и коллекции» в меню домашней страницы системы, а затем найти ссылку нужной коллекции.

Например:

<http://myhost.edu/dspace/handle/12345.6789/42>

Подчеркнутая часть URL есть идентификатор Handle, известный каждому администратору системы. Именно он входит в атрибут collection-handle элемента <name-map>.

Добавление набора форм

Можно добавить новый набор форм путем создания нового элемента формы внутри элемента <form-definitions>. Он имеет атрибут name, значение которого как отмечалось выше, должно совпадать с атрибутом form-name элемента <name-map>, используемого для соответствующей коллекции.

Формы и страницы

Содержание формы есть последовательность элементов страницы. Каждый из которых соответствует определенной веб-странице или форме ввода элементов метаданных, последовательно представленных между начальной страницей «Описание» и финальной страницей «Проверка», представляющую сводный список всех собранных метаданных.

Форма может содержать от одной до шести страниц. Они представлены в том порядке, как они идут в XML-файле. Каждый страничный элемент

должен содержать атрибут `number`, равный порядковому номеру страницы, например:

```
<page number="1">
```

Элемент страницы, в свою очередь, содержит последовательность элементов полей. Каждое поле определяет интерактивный диалог, куда вводится одно из полей элементов метаданных Дублинского ядра.

Композиция полей

Каждое поле содержит следующие элементы, в порядке указания. Также обозначается обязательность данного элемента.

dc-schema (обязательный) – название применяемой схемы, например, для Дублинского ядра – `dc`. Это значение должно совпадать со значением элемента `schema`, определенном в `dublin-core-types.xml`.

dc-element (обязательный) - название элемента ДЯ, вводимого в это поле, например, соисполнитель (`contributor`).

dc-qualifier – квалификатор элемента ДЯ, вводимого в данное поле, например, `contributor.advisor`, значение этого элемента должно быть консультант. Незаполнения означает использование неквалифицируемого ДЯ.

repeatable – принимает значение истина, когда поле может содержать несколько значений, и ложь в противном случае. При указании этого поля как повторяемое (`repeatable`), серверлет пользовательского интерфейса добавит управляющий элемент, чтобы пользователь мог запросить несколько полей для ввода дополнительных значений. Предназначен для использования в произвольных повторяющихся полях, таких как ключевые слова, когда невозможно знать заранее сколько понадобится полей ввода для данного элемента.

label (обязательный) – текст, применяемый для обозначения данного поля, описывающий что нужно вводить, например, «Введите научного руководителя».

input-type (обязательный) – определяет вид интерактивного элемента, который вставляется в форму для сбора значений метаданных ДЯ. Должен принимать одно из следующих значений:

- **onebox** – одно текстовое поле.
- **twobox** – пара простых текстовых полей, используемых для повторяемых значений, например, тема или ключевые слова (`subject`) в ДЯ.
- **textarea** – большой блок для текста, состоящего из нескольких строк, например для резюме.
- **name** - имя с отдельными полями для фамилии и инициалов.
- **date** – календарная дата, требующая, чтобы был внесен по крайней мере год.
- **dropdown** – выбирает значения из ниспадающего списка. Для этого элемента следует также включить значение атрибута `value-pairs-name` для определения списка выбираемых значений меню. Используется для определения ограни-

ченного множества значений, например, для элемента ДЯ язык (`language`).

- **qualdrop_value** – вводит «подходящее значение», включающее как квалификатор из ниспадающего меню, так и произвольное текстовое значение. Используется для ввода полей типа альтернативные идентификаторы и коды для вносимых элементов, например, для элемента ДЯ идентификатор. Как и в случае ниспадающего списка нужно включить значение атрибута `value-pairs-name` для определения списка выбираемых значений меню.

hint (обязательный) – содержит текст, который будет появляться при «наводке», или инструкции, следующие за входными полями. Они могут оставаться пустыми, но все равно должны присутствовать.

required когда этот элемент включен в с каким-либо контентом, это означает что это поле является обязательным для ввода. Если пользователь попытается оставить его пустым, появится предупреждающее сообщение.

Например, `<required> Вы должны ввести название.</required>`. Если оставить этот элемент пустым, это не отмечает это поле как обязательное для ввода. Например, `<required></required>`.

visibility – когда этот необязательный элемент включается вместе со значением, он ограничивает видимость поля в границах, определенных тем значением. Если элемент отсутствует или пустой, поле видимо при любых диапазонах. В настоящее время поддерживаются следующие границы.

- **workflow**: поле видимо только на стадии рабочего процесса внесения. Это используется в тех случаях, когда нужно скрыть от пользователя трудные поля, такие например, как предметная классификация, тем самым облегчая использование системы внесения.
 - **submit**: поле будет только видимо при первоначальном внесении, а не на стадии рабочего процесса.
- Например, `<visibility>workflow</visibility>`.

Пример описания одного из полей ДЯ:

```
<field>
  <dc-schema>dc</dc-schema>
  <dc-element>title</dc-element>
  <dc-qualifier />
  <repeatable>false</repeatable>
  <label>Название</label>
  <input-type>onebox</input-type>
  <hint>Введите основное название материала</hint>
  <required>Вы должны внести основное название!</required>
</field>
```

В системе предусмотрена возможность подключения предметного классификатора [18].

Автоматическое скрытие полей

Следует отметить, что некоторые поля могут быть автоматически скрыты или пропущены, при отображении пользовательской формы. Это зависит от разновидности вносимых документов и происходит потому, что в соответствии с первоначальным описанием документа программа пользовательского интерфейса DSpace пропускает ненужные поля ДЯ.

Например, если пользователь на первой странице «Описание» отмечает, что не будет альтернативных названий (первый из трех флажков), ввод элемента ДЯ `title.alternative` на пользовательской странице внесения автоматически опускается.

Когда пользователь инициирует внесение, DSpace вначале показывает так называемую «страницу начальных вопросов» (рис. 4), которая содержит три обозначенных флажками вопроса:

1. **Документ содержит больше одного названия, например, на разных языках.** Управляет полем `title.alternative`.
2. **Документ был опубликован раньше.** Управляет полями ДЯ:
 - `date.issued`
 - `publisher`
 - `identifier.citation`
3. **Документ состоит из больше чем одного файла.** Не влияет ни на одно входное поле метаданных.

Ответы на первые два вопроса управляют отображением некоторых полей метаданных ДЯ, даже если они определены в качестве полей пользовательской страницы.

И наоборот, если управляемые флажком поля метаданных не определены в пользовательской форме, то для избежания ошибок и недоразумений такой флажок не показан на начальной странице внесения.

Добавление групп значений

Наконец, описание пользовательской формы может потребовать определить «группы значений» для каких-либо полей с входными типами, которые на них ссылаются. Для этого нужно добавить элемент `value-pairs` к содержимому элемента `form-value-pairs`, содержащий атрибуты:

- **value-pairs-name** – имя по которому `input-type` ссылается на этот список.
- **dc-term** – поле квалифицированного ДЯ для которого выбирается значение из этого списка выбора.

Каждый элемент `value-pairs` содержит последовательность пар подэлементов, каждый из которых, в свою очередь, содержит два элемента:

- **displayed-value** – имя, показанное на веб-странице, для ввода меню;
- **stored-value** – запоминаемое значение в элементе ДЯ, когда выбирается данный вход.

В отличие от HTML-тэга `select`, здесь нет возможности обозначить один из входов значением по умолчанию. Поэтому всегда является выбором по умолчанию первое значение.

Пример. Здесь показано меню типов обычных идентификаторов.

```
<value-pairs value-pairs-name="common_identifiers"
dc-term="identifier">
  <pair>
    <displayed-value>№ Прав. Док.</displayed-
value>
    <stored-value>govdoc</stored-value>
  </pair>
  <pair>
    <displayed-value>URI</displayed-value>
    <stored-value>uri</stored-value>
  </pair>
  <pair>
    <displayed-value>ISBN</displayed-value>
    <stored-value>isbn</stored-value>
  </pair>
</value-pairs>
```

Установка пользовательских форм

Веб-приложение DSpace читает пользовательские формы только при запуске, поэтому важно помнить, что следует всегда перезапускать Tomcat (или любой другой используемый контейнер сервлета), чтобы вступили в силу сделанные изменения в файл `input-forms.xml`.

Любые ошибки в синтаксисе или семантике определения формы, типа плохо сформированного XML или ссылки на несуществующие названия полей, вызовет фатальную ошибку в интерфейсе пользователя DSpace.