

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

СОЗДАНИЕ НАУЧНЫХ ЭЛЕКТРОННЫХ БИБЛИОТЕК С ПОМОЩЬЮ СИСТЕМЫ DSPACE

Рассматриваются основные возможности и особенности DSpace как свободно распространяемой системы построения электронных научных библиотек. А также описан опыт построения многоязычной электронной библиотеки на ее основе.

Введение

Данная работа продолжает цикл публикаций [1–2], посвященных созданию электронных библиотек (ЭБ) на основе свободно распространяемого программного обеспечения с открытым кодом. Итогом данных работ стала разработка двух типовых институтских библиотечных веб-сайтов – <http://greenstone.isoftware.kiev.ua>, созданного на основе программного обеспечения (ПО) Greenstone и <http://eprints.isoftware.kiev.ua>, созданного на основе ПО EPrints.

За прошедший период в Институте программных систем НАН Украины изучена и опробована еще одна популярная система для построения научных ЭБ – DSpace. Согласно данным Registry of Open Access Repositories (ROAR) <http://roar.eprints.org> на май месяц 2007 года в мире функционирует 235 электронных архивов, созданных с использованием ПО DSpace. Наиболее крупный из них – DSpace at Cambridge, насчитывающий 188395 записей. Известно, что в Украине есть «Открытый электронный архив гражданского общества» www.e-archive.org.ua, созданный на основе ПО DSpace; в России функционируют две библиотеки на DSpace: Уральского <http://dspace.lib.usu.ru> (общее количество записей – 1265) и Красноярского государственных университетов <https://elib.krasu.ru> (общее количество записей – 422). В настоящее время электронные коллекции, созданные с помощью этой системы, доступны в локальной сети нашего института, которые в дальнейшем предполагается открыть для всеобщего обозрения и доступа. Результатам прове-

денной разработки посвящена данная работа.

DSpace является системой того же класса, что и рассмотренная ранее EPrints [2]. Обе являются бесплатными системами с открытыми исходными кодами, OAI-совместимыми [3], интероперабельными, эквивалентными по функциональности самоархивирования, и написаны одним и тем же программистом из Саутгемптона Робом Тэнсли (Rob Tansley). Сравнительной характеристике этих систем посвящена работа [4].

Предлагаемая работа посвящена описанию ПО DSpace и содержит историю создания и подобные проекты (раздел 2). Описание некоторых аспектов функциональности системы, в частности, описание модели данных, форматы файлов, виды метаданных, характеристики пользователей системы, их группы и права, описание процесса поглощения и занесения документов в систему, поиск и просмотр и др. (раздел 3). Поддержка многоязычности (раздел 4); установка и необходимое ПО (раздел 5); краткое описание проекта (раздел 6).

1. История создания и подобные проекты

Платформа электронных библиотек DSpace разрабатывалась совместно компанией Hewlett-Packard и библиотеками MIT (Massachusetts Institute of Technology). 4-го ноября 2002 года система была запущена как действующая служба, поддерживаемая библиотеками MIT. Также на основании лицензии BSD [5] открыт исходный код с намерением поощрить формирование сообщества открытых кодов вокруг DSpace.

DSpace формувалась под впливом наукових досліджень в області систем електронних бібліотек. Архітектура DSpace походить із фреймворка Кана і Віленського для сервісів розподілених електронних об'єктів [6], а також із робіт Армса і других по архітектурі електронної бібліотеки [7–8]. DSpace ґрунтується також на роботі Лагозе і других по архітектурі FEDORA [9] і прототипі реалізації FEDORA в університеті Вірджинії [10]. Наступна суттєва частина роботи по створенню DSpace пов'язана з OAIS (Open Archival Information System) [11], де визначені всі основні поняття і терміни, що використовуються в DSpace.

Система EPrints [12], розроблена в Університеті Саутгемптона, має багато сподоби з DSpace, але оптимізована так, щоб забезпечити доступ до самостійному депонуванню (внесенню) автором матеріалів (статей, звітів, книг і пр.), в той час як DSpace забезпечує платформу для довготривалого зберігання цифрових матеріалів, що використовуються в академічних дослідженнях. Інтерфейс користувача для депонування матеріалів в DSpace ґрунтується на досвіді, отриманому з розробки і використання інтерфейсу користувача для депонування системи EPrints. Інтегруєбельність з EPrints, в даний час частково може досягатися через використання протоколу OAI для збору метаданих (OAI-PMH) [3, 13], що забезпечує сервіс доступу до різних архівів.

Серед систем того ж класу можна також назвати ПО Greenstone Новозеландського проекту по ЕБ в Університеті Waikato [14] – відкритий інструмент створення ЕБ, головним чином зосереджений на централізованій публікації документів (і інших об'єктів колекціонування), крім того, CERN розробив ПО Сервер Документів CERN (CDSware) [15], який є сервером попередньої публікації наукових документів.

2. Обзор функциональности

DSpace функціонує як централізований сервіс організації. Різні підрозділи в межах установи (лабораторії, центри, школи або відділи) можуть мати свої власні окремі області в межах системи. Члени даних підрозділів безпосередньо вносять контент через веб-інтерфейс користувача, який розроблено так, щоб внесення здійснюється максимально просто. Альтернативно система передбачає імпортування множини елементів для пакетної завантаження контенту.

В кожному підрозділі можна також призначити людей, які можуть переглядати і редагувати внесення перед тим, як вони будуть включені в основний репозиторий. Далі DSpace індексує метадані, що надійшли разом з електронним документом і робить їх доступними згідно з привілеями доступу, визначеними в даному підрозділі.

Функціональні аспекти DSpace [16]:

- для базової організації даних зафіксована визначена модель даних;
- система зберігає і індексує метадані в різноманітних форматах;
- система зберігає інформацію про користувачів системи;
- незважаючи на те, що великі зусилля прикладаються для спрощення доступу до електронних матеріалів установи, тем не менше відкривати повний доступ до вмісту репозитория не завжди вигідно. Крім того, такі функції як депонування і редакційна перевірка повинні бути прив'язані до відповідних користувачів і обмежені ними. Тому система має функцію авторизації;
- система повинна бути спроможна приймати надходять матеріали, такою процесом називається поглинання;
- в деяких випадках потрібно, щоб матеріали або пов'язані з ними метадані, що вносяться в архів, були перевірені або доповнені визначеними користувачами. Ця послідовність

дійствий називається робочим процесом (workflow);

- матеріали в архиві доступні по ссылкам, приведеним в описанні конкретного елемента. По цьому жє описанню можна делати бібліографічніє ссылки на даний матеріал;

- кінечніє пользователи должны иметь возможность просматривать и находить содержимое репозитория. В связи с этим система должна обеспечивать функции поиска и просмотра (навигации);

- для предоставления возможности интегрированного поиска элементов (документов) поддерживается протокол сбора метаданных OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) [3];

- должна существовать возможность уведомления конечных пользователей системы об интересующих их свежих поступлениях в репозиторий, чтобы избавить пользователя от необходимости регулярно посещать репозиторий для выполнения такой проверки. Система предоставляет автоматическую рассылку уведомлений по электронной почте через службу подписки;

- предоставляется возможность обрабатывать данные произвольных форматов, от простых текстовых документов до наборов данных и цифрового видео;

- доступ к перечисленным функциональным возможностям предоставляется посредством веб-интерфейса.

Рассмотрим более подробно некоторые из приведенных функциональных аспектов.

2.1. Модель данных. Способ организации данных в DSpace выбран таким образом, чтобы отразить структуру организации, использующей данную систему (рис. 1).

Каждый сайт DSpace делится на разделы, соответствующие подразделениям организации. Раздел является самым высоким уровнем иерархии DSpace. Разделы могут содержать подразделы, т.е. могут образовывать иерархию. Разделы содержат *коллекции* логически связанных материалов. Каждая коллекция может принадлежать только одному разделу. Коллекция состоит из *элементов*, которые являются основной единицей или “атомом” архивирования. Элемент принадлежит одной и только одной коллекции (связь *принадлежит*), но дополнительно может быть отображен еще и в других коллекциях (связь *включает*). Элемент состоит из сгруппированного, связанного между собой содержимого и соответствующих описаний (метаданных). Метаданные, описывающие элемент, индексируются для навигации и поиска. Далее, элементы, в свою очередь, представляют собой *наборы* (связки) *битовых потоков* (файлов). Цель подобных наборов – хранить тесно связанные файлы вместе. В табл. 1 приведены примеры для каждого типа объекта модели данных.

Каждый элемент имеет запись метаданных квалифицированного Дублинского Ядра (Qualified Dublin Core – QDC). Другие метаданные могут сохраняться в элементе в виде текстового файла, но обязательным набором метаданных для каж-

Таблица 1. Примеры объектов DSpace

Тип объекта	Пример
Раздел (Community)	Институт программных систем
Подраздел (Sub-community)	Отдел компьютерных вычислений
Коллекция (Collection)	Отчеты Публикации конференций
Элемент (Item)	Статья “Сравнение систем электронных библиотек EPrints 3.0 и DSpace 1.4.1”
Набор (Bundle)	HTML файл и связанная картинка, представляющие единый HTML-документ
Файл (Bitstream)	Отдельный HTML файл; отдельный файл изображения
Формат файла (Bitstream Format)	Версия Microsoft Word; изображение в формате JPEG

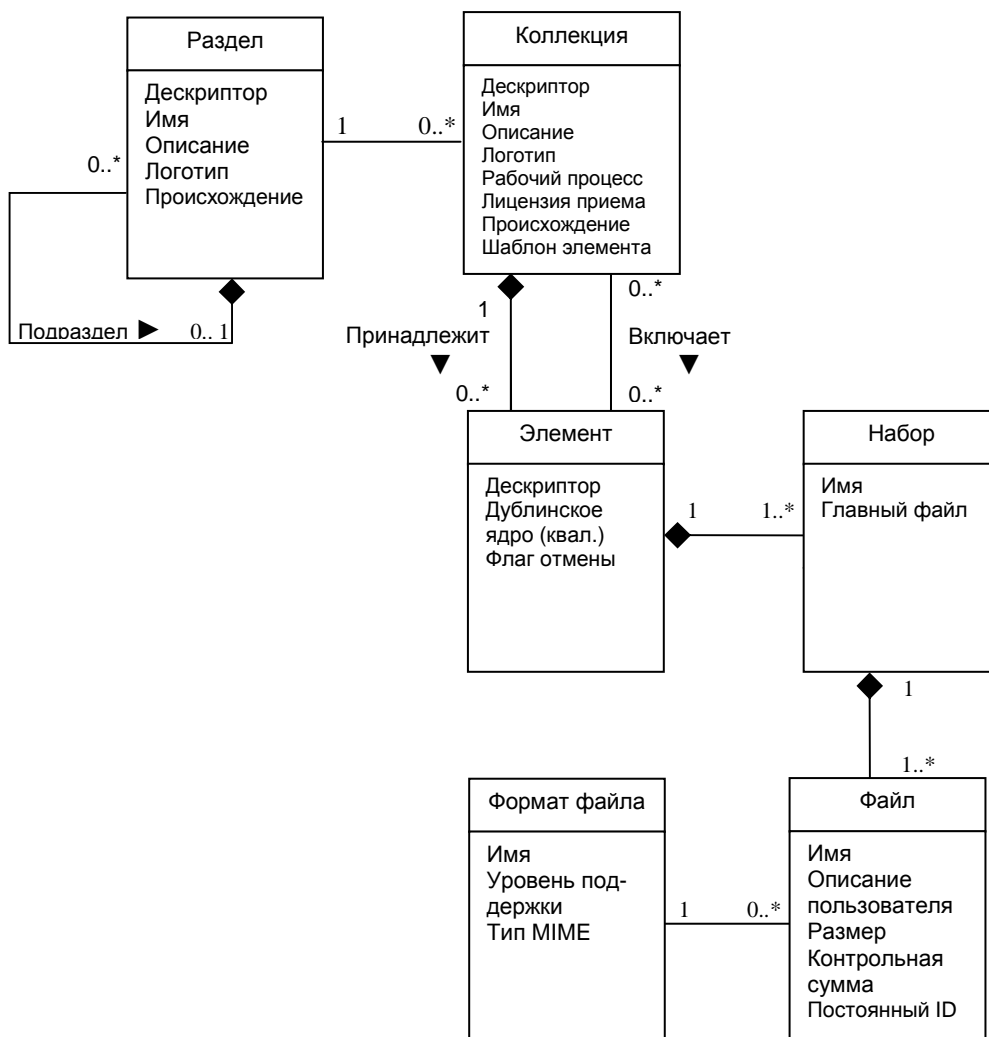


Рис. 1. Модель данных DSpace

дого элемента является QDC, что обеспечивает интероперабельность и простоту нахождения элементов. Метаданные Дублинского ядра могут быть введены конечными пользователями вместе с самими документами или они могут быть получены из других метаданных как часть процесса поглощения (ingest process).

2.2. Форматы файлов. Каждый хранимый в системе файл связан с определенным форматом. Так как служба хранения – основополагающая функция библиотечной системы, то для таких систем существенна возможность распознавать конкретные форматы файлов, загружаемые пользователями. Неотъемлемой частью формата файла является явно или скрыто выраженный способ интерпретации его содержимого.

Список поддерживаемых форматов файлов приведен в табл. 2, в которой представлена система позволяющая сохранить

и предоставить доступ к большинству наиболее известных типов файлов [17].

Таблица 2. Поддерживаемые форматы файлов

DSpace 1.4.1		
Adobe PDF	Mathematica	RealAudio
AIFF	Microsoft Excel	RTF
audio/basic	Microsoft PowerPoint	SGML
BMP	Microsoft Project	TeX dvi
FMP3	Microsoft Visio	Text
GIF	Microsoft Word	TIFF
HTML	MPEG	Video Quick-time
image/png	MPEG Audio	WAV
JPEG	Photo CD	WordPerfect
LateX	Photoshop	XML
MARC	Postscript	

Каждый хранимый в DSpace-системе файл дополнительно характеризуется такими уровнями поддержки:

- поддерживаемый формат распознается и организация предоставляющая

услуги по размещению информации уверена, что файл данного формата можно будет использовать в будущем, используя любую комбинацию методов, соответствующих тем или иным потребностям;

– известный формат распознается и организация предоставляющая услуги по размещению информации может сохранять файл как есть и извлекать его в том же формате. Организация, предоставляющая услуги по размещению информации, попытается получить дополнительную информацию, чтобы модернизировать формат до “поддерживаемого” уровня;

– неподдерживаемый формат нераспознаваем, но организация предоставляющая услуги по размещению информации сохраняет файл как есть и позволяет извлекать его из архива в том же формате.

2.3. Метаданные. DSpace поддерживает три вида метаданных об архивируемом контенте.

Описательные метаданные. Каждый элемент архива имеет одну запись метаданных в формате QDC, который не жестко основан на наборе элементов и квалификаторов профиля библиотечных приложений [18], обеспечиваемого в системе DSpace по умолчанию. Организации с другими требованиями могут его легко изменять, через реестр элементов и квалификаторов.

Другие описательные наборы метаданных, например, UNIMARC, могут поддерживаться в обычных текстовых файлах. Разделы и коллекции имеют несколько простых описательных полей метаданных (имя и несколько описательных предложений), хранящихся в СУБД.

Административные метаданные включают метаданные сохранности (preservation), происхождения (provenance) и политики авторизации данных. Большинство из них хранится в СУБД системы DSpace. Метаданные происхождения хранятся в записях DC. Дополнительно некоторые другие административные метаданные (например, размер файла или MIME-типы) копируются в записях DC так, чтобы они были легко доступны через протокол OAI.

Структурные метаданные включают информацию о том, как представлять элемент или файлы внутри элемента пользователям и отношения между составляющими частями элемента. В качестве примера, рассмотрим диссертацию, состоящую из набора TIFF-изображений, каждое из которых представляет отдельную страницу диссертации. Структурные метаданные должны включать информацию о том, что каждое изображение является одной страницей и о порядке изображений/страниц. Структурные метаданные в DSpace занимают довольно важное место. Внутри элемента, как было вышеописано, файлы могут быть упорядочены в отдельные наборы. Набор может иметь *главный файл* (primary bitstream). Дополнительные структурные метаданные могут сохраняться в простых текстовых файлах.

2.4. Пользователи, группы и права. Многие функции DSpace, например, просмотр и поиск документов в системе, могут выполняться анонимно, но чтобы выполнить внесение документов пользователю нужно зарегистрироваться.

О каждом пользователе DSpace хранит следующую информацию:

- адрес электронной почты;
- фамилию и имя;
- пароль;
- список коллекций, относительно которых пользователь уведомляется о новых поступлениях;
- признак саморегистрации, т.е. создана ли была учетная запись пользователя автоматически системой, например в результате обращения к администратору.

DSpace обладает более продвинутой системой прав пользователя по сравнению с аналогичными библиотечными системами (например, EPrints), тесно связанной с применяемой моделью данных. Выделены такие группы пользователей: депоненты, администраторы, участники процесса депонирования, подписчики и пользователи с разрешением просматривать не общедоступные элементы. Группы пользователей можно также рассматривать в качестве ролей.

Гибкая система прав доступа в DSpace позволяет ограничивать доступ к различным частям архива. Каждому разделу архива можно назначить группу пользователей, которым разрешается доступ к данному разделу. Каждой коллекции назначается множество из отдельных пользователей и групп, которые будут депонентами для этой коллекции, будут иметь доступ к содержимому, играть роль редакторов или администрировать коллекцию.

Пользователь может быть ассоциирован с несколькими группами одновременно. Каждый пользователь группы назначается соответствующими правами. Управление группами и отдельными пользователями осуществляется посредством веб-интерфейса и не требует, во-первых, специальных программистских навыков, во-вторых, доступа к операционной системе, где установлено программное обеспечение DSpace. Аналогично, редактирование прав доступа к разделу или коллекции осуществляется посредством веб-интерфейса. Чтобы иметь доступ к функциям управления, необходимо войти в систему с правами администратора.

Гибкая система авторизации DSpace основана на привязке действий к объектам (*политики ресурсов*) и списков пользователей (групп), которые могут их выполнять. Существует две специальные группы – “администраторы”, где пользователи могут совершать любые действия, и “анонимный”, которая содержит всех пользователей системы. Назначение политики действия для элемента группе пользователей “анонимный” означает, что каждый пользователь системы может совершать это действие. Например, большинство элементов DSpace имеют политику для анонимных пользователей ЧТЕНИЕ.

Чтобы пользователь мог выполнить действие с объектом системы, он должен иметь разрешение, которое должно быть задано явно. В табл. 3 и 4 представлен перечень возможных действий, которые понимает система авторизации и политики ресурсов.

Таблица 3. Возможные действия в DSpace

ЧТЕНИЕ (READ)	Действие связано с тем, что пользователь узнает о существовании объекта в системе, и просматривает какие-либо связанные с ним мета данные
ЗАПИСЬ (WRITE)	Изменение метаданных, связанных с объектом. Здесь не включена возможность удаления
ДОБАВИТЬ (ADD)	Действие добавления объекта (например, элемента) в контейнер (например, в коллекцию). Чтобы внести элемент в коллекцию, пользователь должен иметь разрешение ДОБАВИТЬ в этой коллекции
ИЗЪЯТЬ (REMOVE)	Действие, связанное с изъятием объекта из контейнера

Обращаем внимание, что здесь нет действия УДАЛИТЬ. Для того, чтобы “удалить” объект из архива (например, элемент) нужно иметь разрешение ИЗЪЯТЬ на все объекты (в данном случае, коллекцию), которая его содержит.

В идеале, ничто поступающее в систему DSpace не должно когда-либо удаляться; однако, практические и юридические факторы иногда требуют удаления. Элементы могут быть удалены из DSpace одним из двух способов. Они могут быть “изъяты”, что означает, что они остаются в архиве, но полностью скрыты для обозрения. В этом случае, если пользователь пытается получить доступ к изъятому элементу, ему посылается сообщение, указывающее, что элемент был удален. В другом случае, элемент может также быть “удален”, когда все его следы удалены из архива.

2.5. Процесс поглощения и занесения. На рис. 2 показан процесс поглощения в системе DSpace. Приложение “пакетный импорт элементов” (Batch Item Importer) преобразовывает внешний SIP (Submission Information Package – пакет подачи информации) [11] (XML-документ метаданных и несколько файлов контента) в объект “процесса приема” (In Progress Submission). Аналогично для сборки объекта “процесса приема” используется передача элементов конечным пользователем через веб-интерфейс.

Таблиця 4. Політики ресурсів в DSpace

Раздел (Community) ADD/REMOVE	Добавить или изъять коллекции или подразделы
Коллекция (Collection) ADD/REMOVE	Добавить или изъять элементы (ADD=разрешение на занесение элемента)
DEFAULT_ITEM_READ	Наследуется как READ всеми вносимыми элементами
DEFAULT_BITSTREAM_READ	Наследуется как READ файлами всех вносимых элементов
COLLECTION_ADMIN	Администраторы коллекции могут редактировать элементы в коллекции, изымать, отображать другие элементы в данной коллекции
Элемент (Item) ADD/REMOVE	Добавить или изъять наборы
READ	Можно просматривать элемент (метаданные элемента всегда видны)
WRITE	Можно модифицировать элемент
Набор (Bundle) ADD/REMOVE	Добавить или изъять файлы в наборе
Файл (Bitstream) READ	Просмотреть файл
WRITE	Модифицировать файл

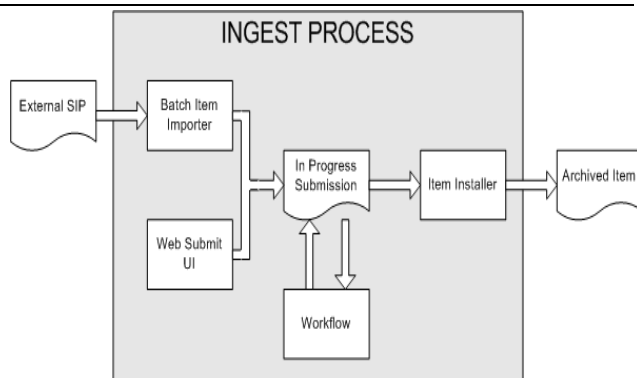


Рис. 2. Процесс поглощения в DSpace

В зависимости от политики коллекции, в которую нацелено внесение, может стартовать рабочий процесс (workflow), что позволяет редакторам или контролерам осуществлять проверку вносимых материалов и гарантировать его пригодность для включения в коллекцию.

Когда завершен процесс внесения пакетный или через веб-интерфейс, то включается следующий шаг поглощения – рабочий процесс (workflow) либо “инсталлятор элемента” (Item Installer). В DC добавляется сообщение о “происхождении” (provenance message), которое включает имена файлов и контрольные суммы вносимого контента. Подобным же образом, всякий раз, когда рабочий процесс изменяет состояние (например, редактор разрешил внесение), добавляется аналогичное предложение “происхождения”. Это позволяет проследить, как элемент изменился с тех пор, как его внес пользователь. Также в DSpace применяется система историй (History system), однако “происхождение” облегчает получение доступа в конкретный момент.

Как только успешно завершен рабочий процесс, объект “процесса приема” поглощается “инсталлятором элемента”, который конвертирует его в полностью соответствующий репозиторию DSpace элемент. “Инсталлятор элемента”:

- назначает дату внесения;
- добавляет значение метаданных “date.available” в запись метаданных DC;
- добавляет дату выпуска, если ничего другого нет;
- добавляет, сообщение “происхождения” (включая контрольные суммы файла);

- назначает постоянный идентификатор дескриптора (Handle);
- добавляет элемент в целевую коллекцию, и добавляет соответствующую политику авторизации;
- добавляет новый элемент к индексам поиска и просмотра.

Когда новый пользователь регистрируется в системе, ему выделяется отдельная рабочая область, куда он может загружать свои документы. Последовательность занесения в DSpace описана далее (конфигурация по умолчанию):

1. Выбрать коллекцию, где будет находиться элемент.
2. Выбрать опции, влияющие на множество полей, доступных для ввода на последующих шагах.
3. Ввести значения основных полей метаданных.
4. Ввести ключевые слова, классифицирующие тему элемента, а также ввести дополнительные поля метаданных.
5. Загрузить файлы.
6. Проверить, все ли в порядке с загруженными файлами и их форматом; отредактировать в случае наличия ошибок.
7. Проверить, все ли в порядке во всех данных, введенных ранее; отредактировать в случае наличия ошибок.
8. Принять лицензионное соглашение.
9. Процесс депонирования может быть приостановлен в любой момент времени без потери введенных данных.

2.6. Поиск и просмотр. DSpace позволяет конечным пользователям обнаруживать контент множеством способов:

- по внешней ссылке, например, дескриптору (Handle);
- поиск по одному или нескольким ключевым словам в метаданных или извлеченном полном тексте;
- просмотр по индексам названия, автору, дате и предметной теме.

Поиск – существенный компонент обнаружения объектов в любой библиотечной системе. Учитывая высокие пользовательские требования к поисковым машинам, предоставление как можно большего числа поисковых возможностей – важная особенность DSpace. Модуль ин-

дексации и поиска DSpace имеет очень простой API, который предоставляет возможность индексации нового контента, регенерации индекса и выполнения поиска по всему архиву, разделу, подразделу или коллекции. За API стоит свободно распространяемая поисковая Java-машина Lucene. Lucene предоставляет поиск по полям метаданных, удаление стоп-слов, выполнение поиска слова во всех его морфологических формах, возможность инкрементно добавлять новый индексируемый контент без регенерации всего индекса. Индексы поиска конфигурируемы, что позволяет организациям настраивать индексируемые поля метаданных. Система предоставляет простой и расширенный поиск (рис. 3).

Следующий важный механизм обнаружения материалов в DSpace является просмотр – процесс где пользователь просматривает конкретный индекс, например, индекс названия, и в процессе поиска интересующих элементов проводит по нему навигацию. Подсистема просмотра обеспечивает простой API, где предоставляется возможность определить индекс и его подраздел. Затем подсистема просмотра раскрывает часть интересующего индекса. Индексы, которые могут быть просмотрены: название элемента (рис. 4), автор элемента, дата выпуска элемента и предметные термины. Дополнительно, просмотр может ограничиваться до элементов внутри определенной коллекции или раздела.

2.7. Поддержка OAI и плагины переходов. Инициатива «Открытые архивы» (OAI, Open Archives Initiative) разработала протокол для сбора метаданных OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) [13]. Это позволило сайтам программно извлекать и собирать метаданные из нескольких источников и предлагать сервисы, используя эти метаданные, например, индексацию или сервисы связи. Такой сервис позволяет пользователям получать доступ к информации от большого числа сайтов, объединенных в центральном каталоге.

Інформаційні системи

DSpace виставляє метаданні в форматі DC для публічно доступних елементів. Дополнительно, структура розділа і колекції також виставлена через механізм наборів протокола. Для забезпечення

такої функціональності використовується структура з відкритим вихідним кодом OAICat організації OCLC (Online Computer Library Center) [19].

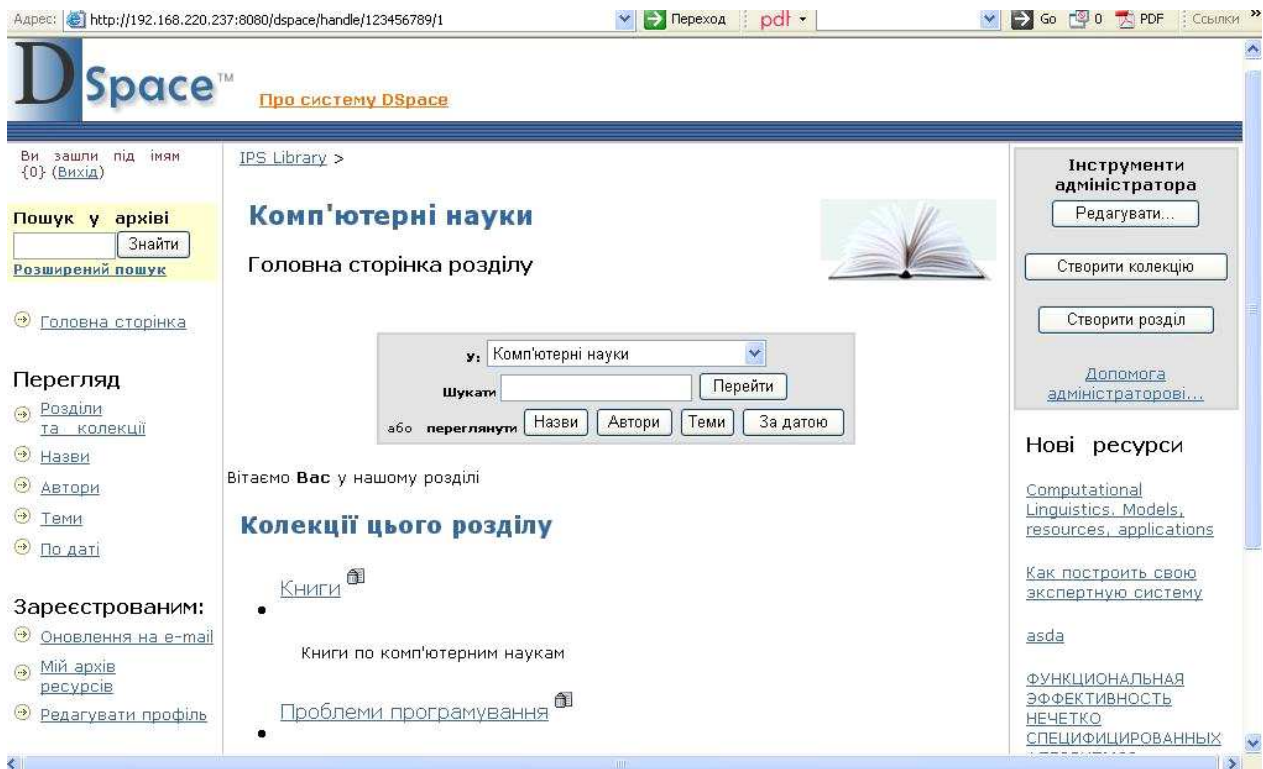


Рис. 3. Главная страница раздела

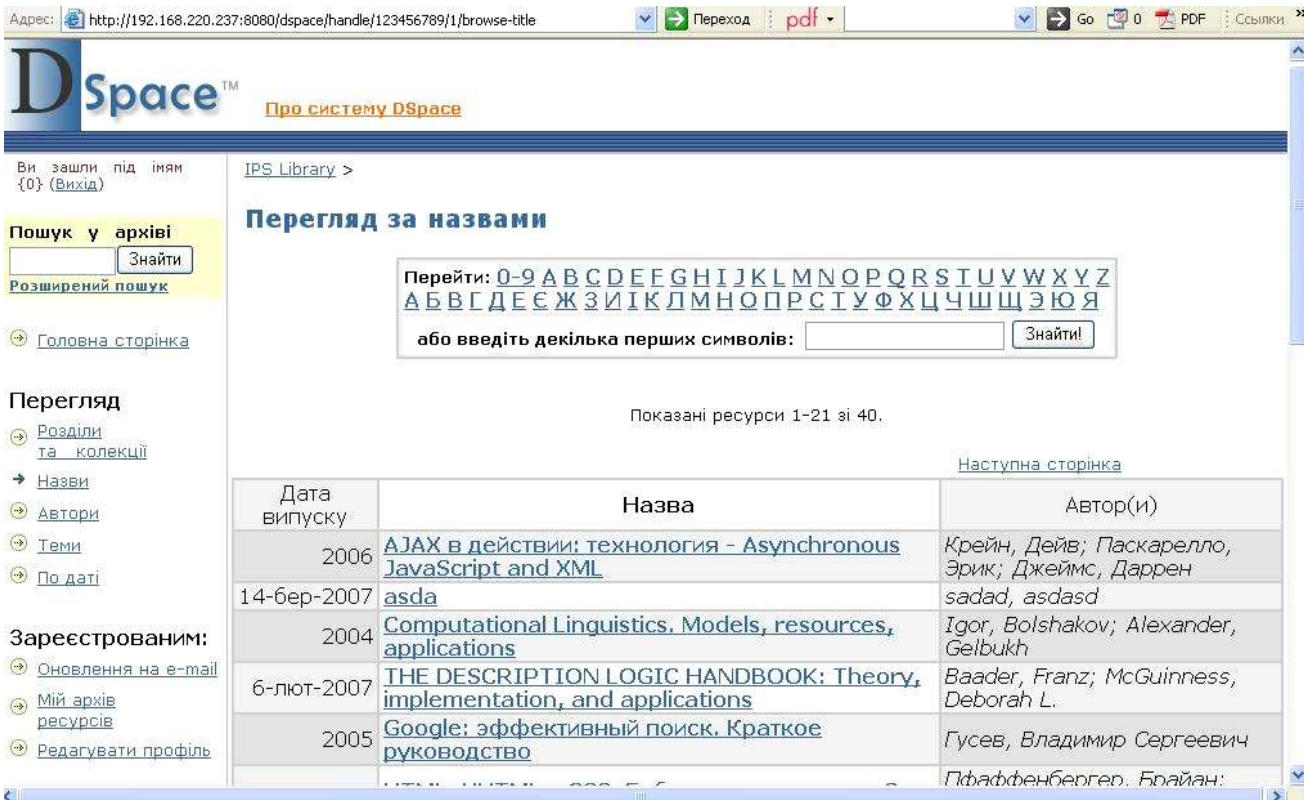


Рис. 4. Просмотр по названиям

Можно также конфигурировать сервис OAI воспользовавшись плагинами перехода (crosswalk plugins) для того, чтобы предложить дополнительные форматы метаданных, например MODS.

Плагины переходов предусмотрены для реализации функций экспорта и импорта в системе. Эти программные модули осуществляют перевод между метаданными DSpace-объектов и определенным внешним представлением. Как правило, это пары плагинов для импорта и экспорта. Например, из формата метаданных MODS во внутренний формат DSpace и наоборот. Используемые плагины перечислены в конфигурационном файле.

Инсталляция DSpace 1.4.1 включает следующие плагины переходов, используемые в рамках протокола OAI-PMH:

- METS – стандарт кодирования и передачи метаданных [20];
- MODS – схема метаданных описания объекта [21];
- QDC – вышеупомянутое квалифицированное Дублинское Ядро, основной набор метаданных системы DSpace [18];
- DIDL – Digital Item Declaration Language, с помощью которого в MPEG-21 описываются сложные электронные объекты [21].

2.8. Подписка. Вышеотмечено, что конечные пользователи могут “подписаться” на коллекции через пользовательский веб-интерфейс, для того чтобы быть уведомленными о поступлении новых элементов в эти коллекции. Ежедневно конечные пользователи, подписанные на одну или несколько коллекций, получают по электронной почте краткое сообщение обо всех новых поступлениях в какую-либо из подписанных коллекций за предыдущий день. Если ни в одну из подписанных коллекций ничего не поступило, то сообщение не посылается. В любое время пользователи могут отказаться от подписки.

2.9. История. Информация о происхождении, предоставляемая в повествовательной форме, очень полезна, но она программно не управляема. Система истории фиксирует временные характеристики

существенных изменений, происходящих в DSpace.

В настоящее время, подсистема истории вызывается явно, когда происходят существенные события (например, DSpace принимает элемент в архив). В этом случае подсистема истории создает данные в формате RDF [22], описывая текущее состояние объекта. RDF-данные моделируются с использованием ABC-модели [23], онтологии для описания временных данных, и сохраняются в файловой системе.

3. Поддержка многоязычности

В DSpace приемлемый уровень локализации достигается с помощью сторонних дополнений. Прежде всего опишем основную версию системы, а затем дополнения и исправления от сторонних разработчиков.

В DSpace есть только один ресурсный файл, содержащий большинство фраз интерфейса. Чтобы локализовать интерфейс, необходимо иметь перевод этого файла с видоизмененным именем. Когда пользователь входит в систему через веб-интерфейс, язык определяется автоматически по предпочтениям, указанным в веб-браузере. Переключение языков недоступно. Шаблоны писем электронной почты и файлы помощи не включены в локализацию. Нет также возможности ввести новости и описания коллекций на разных языках.

Однако, существует несколько сторонних дополнений к системе, добавляющих переключение языков, возможность локализовать сообщения электронной почты и тексты справки. Имена коллекций и разделов, их описания, новости сайта остаются непереводаемыми.

Значения метаданных могут быть добавлены на нескольких языках, хотя и не в обычном процессе депонирования, а на дополнительном шаге редактирования метаданных (это означает, что нужно провести дополнительные необычные манипуляции). Тем не менее, только первое из введенных значений используется для отображения на экране, вне зависимости от языка. Единственное исключение – воз-

возможность вводить альтернативные заглавия документа.

4. Установка и необходимое ПО

Установка и изначальная конфигурация для обеих систем может быть выполнена в течение одного рабочего дня при наличии опыта установки программного обеспечения на базовой операционной системе и собственно установочных файлов системы электронных библиотек. А также свободно распространяемого программного обеспечения:

- *Операционная система.* Unix-подобная.
- *Веб-сервер.* Apache Tomcat или эквивалент.
- *Сервер баз данных.* PostgreSQL или Oracle.
- *Библиотеки языков программирования.* Java, Apache Ant.

После установки программного обеспечения, необходимого для DSpace, систему распаковать и сконфигурировать, отредактировав конфигурационный файл. После чего добавить задания в cron таблицу: периодическая очистка базы данных, индексация, рассылка почты и сбор статистики.

Заключение

Анализ ряда программных продуктов, реализующих системы электронных библиотек был проведен в рамках проекта по информатизации НАН Украины “Разработка проектных решений автоматизированного библиотечного сервиса в интересах организаций НАН Украины”. Выбрано несколько систем, среди которых рассматривались DSpace и EPrints. Обе системы успешно установлены и запущены как действующие службы в локальной сети DSpace, а EPrints доступен по адресу <http://eprints.isoftware.kiev.ua>.

1. *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>.

2. *Новицкий А.В., Резниченко В.А., Проскудина Г.Ю.* Создание научных архивов с помощью системы EPrints. Электронные библиотеки. – 2006. – Том 9. – Вып. 4. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2006/part4/Novitski>
3. *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>
4. *Kudim K.A., Proskudina G.Yu.* Comparison of EPrints 3.0 and DSpace 1.4.1 digital library systems // Pros. of the 12-th Conf. EURASLIC, 2-4 May 2007 (in print).
5. *Open Source BSD License.* Available at <http://www.opensource.org/licenses/bsd-license.php>
6. *Kahn Robert, Wilensky Robert:* A Framework for Distributed Digital Object Services, May 1995. <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>
7. *Arms William Y.* Key Concepts in the Architecture of the Digital Library, D-Lib Magazine, July 1995. <http://www.dlib.org/dlib/July95/07arms.html>
8. *Arms William Y., Blanchi Christophe, Overly Edward A.* An Architecture for Information in Digital Libraries, D-Lib Magazine, February 1997. <http://www.dlib.org/dlib/february97/cnri/02arms1.html>
9. *Payette Sandra, Lagoze Carl* Flexible and Extensible Digital Object and Repository Architecture, in Christos Nikolau and Constantine Stephanidis, eds., Research and Advanced Technologies for Digital Libraries: Proc. of the Second European Conference, ECDL '98, Heraklion, Crete, Greece, September 21-23, 1998, G. Goos J. Hartmanis, J. van Leeuwen, eds., Lecture Notes in Computer Science, 1513. Berlin: Springer, 1998. <http://www.cs.cornell.edu/payette/papers/ECDL98/FEDORA.html>
10. *Staples Thornton, Wayland Ross, Virginia Dons* FEDORA: A Prototype for a Digital Object Repository. D-Lib Magazine 6, 7/8. July/August 2000. <http://www.dlib.org/dlib/july00/staples/07staples.html>
11. *Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS), CCSDS 650.0-B-1, Blue Book, Issue 1, January 2002.* <http://public.ccsds.org/publications/archive/650x0b1.pdf>
12. *GNU EPrints Software.* <http://software.eprints.org/>

13. *Open Archives Initiative*. <http://www.openarchives.org/>
14. *The Greenstone Digital Library Software* <http://www.greenstone.org/>
15. *CERN Document Server Software (CDSoftware)*. <http://cdsware.cern.ch/>
16. *Tansley R., Bass M., Stuve D., Branchofsky M., Chudnov D.* The DSpace Institutional Digital Repository System: Current Functionality. In Proc. of JCDL 2003.
17. *A Guide to Institutional Repository Software*. 3rd Edition. Open Society Institute. 2004. http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf
18. *Dublin Core Library Application Profile*. <http://dublincore.org/documents/2002/09/24/libraryapplication-profile/>
19. *OCLC Research OAICat Open Source Project*. <http://www.oclc.org/research/software/oai/cat.shtm>
20. *Metadata Encoding and Transmission Standard (METS)*. <http://www.loc.gov/standards/mets/>
21. *Understanding Metadata*. National Information Standards Organization. 2004. <http://www.niso.org>
22. *W3C Resource Description Framework (RDF)*. <http://www.w3.org/RDF/>
23. *Lagoze C., Hunter J.* The ABC Ontology and Model // in Proc. of the International Conference on Dublin Core and Metadata Applications 2001. – Tokyo: 2001. – P. 160 – 176.

Получено 24.04.2007

Об авторах:

Резниченко Валерий Анатольевич,
кандидат физико-математических наук,
старший научный сотрудник,

Проскудина Галина Юрьевна,
научный сотрудник,

Кудим Кузьма Алексеевич,
аспирант Института программных систем
НАН Украины.

Место работы авторов:

Институт программных систем
НАН Украины,
03187, Киев-187,
проспект Академика Глушкова, 40.
Тел. (044) 526 5139, 526 6033
Email: reznich@isofts.kiev.ua
gupros@isofts.kiev.ua
kuzma@isofts.kiev.ua