

Создание научных архивов с помощью системы EPrints

*Новицкий А.В., Резниченко В.А., Проскудина Г.Ю.
Институт программных систем НАН Украины*

Аннотация

В работе рассматриваются основные возможности и особенности EPrints как свободно распространяемой системы построения электронных научных архивов на основе Веб. Описан также первоначальный опыт построения многоязычной электронной библиотеки с особенностями адаптации под определенные задачи и требования.

1 Введение

Университеты и исследовательские институты во всем мире активно планируют и реализуют архивы своей научной продукции. Кроме того, Веб-механизм предоставляет новые возможности для своевременного распространения научной информации. Нашей целью является помочь научным организациям страны в создании таких архивов (репозитариев). Правильный выбор программного обеспечения (ПО), которое бы максимально полно удовлетворило потребности учебных и научных организаций, при минимуме прямых и косвенных затрат – один из аспектов решения данной задачи [1].

EPrints – одна из наиболее широко распространенных систем, которая используется для формирования и управления открытыми архивами [2] и предназначена для создания архивов научных исследований с большим разнообразием информационных ресурсов (научные статьи, отчеты, диссертации, монографии, учебно-методические пособия, материалы конференций, данные результатов экспериментов и наблюдений и т.п.).

Открытые архивы, созданные в среде EPrints, поддерживают протокол обмена метаданными OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) [3], который обеспечивает глобальные услуги доступа и поиска. Инициатива «Открытые архивы» (OAI, Open Archives Initiative) разрабатывает и продвигает стандарты интероперабельности с целью эффективного распространения электронных ресурсов, а также повышения доступности обмена научной информацией.

Кроме того, открытые архивы, созданные в EPrints, могут быть зарегистрированы глобальными OAI сборщиками данных (harvesters), например: scirus (<http://www.scirus.com>), myOAI (<http://www.myoai.com>), OAIster (<http://oai.umd.umich.edu>). Они предоставляют глобальные услуги поиска научных публикаций для всех зарегистрированных учреждений, например, OAIster на сегодня имеет данные более чем от 500 университетов и исследовательских организаций всего мира.

Программное обеспечение EPrints работает на операционной системе (ОС) Unix при поддержке Apache, MySQL, Perl. Все это – свободно распространяемое ПО, регистрация провайдера данных в OAI также бесплатна.

В работе описан пример реализации электронной библиотеки Житомирского Государственного Университета им. Ивана Франка с использованием ПО EPrints. Данный проект был зарегистрирован в качестве провайдера данных в OAI и был выбран провайдером сервисов SAIL-EPrints Service Provider (<http://EPrints.bo.cnr.it/>).

Работа выполняется в рамках проекта по информатизации НАН Украины по теме „Разработка и реализация проектных решений автоматизированного библиотечного сервиса в интересах организаций НАНУ”.

Предлагаемая статья содержит: информационную модель протокола OAI-PMH (Раздел 2); краткое описание системы EPrints (Раздел 3); ввода записей в систему EPrints (Раздел 4); описание конфигурационных файлов (Раздел 5); пример архива EPrints (Раздел 6) и заключение (Раздел 7).

2 Информационная модель OAI-PMH

Исходя из того факта, что источником возникновения протокола была электронная публикация, информационная модель OAI-PMH в общем случае интерпретируется в терминах библиографических данных, описывающих академический ресурс, хотя возможны и другие интерпретации [3]. Протокол OAI-PMH имеет простую и гибкую информационную модель (Рис. 1).

В верхней части модели – описываемый ресурс. Это может быть как традиционный библиотечный объект (например, книга, статья), так и другие сущности (например, изображения, понятия). Ниже идет *элемент* – составляющая репозитария, с помощью которой распространяются метаданные о ресурсе. Элемент концептуально является контейнером, который хранит или динамически генерирует метаданные об отдельном ресурсе в нескольких форматах, каждый из которых может быть собран в виде *записи* через OAI-PMH. Каждый элемент имеет *идентификатор записи* или уникальный идентификатор. Уникальный идентификатор однозначно определяет элемент в репозитарии, он используется в запросах OAI-PMH для извлечения метаданных из элементов. Элемент может содержать метаданные в нескольких форматах. Уникальный идентификатор указывает на элемент, и все возможные записи, имеющиеся в одном элементе, совместно используют один и тот же уникальный идентификатор. Записи описывают ресурс в произвольном формате метаданных, который может быть выражен в XML Schema. Протокол OAI-PMH требует обязательного включения в описание ресурса набора метаданных Дублинское ядро (ДЯ). Также желательно включать в описание и более расширенные наборы метаданных (например, MARC).

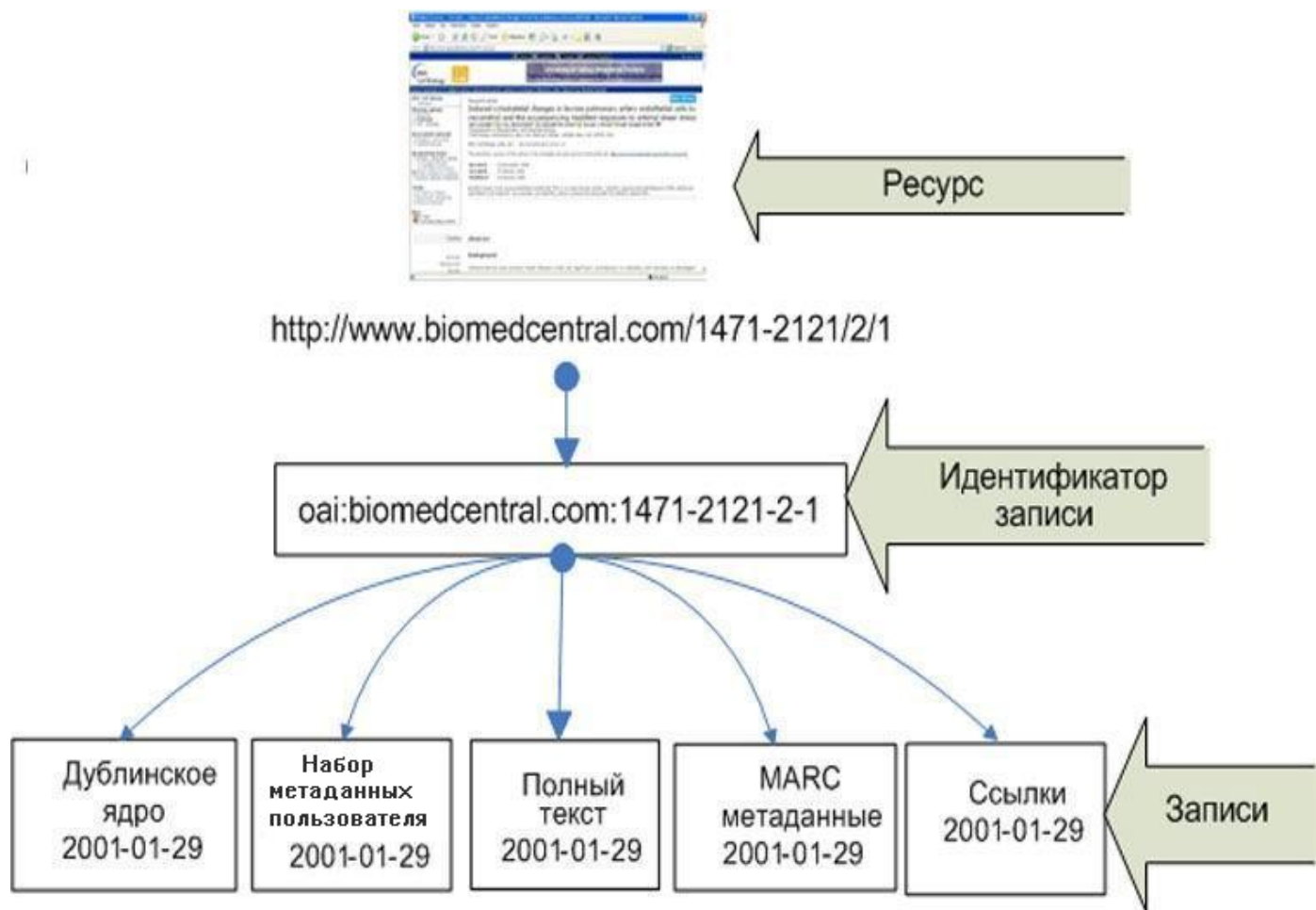


Рис. 1. Информационная модель OAI-PMH

Необходимо подчеркнуть, что идентификатор записи не есть идентификатор документа (объекта). Очевидно, что многие пользователи захотят получить доступ к полному тексту ресурса, описанному записью метаданных. Протокол рекомендует, чтобы архивы использовали элемент записи метаданных для связывания записи с идентификатором (URL, URN, DOI и пр.) ассоциированного документа (объекта). Для этой цели обязательный формат D4 предоставляет атрибут «идентификатор» (DC.Identifier).

3 Краткое описание EPrints

EPrints - свободно распространяемое программное обеспечение под лицензией GNU, используемое для формирования и управления Открытыми Архивами. К настоящему времени в мире создано с использованием EPrints более 200 архивов с более чем 200 000 записями. ПО EPrints может использоваться для создания архивов научных исследований с большим разнообразием информационных ресурсов (научные статьи, отчеты, диссертации, монографии, учебно-методические пособия, материалы конференций, данные результатов экспериментов и наблюдений и т.п.).

ПО EPrints разработано в Школе электроники и информатики Университета Саутгемптона (Великобритания). С созданием системы EPrints тесно связан проект TARDIS (Targeting Academic Research for Deposit and Disclosure) [4], основной задачей которого было исследование всех сторон создания электронного архива с целью разработки типового архива для академических учреждений.

Основными системными требованиями для EPrints версии 2.3.13.1 являются: ОС Unix, язык программирования Perl 5.8.x, сервер баз данных MySQL 4.1.x, веб-сервер Apache 1.x, 2.0.55.

Аппаратные требования - сервер с объемом ОЗУ 1 Гб и процессором с тактовой частотой более 1 ГГц с соответствующим дисковым пространством для хранения полнотекстовых документов желательно с поддержкой SCSI (Small Computer Systems Interface) при большой нагрузке на сервер.

Данный программный продукт отвечает основным требованиям, выдвигаемым к электронным библиотекам (ЭБ) учебных и научных организаций, что включает:

- создание электронных архивов информационных ресурсов разного вида;
- обеспечение быстрого доступа к ЭБ;
- предоставление результатов исследований широкому кругу пользователей;
- сохранение результатов научных исследований;
- обеспечение обмена информацией на глобальном уровне [3].

3.1 Функции и возможности

ПО EPrints предоставляет следующие возможности [5]:

- создание электронных архивов;

- поддержка файлов разного формата;
- индексация файлов PDF, ASCII, Microsoft Word, HTML;
- просмотр формул в документах, созданных на языке LaTeX;
- выполнение полнотекстового и расширенного поиска (по метаданным);
- гибкое администрирование прав доступа;
- гибкая интеграция с основным сайтом (с использованием основного стиля оформления веб-сайта организации).

EPrints имеет подробную документацию по всем аспектам проекта. Сайт демонстрации *demoprints.EPrints.org* предоставляет разнообразную интерактивную помощь. Кроме того, документация использует технологию Wiki (<http://www.wiki.org>), где пользователи EPrints размещают практические советы, сценарии и другую полезную информацию.

Инсталляция ПО выполняется конфигурационными Perl-скриптами. Настройка архива также выполняется скриптами. Кроме того, для установки в данной ОС необходимо иметь права root. Это связано с установкой необходимых модулей Perl. Инсталляция довольно простая, со многими сервисными уведомлениями.

В системе EPrints разработана стандартная конфигурация, которая подходит для большинства архивов. Она касается полей метаданных и типов документов.

Сообщение о последних поступлениях в архив может быть выполнено по протоколу RSS (RDF Site Summary, <http://www.rssreader.com>).

В системе реализована поддержка протокола обмена метаданными OAI-PMH версии 2.

3.2 Примеры применения

С помощью EPrints созданы такие мощные библиотеки, как «DLR electronic library» (<http://elib.dlr.de/>, Германия), с общим количеством записей на момент написания около 42 тыс. (Рис. 2), ЭБ немецкого аэрокосмического центра. Электронный репозиторий «Archive of European Integration - AEI» (<http://aei.pitt.edu/>) - архив исследовательских материалов по проблемам европейской интеграции и объединения с общим количеством записей – 4.5 тыс. Статистика наполнения архива приведена на Рис. 3.

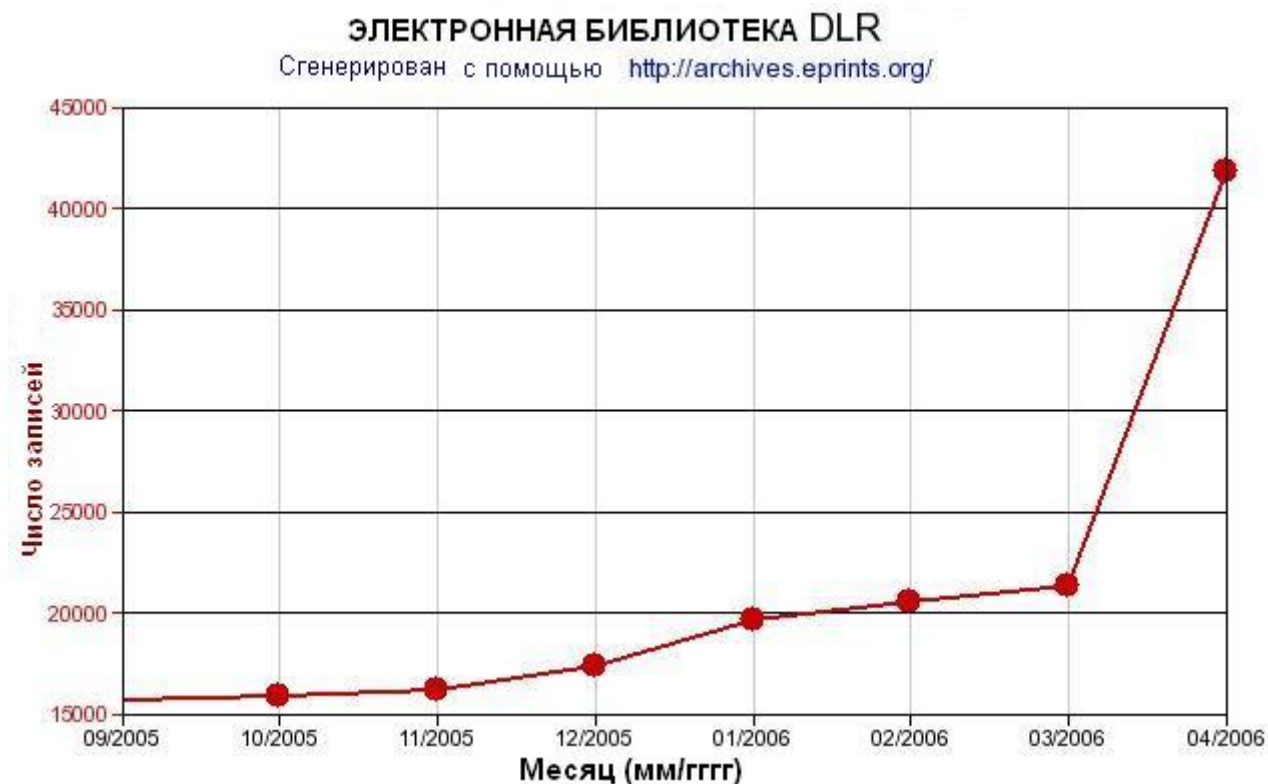


Рис. 2. Статистика наполнения ЭБ DLR

ЭЛЕКТРОННЫЙ АРХИВ "Archive of European Integration"

Сгенерирован с помощью <http://archives.eprints.org/>

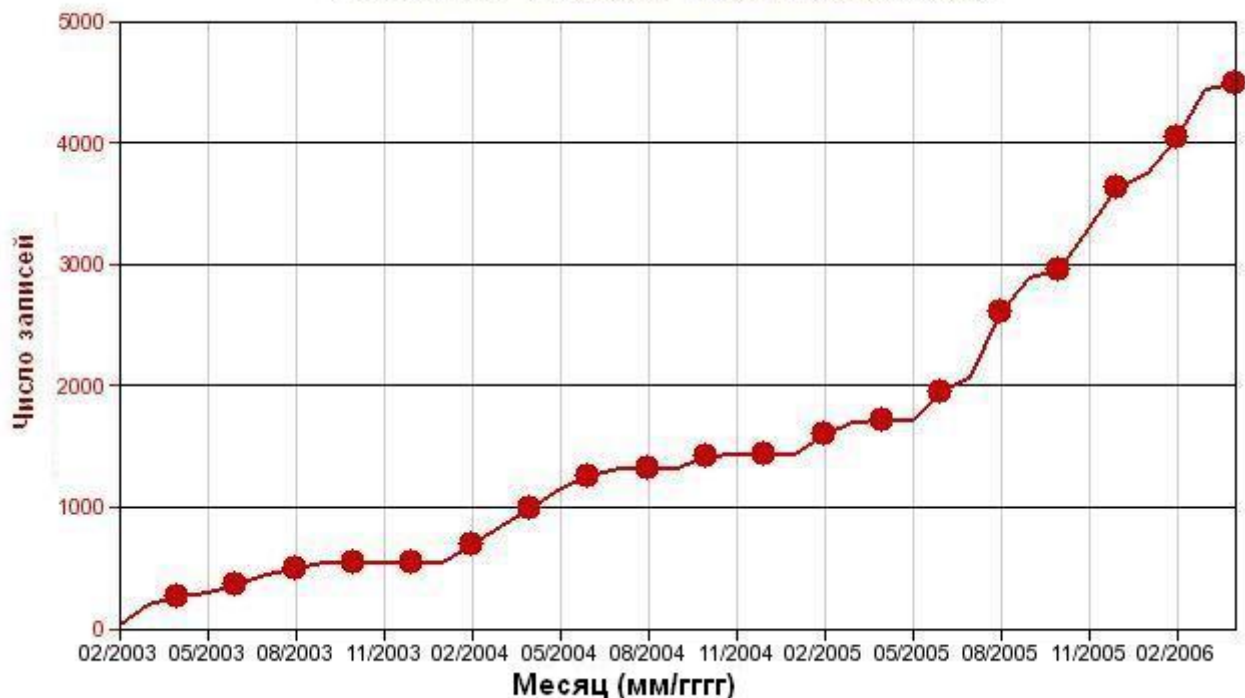


Рис. 3. Статистика наполнения электронной библиотеки AEI

В настоящее время в мире сложилась следующая ситуация по использованию ПО EPrints (Рис.4). На рисунке хорошо видна мировая тенденция не к увеличению числа архивов, а к увеличению количества записей в них (конец графика). Число архивов растет незначительно, в то время как основное внимание сконцентрировано на их наполнении. Более 50% мировых университетских репозитариев используют EPrints; он имеет только одного серьезного конкурента Dspace [6].

Рост количества EPrints архивов и записей

Сгенерирован с помощью <http://archives.eprints.org/>



Рис. 4. Динамика зависимости построения архивов с использованием ПО EPrints и количества записей в них

Еще один пример практического применения EPrints - <http://eprints.agentlink.org> (Рис. 5). Ресурс общего и свободного доступа AgentLink выполняет функцию единого центра сбора (объем 1.5 тыс. записей) и обмена информацией по исследованиям, направленным на внедрение агентных технологий. Все публикации индексируются и классифицируются согласно AgentLink с целью упрощения доступа пользователей к необходимым материалам.



Рис. 5. Внешний вид стартовой страницы AgentLink

3.3 Архивы

Все электронные документы, размещаются в EPrints в виде архива. Архив EPrints имеет свой конфигурируемый Web-сайт и данные.

Под *архивом (репозитарием)* в EPrints мы понимаем коллекцию электронных документов [7]. Архив содержит документы или объекты и записи метаданных, описывающие эти документы.

Архив может содержать разные типы документов, разных форматов и размеров. Все документы архива можно сгруппировать по определенным метаданным. Более точная настройка архива осуществляется при помощи редактирования конфигурационных файлов.

Создавать открытый научный архив можно двумя способами – когда этим занимается непосредственно редакция, выпуская электронные издания с открытым доступом по аналогии с печатными изданиями, или же когда сами авторы самостоятельно вносят свои статьи в архивы, этот процесс называется самоархивированием (self-archiving) [8]. Как показывают исследования [9-10], при самоархивировании уровень цитирования работ становится намного выше, чем у закрытых архивов.

3.4 Метаданные и типы документов

EPrints обрабатывает разные типы записей, описывающие документы. Каждый тип документов имеет свой набор полей метаданных (поднабор всех полей метаданных системы EPrints), который будет максимально полезен при описании.

При разработке основных типов документов предполагалось, что архив не будет включать учебные материалы, программное обеспечение и внутреннюю административную документацию (например, финансовые отчеты). Страницы введения метаданных формируются таким образом, что поля для заполнения максимально отвечают типу документа. В системе сделан акцент на интернациональность архива и его широкую доступность.

В системе EPrints выделены следующие типы документов:

- *статья*, подготовленная для печати в журнале, информационном бюллетене, газете, которая будет доступна через Интернет или бумажный носитель, причем она может быть и не рецензированной;
- *книга* или публикуемый том конференции;
- *раздел книги*, глава или отдельная часть в опубликованных трудах (например, конференции);

- *монография*, что включает технический отчет, проектный отчет, документацию и руководства, рабочие бумаги и дискуссионные материалы;
- *статья для конференции или семинара* – статьи, лекции, плакаты, презентации, представленные на конференции или семинаре; если они были опубликованы, то следует использовать «Раздел книги» или «Статья»;
- *диссертация* – автореферат или диссертация;
- *патент* – опубликованный патент (не включает неопубликованных патентов);
- *другой* – издание, которое принадлежит архиву в плане его направленности, но не может быть отнесено ни к одной из предыдущих категорий.

Для полноты архива некоторые поля метаданных являются обязательными для заполнения. Кроме того, каждое поле имеет детальную справку в зависимости от типа документа.

3.5 Поиск

Сервис поиска EPrints индексирует все имеющиеся файлы во всех архивах, что позволяет реализовать базовый (simple search) и расширенный (advanced search) поиск (Рис. 6).

Ukraine Eprints

[Home](#) [About](#) [Browse](#) [Search](#) [Register](#) [User Area](#) [Help](#)

Welcome to Ukraine Eprints

Welcome to this generic eprint server, running the default configuration for [GNU EPrints repository software](#).

[More information](#) is available about the repository.

Search items:

Search full text: [Search](#)

[Browse](#)

Browse the repository by [Subject](#) or [Year](#).

[Latest Additions](#)

View items added to the repository in the past week.

[Simple Search](#)

Search the repository using the most common fields.

[Advanced Search](#)

Search the repository using a more complex range of fields.

[Registered Users Area](#)

Registered users may submit items to the repository.

[Register](#)

Register for an account (or reset a forgotten password).

Ukraine Eprints is running on *GNU EPrints* repository-creating software, which generates eprints repositories that are compliant with the [Open Archives Protocol for Metadata Harvesting OAI 1.1 and 2.0](#).

The GNU EPrints repository-creating software is available for free at <http://software.eprints.org/>.



Рис. 6. Стартовая страница EPrints

Базовый поиск дает возможность выполнять запрос по основным группам метаданных таким как «автор/описание/создатель/дата», «текст/название/описание/автор/создатель/дата», «автор /создатель/редактор» и «дата». Причем вывод результата поиска можно отсортировать по «году издания», «автору» или «названию». На Рис. 7-8 приведен пример базового поиска и результата его выполнения.

Open archive
Home
About
Browse
Search
Register
User Area
Help

Simple Search

[Click here for an advanced search](#)

Title/Abstract/Authors/Creators/Date
Enter a term or terms to search for.

Full Text/Title/Abstract/Authors/Creators/Date
Enter a term or terms to search for.

Authors/Creators/Editors
Enter a name or names to search for. Either the family name, or the family name followed by a comma and the first name or initial. Names with spaces may be indicated by surrounding them with double quotes, eg. "van Damme, J".

Date
Enter a date or date range.
Examples: "1985-", "2001-05-17-2002-05-16", "-11-1980"

Retrieved records must fulfill of these conditions.

Order the results:

[Contact Information](#)

Рис. 7. Базовый поиск в EPrints

Open archive
Home
About
Browse
Search
Register
User Area
Help

Results for Simple Search

Title/Abstract/Authors/Creators/Date matches all of "Semantic". Results ordered by year (most recent first).
Displaying results **1 to 4 of 4**. Search time: 1s.
(index last built at: Пнд Окт 30 14:09:39 EET 2006)

Srinivasan, Naveen
[Semantic Web Service Discovery in the OWL-S IDE](#)
Article. Item availability restricted. [Deposited 13 October 2006]



Yonglei, Yao
[Service Matching Based on Semantic Descriptions](#)
Article. [Deposited 13 October 2006]



Марчук, А.Г.
[Semantic Web – на пути к новому поколению информационных систем](#)
05 March 2005, Article. Item availability restricted. [Deposited 09 October 2006]



Zhang, Ming
[DLonto: A Semantic Information Model for Digital Libraries](#)
Article. Item availability restricted. [Deposited 13 October 2006]



[Contact Information](#)

Рис. 8. Результат базового поиска по запросу «Semantic»

Расширенный поиск в отличие от базового дает возможность выполнить запрос по отдельным данным, например, «полнотекстовый поиск», «название», «автор», «резюме», «ключевые слова», «предметный классификатор», «тип публикации», «место издания», «редактор», «статус публикации» и «тип публикации». И как в предыдущем случае, вывод результата поиска можно отсортировать по «году издания», «автору» или «названию». На Рис. 9-10 приведен пример расширенного поиска и результата его выполнения.

Open archive [Home](#) [About](#) [Browse](#) [Search](#) [Register](#) [User Area](#) [Help](#)

Advanced Search

Don't panic! Just leave the fields you don't want to search blank. [Click here for a simple search.](#)

Full Text
Enter a term or terms to search for.

Department
Enter a term or terms to search for.

Editors
Enter a date or date range.
Examples: "1985", "2001-05-17 2002-05-16", "-11-1980"

Retrieved records must fulfill of these conditions.

Order the results:

[Contact Information](#)

Рис. 9. Расширенный поиск в EPrints

Open archive [Home](#) [About](#) [Browse](#) [Search](#) [Register](#) [User Area](#) [Help](#)

Results for Advanced Search

Department matches all of "кафедра квантовой электроники". Results ordered by year (most recent first).
Displaying results **1 to 1 of 1**. Search time: 1s.
[Index last built at: Пнд Oct 30 14:09:39 EET 2006]

ЧУЧМАН, М.П.
Эмиссионные характеристики и физика процессов в лазерной плазме галлия, индия, свинца, сурьмы и соединений на их основе.
14 October 2005, Thesis. Item availability restricted. [Deposited 09 October 2006]

[Contact Information](#)

Рис. 10. Результат расширенного поиска по запросу «кафедра квантовой электроники» в поле «отдел»

3.6 Просмотр

Некоторые пользователи предпочитают просмотр (навигацию) поиску. Страницы просмотра создаются сценарием автоматически. Эти страницы статические, т.е. генерируются не динамически из базы данных, а статически - скриптом, который запускается через равные промежутки времени. По умолчанию поля, по которым можно выполнять просмотр это - год и предметный классификатор (subject).

Предметный классификатор рассматривается как частный случай тезауруса информационных систем, т.е. он описывает логику отношений между предметными областями. Каждая единица хранения имеет свой статический URL, по которому отображаются расширенные метаданные, то есть дополнительные данные, которые не отображаются при навигации или при просмотре результатов поиска (например, аннотация).

Поля отображаемых метаданных могут быть отредактированы. Выполнять просмотр можно не только по определенным полям (год, предметная область), но и по полям более сложного, вложенного типа, например в выбранной предметной области можно выполнить просмотр по годам.

Кроме того, ПО EPrints предоставляет возможность просмотра новых поступлений за последнюю неделю.

3.7 Международная поддержка

EPrints использует несколько технологий для интернационализации архива документов, а именно, для сохранения метаданных используется кодировка Unicode, обеспечивающая поддержку различных языков [5].

Следует отметить, что корректное отображение кириллицы нам удалось получить только при использовании определенных версий сопутствующего ПО (см. вводную часть Раздела 3).

Имеется также возможность, которая была использована и в нашем проекте, создания многоязычного интерфейса. Благодаря тому, что интерфейс формируется на XML файлах, легко осуществлять локализацию. Чтобы получить нужный язык интерфейса, достаточно сделать перевод отдельных конфигурационных файлов и продублировать их для каждого языка.

4 Ввод записей в систему EPrints

Процесс наполнения библиотеки проходит в несколько этапов. Все операции осуществляются через Web-интерфейс. Вносить электронные документы может только зарегистрированный пользователь.

4.1 Регистрация

Пользователем заполняется соответствующая Web-форма, где указывается имя пользователя (username), пароль (password) и его электронный адрес (email). По данному адресу высылается активационный код. И если адрес верный, то только в этом случае создается новая учетная запись (account).

4.2 Информация пользователя

После регистрации пользователь заполняет форму, где указывает свое имя, организацию и т.п. Поля, которые нужно заполнить, могут дополнительно настраиваться. Редактировать записи пользователей может администратор библиотеки.

4.3 Страница пользователя

EPrints выделяет 3 группы пользователей: обыкновенные пользователи, редакторы и администраторы. В зависимости от типа зарегистрированного пользователя данная страница будет иметь разный вид, соответственно предоставляя меньше или больше сервисов.

Для доступа к странице пользователя нужно ввести имя (username) и пароль (password).

При помощи Web-интерфейса администратор может выполнять следующие действия: подписка на последние введенные записи; установка пароля доступа; ввод записей; изменения адреса электронной почты; просмотр статуса архива; редактирование архива; внутрисистемный поиск; редактирование предметного классификатора; редактирование записей пользователей.

4.4 Ввод записи

На своей странице пользователь может вводить запись в архив EPrints. При создании новой записи пользователю нужно указать тип документа. Это необходимо для дальнейшего формирования метаданных. Как только соответствующие метаданные будут заполнены, запись перемещается в редакционный буфер, где он пребывает до того времени, пока его обработает редактор.

Особенностью ввода записи является постраничное заполнение метаданных, объединенных в соответствии с их логической структурой.

В EPrints определены форматы документов, которые желательно принять по умолчанию: PDF, ASCII, Microsoft Word и HTML. Это связано с тем, что эти документы могут быть проиндексированы поисковыми машинами (например, Google), кроме того, есть возможность вносить изображения и в случае, если ни один из форматов не подошел, можно выбрать тип «другой», с тем чтобы пользователь смог самостоятельно определить формат вносимого документа.

В тех случаях, когда набор полей метаданных не удовлетворяет пользователя, есть возможность самостоятельно определить новые метаданные для конкретного типа входных документов. Некоторым полям метаданных для упрощения и правильности занесения можно присваивать значения по умолчанию, эти изменения требуют прав администратора.

При вводе конкретной записи предусмотрена возможность ее приписывания к нескольким предметным классификаторам. Для упрощения навигации по дереву предметного классификатора, которое в отдельных случаях может быть очень разветвленным, используется быстрый поиск (Рис. 11).

Subjects

Current item: **Спірін, О.М. (2000) *Практична інформатика*. методичний посібник. Житомирський державний університет ім. Івана Франка, Україна. ISBN 966760330X**

Please enter the information about this item. Fields marked with a * are fields that must be filled out before your deposit will be accepted.

[< Previous](#)[Save for Later](#)[Next >](#)

Subjects *

Please select at least one main subject category, and optionally up to two other subject categories you think are appropriate for your submission, from the list below.

Selections: none

Quick Search:

[Go!](#)

- [A General Works...](#)
- [B Philosophy, Psychology, Religion...](#)
- [C Auxiliary Sciences of History...](#)
- [D History General and Old World...](#)
- [E History America](#) [\[add\]](#)
- [F History United States, Canada, Latin America...](#)
- [G Geography, Anthropology, Recreation...](#)
- [H Social Sciences](#)

Рис. 11. Навигация по предметному классификатору с поиском

Рассмотрим типичный набор метаданных, например для книги:

- НАЗВАНИЕ КНИГИ (обязательное поле);
- АВТОРЫ/СОЗДАТЕЛИ (обязательное поле);
- РЕДАКТОРЫ (обязательное поле);
- ЗАГОЛОВОК ЖУРНАЛА/ПУБЛИКАЦИИ (обязательное поле);
- ТОМ - номер тома журнала, где опубликована статья;
- НОМЕР выпуска журнала;
- ДИАПАЗОН СТРАНИЦ;
- ДАТА ВЫПУСКА - когда статья была выпущена или издана (обязательное поле);
- ДАТА ПРЕДСТАВЛЕНИЯ - когда статья была представлена издателю (обязательное поле);
- ИДЕНТИФИКАТОР - уникальный номер-идентификатор или DOI (The Digital Object Identifier);
- ОФИЦИАЛЬНЫЙ URL;
- ISSN - интернациональный стандартный серийный номер;
- СОСТОЯНИЕ (обязательное поле):
 - опубликована;
 - в печати;
 - представлена на рассмотрение;
 - не опубликована.
- НУЖДАЕТСЯ ЛИ ЭТА РОБОТА В РЕЦЕНЗЕНТЕ:
 - да;
 - нет.
- РЕЗЮМЕ;
- СПИСОК ЛИТЕРАТУРЫ;
- КЛЮЧЕВЫЕ СЛОВА;
- ПРЕДМЕТНЫЙ КЛАССИФИКАТОР (обязательное поле);
- ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ - информация, которая может быть полезной о вашем депозите, которая не может быть введена где-нибудь еще. Эта информация появится на итоговой странице.
- КОММЕНТАРИИ И ПРЕДЛОЖЕНИЯ - любые комментарии редактору. Эта информация не будет отображена публице.

В случае если заполнение полей метаданных выполнено не до конца, возможно автоматическое сохранение результатов для дальнейшей работы по вводу записи.

Также предусмотрена проверка полей, обязательных к заполнению, что очень удобно при вводе записи.

Процесс заполнения метаданных осуществляется за несколько шагов. На первом заполняется общая информация: название записи, автор. На втором - место и дата публикации, а также серия издания (тут мы выделяем методическое пособие, учебное пособие, а также научная серия). На третьем шаге указываем состояние книги относительно публикации и необходимость в рецензенте. На следующей странице вводим аннотацию и ключевые слова. Вследствие особенности университета и электронной библиотеки мы расширили предметный классификатор в соответствии с нашими потребностями. Например, ввели «История Украины» в предметный классификатор «History General and Old World» или «Методика преподавания» в соответствующие разделы предметных классификаторов (Рис. 12).

Location(s) in Hierarchy

- (Top Level)
 - Library of Congress Subject Areas
 - Q Science
 - QA Mathematics
 - QA75 Electronic computers. Computer science
 - **Technique of computer science**

Рис. 12. Пример структуры предметного классификатора

Наконец, последним этапом ввода записи является загрузка самого файла. При этом можно указать, будут ли иметь доступ к файлу все пользователи или только зарегистрированные. После всех перечисленных выше операций запись перемещается в редакционный буфер.

В случае если вносится группа записей, которые имеют некоторые общие метаданные, EPrints предоставляет возможность использовать любые записи в качестве шаблонов, что существенно экономит время при наполнении архива.

EPrints позволяет строить дерево версий одного и того же документа в хронологическом порядке. При этом если мы для данной записи создадим новую версию с отличным от оригинала типом записи (например, новая версия для книги (book) будет частью книги (Book Section)), то соответствующие метаданные будут автоматически присвоены новой версии записи, но с возможностью редактирования.

4.5 Санкционирование записи

Пользователь, работающий с правами редактора, может просматривать редакционный буфер, при необходимости вносить туда изменения, а также может возвращать запись на доработку, высылая при этом по электронной почте замечания по внесенной записи. Только после одобрения редактором будет возможен доступ к записи широкому кругу пользователей библиотеки. В учреждениях, где скорость обновления существенна, предусмотрена возможность отключения редакционного буфера для непосредственного ввода записи в основной архив.

Поскольку все странички библиотеки статические, необходим некоторый промежуток времени пока их сгенерирует соответствующий скрипт. В нашем случае, поскольку динамика наполнения небольшая, генерация страниц выполнялась через 24 часа, что пока удовлетворяет потребности.

4.6 Подписка

Зарегистрированный пользователь может подписаться на новые поступления по интересующим его вопросам и получать сообщения раз в день, неделю или месяц. Для реализации данной возможности на сервере должен работать почтовый домен.

5 Настройка конфигурационных файлов

После создания нового архива можно задать дополнительные настройки, относящиеся к многоязычной поддержке, метаданным, интерфейсу электронной библиотеки. Все конфигурационные файлы имеют детальные описания. Эти файлы содержат шаблон Web-сайта, статические страницы и др.

При настройке каждого архива используется несколько основных конфигурационных файлов. Перечислим их основное назначение.

ArchiveConfig.pm – общая настройка системы, формирование поисковой формы, навигация, права пользователей, настройка поискового кеша.

ArchiveMetadataFieldsConfig.pm – один из главных конфигурационных файлов, отвечает за метаданные. С его помощью можно установить новые типы метаданных, их значение по умолчанию, а также обязательность поля для заполнения.

ArchiveOAIConfig.pm отвечает за работу с протоколом OAI, здесь вносится идентификатор архива и метаданные, которые архив будет предоставлять, по умолчанию метаданные отвечают стандарту Дублинского ядра.

ArchiveTextIndexingConfig.pm – здесь устанавливаются опции индексирования, например, возможность задать список слов, которые будут игнорироваться при обработке файла.

ArchiveValidateConfig.pm – настройка проверки данных внесенных пользователем [5].

Файлы citations-languageid.xml, phrases-languageid.xml и template-languageid.xml отвечают за интерфейс системы, и при локализации переводить следует именно их.

6 Пример архива EPrints

EPrints был выбран как прототип электронной библиотеки для Житомирского государственного университета им. Ивана Франка. Доступ к

данному серверу возможен по адресу <http://EPrints.zu.edu.ua/>. Во время написания этой статьи библиотека содержала небольшое число документов: учебно-методические материалы, научные работы студентов и преподавателей университета. По каждой записи внесены обязательные метаданные, например, ключевые слова, что позволяет более просто ориентироваться в архиве даже неопытному пользователю.

Также частично произведена локализация интерфейса с поддержкой украинского языка.

Позитивными моментами при использовании EPrints можно назвать: свободный доступ студентов к информационным ресурсам; поддержка протокола OAI позволяет получать более глубокие и актуальные знания во всех областях науки, в будущем возможен доступ к информационным ресурсам других университетов; самоархивирование научных работ повысит общий уровень исследований проводимых в университете.

Единственный острый вопрос, который остается не разрешимым, это наполнение библиотеки. Идеальным вариантом было бы решение об обязательном внесении трудов преподавателей университета в архив. Поскольку многие исследования выполняются за государственные средства, это было бы логичным решением.

Заключение

EPrints это удобное и легко настраиваемое программное обеспечение, которое имеет все качества, присущие современным программным продуктам в области создания электронных библиотек. Лидерство EPrints при создании научных и учебных архивов – тому доказательство.

При объединении СУБД MySQL, Perl и самого стабильного на сегодня веб-сервера Apache можно говорить о высокой надежности и хорошем быстродействии EPrints. Кроме того, EPrints прекрасно зарекомендовал себя как мощное средство создания и управления архивами.

В будущем разработчики EPrints обещают улучшенную поддержку технологии VLit [5] (<http://xanadu.net/>), которая обеспечит непосредственное цитирование документа, что позволит вставлять в основной документ цитаты из первоисточника, не загружая весь документ.

Разработчики ПО EPrints предполагают в самом близком будущем реализовать поддержку операционной системы Windows, а также расширить локализацию, в том числе и русского языка.

Литература

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. К. Лагозе, Г. Ван де Зомпель. Инициатива «Открытые архивы»: создание среды с высокой степенью интероперабельности. Электронные библиотеки. 2001. - 4. Вып. 6. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2001/part6/LS>
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. Gutteridge C., Hitchcock S., Simpson P., Hey J. Report on the technical issues of using GNU EPrints software for the development of an institutional e-Print repository at the University of Southampton: TARDIS deliverable D.2.3.2. 2003. <http://tardis.EPrints.org/>
5. Gutteridge C. EPrints 2.3 Documentation. October 12, 2005. <http://www.EPrints.org/documentation/tech/EPrints-docs.pdf>
6. A Guide to Institutional Repository Software, 3rd Edition, Open Society Institute, August 2004.
7. EPrints Self-Archiving FAQ, <http://www.EPrints.org/openaccess/self-faq/>
8. EPrints Open Access. <http://www.EPrints.org/openaccess/>
9. Tim Brody, Heinrich Stamerjohanns, Francois Vallieres, Stevan Harnad, Yves Gingras, Charles Oppenheim. The Effect of Open Access on Citation Impact. <http://www.ecs.soton.ac.uk/~harnad/Temp/OATAnew.pdf>
10. Sale A. Eprint website for the University of Tasmania. August 2004. <http://EPrints.comp.utas.edu.au:81/archive/00000011/>

Об авторах

Резниченко Валерий Анатольевич - старший научный сотрудник Института программных систем НАН Украины, г.Киев,
E-mail: reznich@isofts.kiev.ua

Проскудина Галина Юрьевна - научный сотрудник Института программных систем НАН Украины, г. Киев,
E-mail: gupros@isofts.kiev.ua

Новицкий Александр Вадимович - аспирант Института программных систем НАН Украины, г. Киев,
E-mail: alex@zu.edu.ua