

Подход к построению научной электронной библиотеки СО РАН

© Леонова Ю.В.

Институт вычислительных технологий СО РАН
julil@ict.nsc.ru

Аннотация

Работа посвящена задаче создания концептуальной модели электронной библиотеки для формирования информационных ресурсов и организации совместной работы, включая средства управления Web-контентом. Представлены практические результаты реализации предложенных технологических решений.

1 Введение

В настоящее время становление современного информационного общества немисливо без использования информационных ресурсов в электронном виде. Организация доступа к источникам информации в электронной форме стала одной из важнейших задач информационного обслуживания науки и образования. Создание электронных библиотек представляет собой качественно иной уровень производства, хранения, организации и распространения самой разнообразной информации. Реализация технологических возможностей создания электронных библиотек предоставляет широкие возможности для управления большими объемами данных и их обработки.

Понятие электронной библиотеки сейчас неотделимо от Интернета и поисковых систем. Именно развитие этих двух информационных технологий открыло новые возможности и породило новые проблемы. Однако наряду с ценной информацией в Интернете накопилось огромное количество ненужной и неактуальной. При любом поиске вываливается огромное количество лишней информации, в основном рекламного характера, приходится просматривать много ссылок, чтобы извлечь нужное. При этом не всегда удается найти серьёзные и содержательные источники. Электронная библиотека представляет собой информационное подпространство Интернета, очищенное от ненужной информации и наполненное содержательными материалами.

В данной статье рассматриваются принципы построения научной электронной библиотеки СО РАН.

2 Система управления контентом

Для более эффективного управления, создания и представления контента в сети библиотеки используют системы управления контентом. Одной из таких систем управления контентом является Drupal (drupal.org). Преимущество Drupal - использование восходящей стратегии для веб-дизайна, отделяющей контент сайта от форматирования. Drupal применяется во многих библиотеках поскольку является бесплатным программным обеспечением с открытым исходным кодом и позволяет пользователям легко вносить контент на Веб-сайт библиотеки.

Традиционно организации имеют Веб-мастера, отвечающего за форматирование и наполнение контента на Веб-сайте. Drupal присваивает полномочия различным пользователям, основанные на ролях (администратор или другой пользователь), позволяя пользователям самостоятельно редактировать Веб-сайт без обращения к Веб-мастеру. В зависимости от присвоенной роли пользователи могут редактировать или просмотреть Веб-сайты, блоги, корпоративные сайты со служебной информацией о сотрудниках.

Drupal предоставляет возможности быстрого создания контента, однако для полноценного использования требуется настройка дополнительных модулей-плагинов, расширяющих функциональность Drupal, как добавление блогов, книг, изображений и интерактивных опросов. Базовая функциональность Drupal также позволяет библиотекам изменять стиль сайта и контент, отображаемый конечному пользователю.

Функциональность Drupal организовывается посредством модулей. В данной статье будут рассмотрены:

- модули ядра, включенные в базовый дистрибутив
- основные дополнительные модули, которые могут использоваться для создания сервисов для библиотек или архивов,
- модули, которые были созданы специально для обеспечения поддержки функций библиотек, архивов или других сервисов информации

Очевидно, модули ядра могут находить применение в любой библиотеке, так как они обеспечивают основную функциональность для разработки проектов управления содержанием: узлы, контроль пользователей, статистика и т.д.

Тем не менее, среди них необходимо отметить модуль Taxonomy, т.к. он позволяет создать схемы классификации, используемые тезаурусом. Этот модуль также может использоваться в сочетании с другими модулями для создания новых схем классификации, создания облака тегов или в сочетании с классификационными схемами, взятыми из других источников.

Основные дополнительные модули не включены в ядро Drupal, однако они существенно расширяют возможности представления информации. Имеется много модулей, которые могут быть использованы для реализации полезных функций, но среди них следует отметить следующие:

- *Content Construction Kit*, ССК: фундаментальный модуль, позволяет создавать структуры данных для новых типов узлов, разработку и интеграцию любого типа информации. В свою очередь, разработаны расширения для ССК, что повышает его функциональность и производительность.
- *Views*: является дополнением к ССК, так как он используется для создания визуализации узлов или комбинаций узлов, используя в качестве отправной точки поиска выражения или набор узлов в соответствии с заданными условиями.
- *Apache Solr*: модуль делает возможным использование поисковой системы Solr Apache Foundation, реализованной на Java. Это позволяет, например, выполнять фасетный поиск, что увеличивает использование ее в качестве инструмента для поиска в библиотечных каталогах и аналогичных ресурсах.

Drupal – это не только платформа, но и активное сообщество пользователей. Подобно многим проектам ПО с открытым кодом, Drupal имеет крупное сообщество пользователей, которое непрерывно разрабатывает модули и улучшает функциональность программного обеспечения.

Группа Drupal для Библиотек (<http://groups.drupal.org/libraries>) содержит как различную информацию о модулях для библиотек, их обновлениях и различных представлениях. Связали модули и предоставляют информацию и обновления для различных. Сайт содержит каталог общедоступных, академических, школьных, и специальных библиотек, которые используют Drupal. Сообщество Drupal уже создало несколько специфичных для библиотеки приложений (API):

- *Faceted Search*: продвинутый поиск на основе фасетного метода. Фасетный метод классификации подразумевает параллельное разделение множества объектов на независимые классификационные группировки. При этом не предполагается жесткой классификационной структуры и заранее построенных конечных группировок. Классификационные группировки образуются путем комбинации значений, взятых из соответствующих фасетов. Основным преимуществом классификации с использованием фасетного метода является гибкость структуры ее построения. При фасетной классификации появляется возможность агрегации объектов и осуществления информационного поиска по любому сочетанию фасетов.
- *Biblio*: позволяет создавать сервер библиографии, поддерживающий импорт/экспорт данных в различных общих форматах.
- *BiblioFacets*: Позволяет интегрировать Faceted Search с модулем Biblio.
- *Book Post*: позволяет объединять информацию о книгах посредством ISBN, используя данные Open Library Project – открытый каталог, позволяющий пользователям добавлять информацию о книгах, или использовать ссылки на каталоги WorldCat, LibraryThing, Google Books.
- *Library*: версия простой системы управления библиотекой. Позволяет пользователям управлять наличием материалов в качестве активов.
- *MARC*: выполняет импорт данных в формате MARC.
- *HILCC Taxonomy Autotag* (Hierarchical Interface to the Library of Congress Classification): анализирует материалы для поиска полей с идентификаторами библиотеки конгресса США (например, QA 76.73 .P224 2005) и загружает соответствующую иерархическую классификацию с сервера библиотеки. Полученная классификация может быть использована для поиска или для иерархической организации контента.
- *Z39.50*: выполняет поиск на серверах Z39.50 и объединяет результаты.

- *Millenium Integration*: модуль импорта библиографической информации из Millennium WebOpac (библиотечного каталога), генерирует материалы типа Biblio. Позволяет интегрировать информацию с нескольких Opac-серверов. Обновляет автоматически реестры, позволяет выполнять поиск информации в Google Books и импортировать индексы Библиотеки конгресса США, строить таксономии и т.д.
- *OAI2 for CCK*: обеспечивает внедрение Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) в Drupal с поддержкой CCK типов и их полей.
- *OAI-PMH*: Интерфейс OAI-PMH для модуля biblio.
- *EZProхy*: представляет собой шлюз Drupal и EZProхy. EZProхy обеспечивает пользователям удаленный доступ к лицензируемому контенту, предоставляемому библиотеками.
- *SOPAC*: обеспечивает интеграцию с SOPAC (Social Online Public Access Catalog) системой каталогов сайтов.

Таким образом, Drupal предлагает всю необходимую функциональность для управления цифровой информацией. Комбинация специализированных библиотечных модулей с другими модулями позволяет поддерживать разнообразные сервисы для работы с информацией в Drupal. Примерами этого - является интеграция в Drupal контента стороннего ПО, как например каталоги библиотек.

В качестве недостатков можно указывать на отсутствие у специфических модулей достаточных возможностей для управления документами для интегрированной информации (что могло бы реализовываться посредством комбинации и конфигурации других модулей) и отсутствие механизма индексирования, способного работать непосредственно на присоединенных к контенту файлов, хотя интеграция третьих программных средств делает это возможным.

При выборе систем управления контентом библиотеки должны определить свои требования. Например, есть существенное перекрытие в функциональности между Drupal и Joomla – другой системой управления контентом с открытым исходным кодом. Много библиотек выбрали Joomla из-за простоты использования и более развитого административного интерфейса. Однако сторонники Drupal утверждают, что после инсталляции Drupal обеспечивает больше функциональности и возможностей для разработки. Последний выпуск программного обеспечения - Drupal 6 и Drupal 7

содержит большое число улучшений, включая большее количество базовых модулей, улучшенную безопасность и улучшенный административный интерфейс.

3. Опыт создания «Электронной библиотеки по моделям динамики изменения биосферы» на основе средств CMS Drupal

В данной работе мы рассмотрим текущее состояние проекта – Электронной библиотеки по моделям динамики изменения биосферы (рис.1), проанализируем модули и самые важные разделы проекта и преимущества, предоставляемые Drupal. Проект «Электронная библиотека по моделям динамики изменения биосферы» был создан в 2009 году на основе ПО для электронной библиотеки СО РАН. Однако движок данного ПО не предоставляет всей необходимой функциональности и не обладает гибкостью. Для обеспечения необходимой функциональности был выполнен перенос ЭБ на платформу CMS Drupal.

3.1 Панель Администратора

После входа в систему через пользователя администратор (далее объяснятся различные типы пользователей), мы встречаемся с “Панелью Администратора”, которая предоставляет 2 способа отображения содержания:

- По задачам
Посредством этой формы визуализации инструментарий систематизируется по 5 фундаментальным блокам: Управление содержанием, Строительство сайта, Конфигурация сайта, Управление пользователями и Отчеты.
- По модулям

Посредством этой формы визуализации можно видеть все модули, установленные на сайте, и управлять всей информацией, относящейся к ним, как например конфигурация прав доступа к каждому модулю.

3.2 Управление содержанием

В этом разделе можно управлять содержанием сайта. По степени важности рассмотрим следующие подразделы и модули:

- Категории (модуль Тахоному)
Посредством категорий систематизируется содержание электронной библиотеки, отображаемое на веб-страницах. Главная особенность Drupal – реализация уникального механизма таксономии, то есть иерархических связей. В большинстве CMS

первоначально определяется структура сайта через меню, а затем документы прикрепляются к разделам. В Drupal же все наоборот. С помощью таксономии можно определить произвольное число рубрик, в которых будут в дальнейшем помещаться материалы сайта. Эти рубрики могут быть представлены как плоские списки или иерархические структуры произвольной вложенности (древовидные, произвольные).

- **Контент (модуль Node)**
Позволяет просматривать и редактировать информацию, отображаемую внутри веб-страниц.
- **Метаданные (модуль Meta tags)** – важный модуль, так как позволяет установить метатеги отдельно для каждого документа, также позволяет определить индексировать или не индексировать сайт роботами web. В этом модуле можно выделить общие ключевые слова и ключевые слова для каждой категории.
- **Библиография (модуль Bibliography)**
Этот модуль позволяет пользователям управлять и просматривать списки публикаций.

Рассмотрим подробнее данный модуль.

3.3 Библиография

Модуль библиографии (Biblio) обеспечивает

- Импорт форматов: BibTex, RIS, MARC, EndNote, XML.
- Экспорт форматов: BibTex, EndNote, XML.
- Стили отображения: AMA, APA, Chicago, CSE, IEEE, MLA, Vancouver.
- Использование ссылок цитирования литературы.
- Интеграция с таксономией.
- Поддержка различных языков

Модуль библиографии позволяет создавать и поддерживать библиографические списки публикаций. По желанию можно хранить полный (HTML) текст публикации. Также могут храниться файлы в других форматах - PDF или Word.

Библиографические записи могут импортироваться/экспортироваться в формате EndNote (размеченный XML) и BibTex. Библиографические списки могут быть отформатированы с помощью различных стилей: American Psychological Association (APA), Council of Science Editors (CSE) and Institute of Electrical and Electronics Engineers (IEEE).

Типы публикаций

Модуль библиографии поставляется с большим количеством заранее определенных типов публикаций, таких как книги, журнальные статьи, доклады конференций, газетные статьи, патенты, отчеты и т.д. Для каждого вида определен набор обязательных и дополнительных полей и их представление конечному пользователю. Имеется возможность более точно настроить параметры, для того чтобы избежать излишней сложности в форме ввода публикаций, можно изменять поля для всех типов публикаций, как определенных по умолчанию, так и пользовательских.

Контроль доступа

Права доступа, связанные с модулем библиографии, определяются администратором: добавление/редактирование библиографических записей (своих и записей других пользователей); импорт записей из текстового файла (например, BibTex или EndNote); показывать ссылки для скачивания файлов, присоединенных к библиографической записи; экспорт (в Bibtex, RIS, Endnote и др.); показывать вкладки-фильтры для фильтрации библиографических списков, полученных при поиске; показать ссылки для сортировки (автор, заглавие, тип, год) списка библиографических записей; просмотр полного текста (HTML) записи.

Навигация

В меню имеется 3 подпункта для навигации по базе публикаций.

- **Авторы** – отображение всех авторов в базе данных. Этот список формируется автоматически из всех авторов, содержащихся в публикациях.
- **Ключевые слова** - работает аналогично с перечислением всех ключевых слов, которые были вставлены в публикации.
- **Импорт** позволяет автоматически заполнить данные публикации из текстового файла.

Добавление списков цитирования

Drupal позволяет добавлять ссылки на библиографические ресурсы в текст, используя синтаксис `<bib>citekey</bib>`

3.4 Управление пользователями

В рамках данного раздела можно управлять пользователями сайте, группами пользователей и их доступом к функциям сайта. В ЭБ можно выделить следующие важные подразделы и модули:

- **Контроль доступа (модуль user)**

Позволяет контролировать доступ к различным модулям веб-сайта в зависимости от различных ролей пользователей



"Электронная библиотека по моделям динамики изменения биосферы"



Интеграционный проект СО РАН № 50 (2009 - 2011 гг.)

Модели изменения биосферы на основе баланса углерода (по натурным и спутниковым данным и с учетом вклада бореальных экосистем)

Интеграционный проект СО РАН № 50.

Научные координаторы проекта:

Ваганов Е.А., академик, д.б.н.

Сибирский федеральный университет, г. Красноярск

Федотов А.М., член-корреспондент РАН, д.ф.-м.н.

Институт вычислительных технологий СО РАН, г. Новосибирск

Новосибирский государственный университет, г. Новосибирск

Список организаций-исполнителей проекта:

- Институт биофизики СО РАН
- Институт вычислительного моделирования Сибирского отделения РАН
- Институт вычислительных технологий СО РАН
- Институт географии СО РАН
- Институт леса СО РАН
- Институт мониторинга климатических и экологических систем СО РАН
- Институт почвоведения и агрохимии СО РАН
- Институт цитологии и генетики СО РАН
- Новосибирский государственный университет
- Сибирский федеральный университет
- Учреждение Российской академии наук. Институт угля и углейхимии Сибирского отделения РАН
- Центральный сибирский ботанический сад СО РАН

Необходимость разработки механизмов, обеспечивающих как функционирование общей информационно-аналитической рабочей среды, является приоритетным направлением для задач информационной поддержки научных исследований. Эти вопросы приобретают особую важность для исследований экологических систем, когда различные группы исследователей (вследствие особенностей решаемых проблем, а также природы вопросов, связанных с окружающей средой), разделенные географически, должны осуществлять совместную работу, обмен данными и знаниями и координировать свои действия с целью оптимизации использования информационно-вычислительных ресурсов, сервисов и приложений.

Тесное кооперирование информационных технологий и наук о Земле способствует пониманию как глобальных, так и региональных природных процессов, взаимодействия процессов, формирующих природную окружающую среду.

Накопленные в настоящее время массивы экспериментальных данных столь велики, а организация и динамика экосистем так сложна, что без современных информационных технологий, методов анализа данных и моделирования осмысление их займет слишком много времени. Создание распределенных информационно-аналитических сред для описания биоразнообразия и его пространственно-временной динамики, накопления в базах данных информации об экосистемах, моделирования и прогноза их динамики на основе аналитических методов приобретают особую актуальность.

Оценка темпов, масштабов и степени необратимости наблюдаемых изменений глобальных параметров биосферы представляет собой, без сомнения, одну из актуальнейших задач, стоящих перед современной наукой.

Большинство моделей системы "биосфера-климат" направлено на получение наиболее вероятного варианта будущей динамики. При этом в большинстве математических моделей климата и биосферы земной шар пространственно разделяется на множество ячеек, для которых составляются уравнения ее состояния. Однако на этих весьма сложных распределенных моделях лежит «проклятие размерности» - стремление к точности прогноза побуждает увеличивать число переменных и параметров модели и часто приводит к противоположному результату.

В рамках вышесказанного основной целью проекта является интеграция всех сведений о биосфере с целью оценки ее состояния и построения прогнозных моделей динамики изменения состояния биосферы.

На страницах предлагаемой библиотеки представлены основные материалы, документы, публикации и информационные ресурсы, связанные с реализацией проекта.

А.М.Федотов

- Модели изменения биосферы на основе баланса углерода (по натурным и спутниковым данным и с учетом вклада бореальных экосистем)
Ваганов Е.А., Федотов А.М.
Содержание интеграционного проекта СО РАН № 50
- Список организаций в ЭБ по моделям динамики биосферы
Для просмотра ресурсов организации перейти на ссылку
- Список авторов ЭБ по моделям динамики биосферы
Для просмотра документов автора перейти на ссылку
- Основные мероприятия, проводимые по проекту
Конференции, совещания, новости и объявления.
- Планы работ по проекту на 2009 год.
- Информационная система спутниковых данных (ИССД).
Информационная система на ГИС-сервере ИВТ СО РАН.

Рис.1. Главная страница электронной библиотеки

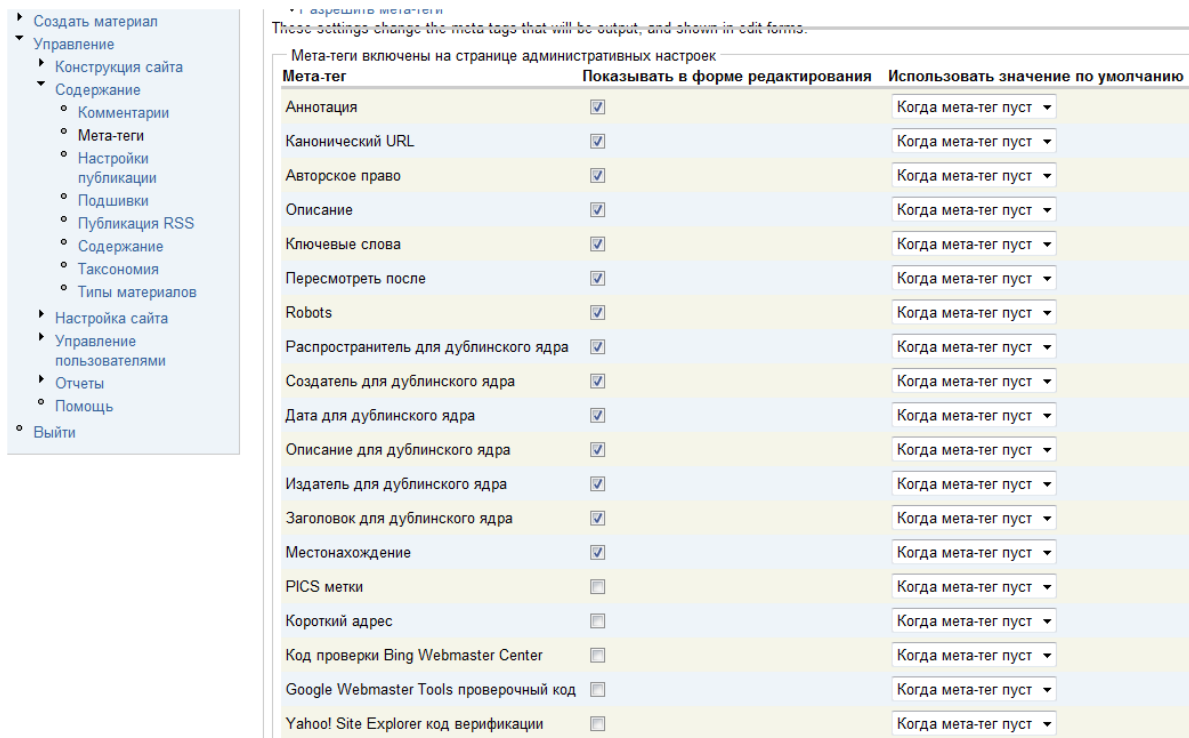


Рис.2. Модуль Meta tags. Настройка тегов для метаданных

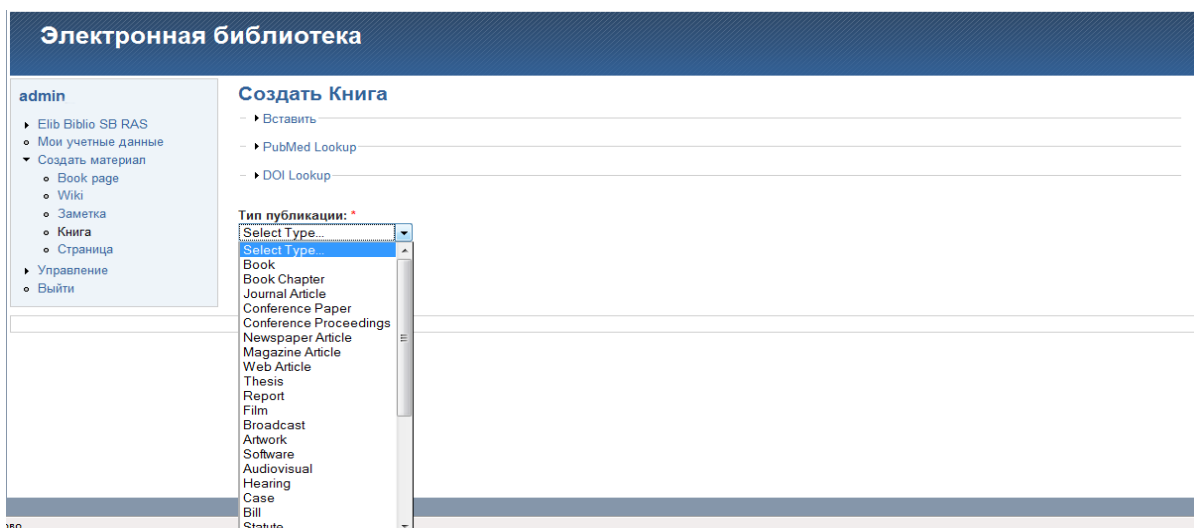


Рис.3. Добавление библиографической записи – выбор типа публикации

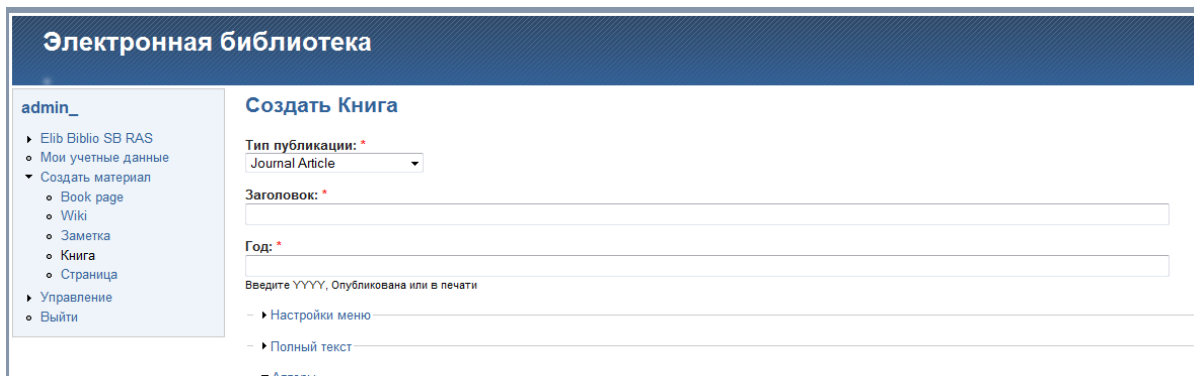


Рис.4. Добавление библиографической записи

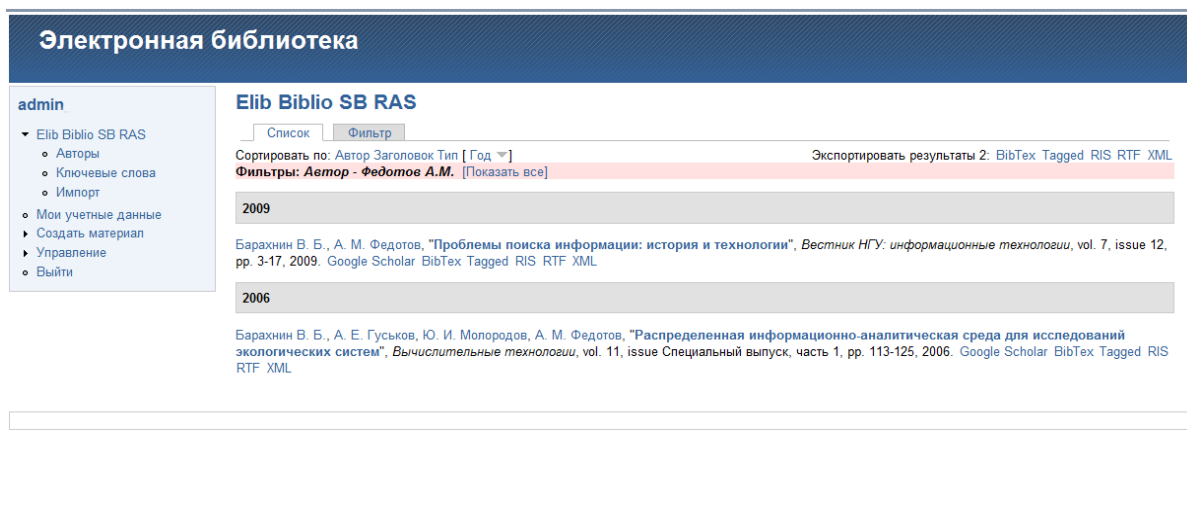


Рис.5. Навигация по библиографической базе

- Роли (модуль user)

Роли позволяют контролировать функции и виды доступа к различной функциональности и модулям, установленные на сайте. В ЭБ используется четыре типа ролей:

- Администратор
- Анонимный пользователь
- Зарегистрированный пользователь
- Привилегированный пользователь

- Пользователи (модуль user)

В этом разделе можно управлять пользователями и их функциями, ролями и получать информацию об их последних доступах к сайту.

Выводы

CMS Drupal предоставляет широкий набор функциональных возможностей, которые позволяют использовать его в качестве цифровой платформы для управления информацией в библиотеках и архивах. В данной работе рассматриваются его основные характеристики, модули и подчеркиваются основные информационные ресурсы для пользователей инструмента.

Литература

- [1] Drupal.org – Community plumbing.
<http://drupal.org>
- [2] Drupal России. <http://drupal.ru/>
- [3] VanDyk, J., Westgate, M. 2007. Pro Drupal Development. Berkeley: Apress
- [4] Стандарт СТ РК 5.0-95 Единая система классификации и кодирования технико-экономической и социальной информации. Основные положения.