

В.А. Бессонов, В.В. Ланин¹

Пермский государственный национальный исследовательский
университет,

Национальный исследовательский университет
«Высшая школа экономики» (Пермский филиал)

v.bessonov@hotmail.com, lanin@perm.ru

ПОДСИСТЕМА УПРАВЛЕНИЯ ДОКУМЕНТАМИ ПОРТАЛА, ПОСВЯЩЕННОГО МОДЕЛИРОВАНИЮ ИНФОРМАЦИОННЫХ СИСТЕМ

Введение

В рамках проекта разработки учебно-исследовательского портала по моделированию информационных систем планируется создание «саморазвивающегося» ресурса, предоставляющего в распоряжение пользователей средства интеллектуального поиска и автоматизированной обработки полученных результатов (документов, источников), удобные средства навигации по найденным ресурсам. Пользователи портала должны получить возможность настройки этих средств в соответствии со своими потребностями, а также возможность оперативного взаимодействия, обмена информацией и публикации результатов своих исследований для обсуждения, коллективной работы над проектами. Основными обрабатываемым на портале объектами будут являться электронные документы, поэтому подсистема управления документами является крайне важным компонентом портала. При проектировании портала было принято решение использовать для его создания одну из существующих ЕСМ-систем.

Современные ЕСМ системы

За последнее время электронные документы стали одним из основных средств обмена информацией. Всё больше организаций прини-

¹ Работа выполнена при поддержке Программы «Научный фонд НИУ ВШЭ» (проект № 12-09-0102)

© Бессонов В.А., Ланин В.В., 2012

мают решение об использовании специализированных средств для управления их информационными ресурсами. И все чаще в качестве таких специализированных средств выступают системы ECM (Enterprise Content Management). Система Enterprise Content Management (ECM) представляет собой набор технологий, которые предназначены для упрощения процессов сбора, хранения, и предоставления важной информации и документов, связанных с организационными процессами [4].

По данным агентства Gartner [1] рынок ECM-систем продолжает расти. Его суммарный доход увеличился на 11.1% по сравнению с 2011 г., в котором он составил \$4,3 млрд. Согласно данным агентства Forrester 71% организаций планируют увеличить число средств, затраченных на ECM-системы [3].

Кроме того, согласно Gartner [1], всё больше крупных организаций начинают переходить от использования ECM-систем в качестве хранилищ данных к более глубокой интеграции их в свои бизнес-процессы. Именно поэтому одним из основных направлений развития ECM-систем должно являться Enterprise Information Management (EIM) [4]. Использование EIM направлено, в первую очередь, на оптимальную организацию информационных ресурсов, включающих как структурированные, так и неструктурированные источники внутри организации. Внедрение ECM-системы предполагает наполнение репозитория документами, составление таксономии документов, разработку набора метаданных.

Для выбора ECM-системы, на базе которой будет построена подсистема разрабатываемого портала, рассмотрим конкретные системы, их характеристики и предоставляемые ими возможности.

Система Alfresco

Alfresco – это Open Source ECM-система одноимённой британской компании. Система Alfresco существует в трёх редакциях:

- Alfresco Community Edition является свободным программным обеспечением и распространяется на условиях LGPL.
- Alfresco Enterprise Edition является платной коммерческой версией продукта, которая распространяется под своей проприетарной лицензией, имеет открытый исходный код и соответствует открытым стандартам.
- Alfresco Cloud Edition – SaaS-версия Alfresco.

Согласно последнему опубликованному отчёту агентства Gartner [1] по ECM-системам в знаменитый «магический квадрант» попала единственная система Open Source из всех представленных –

система Alfresco. Gartner отметили инновационный подход [1] – именно поэтому Alfresco располагается в правом нижнем углу «магического квадранта». Также Alfresco Software попала и в отчет агентства Forrester. Согласно Forrester, Alfresco может составить реальную альтернативу таким крупным игрокам рынка ECM-систем, как IBM, Oracle, Open Text и EMC [2].

Пользователи могут начать работать с Alfresco в их повседневной деятельности сразу после установки, однако это не позволит использовать преимущества, предоставляемые данной системой. Одним из основных преимуществ является гибкая система управления контентом.

Модель содержимого в Alfresco включает в себя 5 основных составляющих [5]:

- 1) пользовательские типы (custom types);
- 2) свойства (properties);
- 3) ограничения (constraints);
- 4) ассоциации (association);
- 5) аспекты.

Пользовательские типы

Пользовательские типы предназначены для создания таксономии содержимого репозитория. С их помощью пользователь определяет типы документов, которые будут содержаться во внутреннем хранилище Alfresco. При этом пользовательский тип может обозначать не только конкретный тип документа, но и некоторую абстрактную категорию документов (например, «финансовый документ»), что позволяет гибко настраивать модель содержимого под нужды конкретной организации. Кроме того, для пользовательских типов определен механизм наследования, с помощью которого конкретные типы документов могут наследовать свои свойства от более абстрактного типа.

Свойства

Обычно для создания полноценной модели контента конкретной организации бывает недостаточно использования только одних пользовательских типов. Часто пользователю требуется сопоставить с каждым из пользовательских типов некоторый набор метаданных, качественно характеризующий данный конкретный тип. Для этого в Alfresco используются свойства. При определении нового свойства пользователь должен указать его имя и используемый тип данных (числовой, строковый, дата и/или время, булевский и т.д.).

Работа со свойствами в Alfresco напоминает работу с полями классов в объектно-ориентированном программировании (ООП). Однако в последнем случае пользователю необходимо программировать

всю бизнес-логику самостоятельно, в то время как Alfresco позволяет декларативно описывать ограничения, накладываемые на значения свойств.

Ограничения

Alfresco предоставляет возможность задания следующих ограничений:

- ограничение на минимальное/максимальное значение свойства (в случае использования числового типа данных);
- ограничение на длину значения свойства (в случае использования строкового типа данных);
- ограничение, контролирующее количество и вид элементов в случае использования типа данных «список»;
- ограничение, задаваемое с помощью регулярных выражений.

Alfresco предоставляет также возможности для переиспользования ограничений. Пользователь может определить набор ограничений и затем использовать их для различных свойств.

Ассоциации

Помимо определения таксономии документов организации и набора их метаданных, зачастую бывает необходимо определить связи между различными типами документов. Для этого в Alfresco используются ассоциации, которые можно разделить на два класса:

1. *Простые ассоциации*, с помощью которых пользователь указывает, что два конкретных пользовательских типа связаны между собой.

2. *Ассоциации, определяющие «родственные» связи* между двумя пользовательскими типами (child associations). Ассоциации данного типа налагают дополнительное ограничение, предполагающее, что связанный документ зависит от документа, с которым он связан. Например, при удалении родительского документа, дочерний также должен быть удален. Это напоминает правило каскадного удаления в реляционных СУБД. В Alfresco есть встроенная ассоциация данного типа – Contains, которая используется для определения связи между пользовательским типом, выступающим в качестве контейнера, например «Каталог», и его содержимым.

Аспекты

С помощью использования аспектов, пользователь может расширить набор свойств отдельных пользовательских типов. В [5] приводится пример, поясняющий необходимость применения аспектов. Зачастую возникает необходимость сделать доступными извне только

некоторые документы. Например, отдел кадров, хочет сделать доступными только объявления новых вакансий. Для этого необходимо определить набор дополнительных свойств, с помощью которых Alfresco могла бы определять, должен ли быть доступен данный документ извне, или нет. Существует два различных подхода, каким образом можно добавить данные свойства в модель:

1. Можно добавить данные свойства базовому пользовательскому типу. Однако в этом случае все дочерние типы также будут обладать этими свойствами, что приведет к дополнительным накладным расходам.

2. Можно добавлять эти свойства только к отдельным типам, которые должны быть помещены на портале. Однако в этом случае модель перестанет быть согласованной, так как, в случае если возникнет необходимость изменить или расширить данный набор свойств, это вновь потребует выполнения этой операции у всех типов снова.

Для решения этой проблемы можно использовать аспекты. Использование аспектов напоминает использование интерфейсов в объектно-ориентированном программировании. С их помощью пользователь может добавить ряд дополнительных свойств к отдельным типам, не меняя иерархию в целом.

Microsoft SharePoint 2010

Microsoft SharePoint – это инструмент для создания сайтов, предоставляющих пользователям возможность для совместной работы. Создаваемые на платформе SharePoint сайты могут быть использованы в качестве хранилища информации, знаний и документов, а также использоваться для исполнения облегчающих взаимодействие Web-приложений, таких как вики и блоги.

Модель данных SharePoint 2010 состоит из следующих понятий [6]:

- 1) списки (columns);
- 2) библиотеки (libraries);
- 3) столбцы (lists)
- 4) типы контента (content types).

Списки

Список представляет собой набор сведений, предназначенных для совместного использования различными пользователями. Содержимое списка определяется его типом, например:

- *Список объявлений* используется для публикации новостей и объявлений.

- *Список контактов* используется для хранения информации о контактных данных сотрудников и клиентов.
- *Список задач* используется для отслеживания текущих задач.
- и др.

Библиотеки

Библиотеки предназначены для хранения контента организации. Каждая библиотека отображает список документов и некоторый набор сведений о них, которые помогают пользователям в работе.

Пользователи могут управлять отображением документов, а также контролировать доступ к ним, ограничивая круг пользователей, которым разрешается просматривать документы до их утверждения. Кроме того, библиотеки являются средством контроля версий: с их помощью возможно одновременно работать с различными версиями одного и того же документа.

Содержимое библиотеки, также как и содержимое списка, определяется ее типом. Перечислим некоторые из них:

- *Библиотека документов* используется для хранения многих типов файлов, включая текстовые документы, электронные таблицы, презентации и т.п.
- *Библиотека рисунков* используется для предоставления общего доступа к коллекциям цифровых фотографий или рисунков.
- *Библиотека форм* используется для хранения форм, созданных с помощью Microsoft Office InfoPath.

Столбцы

Столбец представляет собой атрибут метаданных, используемый для описания отдельных элементов в списке или библиотеке.

Для столбца можно задать следующие свойства:

- *Имя* используется для идентификации столбца в списках и библиотеках.
- *Группа столбцов*, к которой принадлежит столбец, используется для организации столбцов по логическим категориям.
- *Тип данных* (числовой, валюта, дата и время и т.п.).

Типы контента

Типы контента используются для описания свойств элементов списков и библиотек и могут включать следующие элементы:

- метаданные или свойства, пользовательские формы, рабочие процессы, шаблоны документов которые необходимо присвоить этому типу, представленные столбцами, добавляемыми в

- список или библиотеку документов при добавлении типа контента;
- операции создания, редактирования и отображения, которые следует применять с этим типом контента;
- процессы, доступные для элементов этого типа контента (процессы могут запускаться пользователем или автоматически по наступлению выбранного события или условия);
- форматы документов (например, документы Microsoft Word, Excel и т.д.), ассоциирующиеся с данным типом.

Так как типы контента определяются независимо от конкретных списков или библиотек документов, пользователь может сделать определенный тип контента доступным для списков на нескольких Web-сайтах SharePoint. Это позволяет централизованно определять типы контента, хранящегося в семействе Web-сайтов, и управлять ими. Например, пользователь может определить тип контента «Юридический документ», чтобы гарантировать, что все юридические документы будут отслеживаться одни и те же метаданные, даже если эти документы создаются и хранятся на нескольких Web-сайтах.

С помощью типов контента пользователь может, например, определить таксономию документов в библиотеке документов. Например, описать три конкретных типа документов («отчет о расходах», «заказ на покупку» и «счет») и один абстрактный – «финансовый документ», который бы определял ряд общих характеристик, таких, например, как тип используемых данных.

Заключение

В статье как возможные альтернативы реализации подсистемы управления документами портала были рассмотрены наиболее популярные в нашей стране ECM-системы: Alfresco и Microsoft SharePoint. В результате проведенного сравнения принято решение использовать систему Alfresco. Немаловажными факторами при выборе системы стали открытость исходного кода и использование открытых стандартов. Кроме того, система Alfresco реализована на языке Java, что значительно упрощает интеграцию с компонентами обработки онтологий и инструментами Semantic Web, так же реализованными на данном языке. Система SharePoint, в свою очередь, ориентирована на корпоративный сегмент и решения Microsoft, обладает ограниченными возможностями интеграции семантических технологий.

Библиографический список

1. *Gilbert M.R., Shegda K.M., Chin K., Tay G., Koehler-Kruener H.* Gartner's Magic Quadrant for Enterprise Content Management. 18 October, 2012.
2. *Weintraub A.* The Forrester Wave™: Enterprise Content Management, Q4 2011. – November 1, 2011.
3. ECM: Forrester Predicts 2012 to be a Break-Out Year for Content [Электронный ресурс]. URL: <http://www.forrmttek.com/blog/?p=2845> (дата обращения: 11.12.2012).
4. *Curry E., McDaniel B., Shingarev D., Caires M.C., Leyden M., O'Riain S., Flannery K., Kirrane S., Green C., Walsh B., O'Morain L.* A Business Case for Enterprise Content Integration using Ontology-based Content Analytics / Proceedings of the 3rd Asian Semantic Web Conference on The Semantic Web (Industrial Track), 2007.
5. *Potts J.* Alfresco Developer Book. Customizing Alfresco with actions, web scripts, web forms, workflows, and more // Packet publishing. 2008.
6. *Perran A., Perran S., Mason J., Rogers L.* Beginning SharePoint 2010 – Building Business Solutions with SharePoint. – 2010.
7. Managed metadata overview (SharePoint Server 2010) [Электронный ресурс]. URL: [http://technet.microsoft.com/en-US/library/ee424402\(v=office.14\).aspx](http://technet.microsoft.com/en-US/library/ee424402(v=office.14).aspx) (дата обращения: 11.12.2012).
8. *Норт Д.* Совокупный экономический эффект при внедрении Microsoft SharePoint Server 2010 / 2010.