

УДК 004.75:681.3

УПРАВЛЕНИЕ КОНТЕНТОМ В ИНФОРМАЦИОННЫХ СИСТЕМАХ. ИСПОЛЬЗОВАНИЕ СТАНДАРТА CONTENT MANAGEMENT INTEROPERABILITY SERVICES

Бутенко М.Л.

Восточноукраинский национальный университет
им. Владимира Даля

Современные системы управления контентом имеют примерно одинаковую функциональность, но различную внутреннюю структуру данных. При возникновении задач по интеграции или переносу данных могут возникнуть существенные сложности. В 2010 году принят стандарт Content Management Interoperability Services, который позволяет обеспечить взаимодействие различных систем управления контентом. В данной публикации сделан анализ возможного использования стандарта CMIS при проектировании информационных систем.

Введение

Функции систем управления контентом (Content Management Systems – CMS) уровня предприятия можно разделить на несколько основных категорий [1]:

- Создание – предоставление авторам удобных и средств создания контента.
- Управление – хранение контента в едином хранилище. Обеспечивается интеграция с существующими информационными источниками и ИТ-системами.
- Публикация – автоматическое размещение контента на терминале пользователя.
- Представление – дополнительные функции, позволяющие улучшить форму представления данных; например.

Цель публикации – сделать анализ процессов управления контентом в CMS, а именно обмена данными между CMS и интеграции CMS в другие информационные системы. Данные проблемы особенно **актуальны** при практическом использовании нескольких CMS на одном предприятии или при миграции от одной CMS к другой. Подобные проблемы также возникают при работе с системами электронного документооборота (в западных публикациях CMS уровня предприятия и система электронного документооборота – это одно и то же).

В 2008 году корпорации EMC, IBM и Microsoft объявили о начале совместного проекта по разработке спецификаций Content Management Interoperability Services (CMIS) на интерфейс Web-сервисов, призванных обеспечить взаимодействие между системами CMS. В качестве главной цели было заявлено снижение затрат на создание среды управления контентом предприятия на базе нескольких хранилищ от разных производителей. Все названные компании подтвердили возможность взаимодействия их продуктов в рамках предложенного проекта спецификаций и приняли решение направить согласованный проект на одобрение в OASIS (Organization for the Advancement of Structured Information Standards). Стандарт CMIS v1.0 был опубликован в сентябре 2010 года [2].

Цели нового стандарта таковы:

- обеспечение работы новых приложений с имеющимися хранилищами и накопленным в них контентом;
- обеспечение независимого управления контентом разных хранилищ с помощью Web-сервисов;
- обеспечение общих Web-сервисов и интерфейсов Web 2.0 для упрощения разработки новых приложений;
- создание платформы, независимой от языковой принадлежности контента;
- поддержка разработки композитных приложений и «коллажей», то есть нового контента, составленного из контента нескольких источников и выглядящий как единое целое.

Общий замысел CMIS иллюстрирует рис. 1. Приложение Interoperable Content Application,

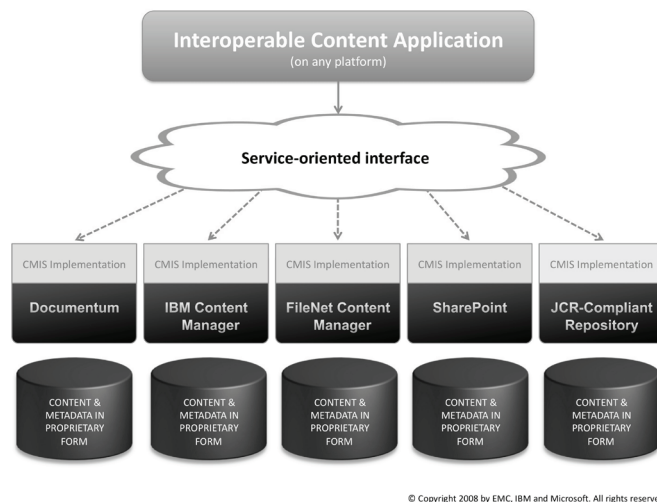


Рисунок 1. Принцип работы стандарта CMIS (рисунок взят с сайта компании EMC)

взаимодействующее с контентом из разных хранилищ, может размещаться на любой платформе и осуществлять это взаимодействие с помощью сервисного интерфейса и специальной надстройки (CMIS Implementation), которая делается каждым участником CMIS для своих хранилищ самостоятельно.

Конечно, стандарт CMIS создавался не на пустом месте. Методологии управления контентом у большинства производителей CMS примерно совпадают. Также не стоит забывать о стандартах JCR, WebDAV и Atom, которые также нацеливались на решение задачи интеграции.

JCR (JCR-170/JCR-283). Java Content Repository делает акцент на тех же проблемах упорядоченного доступа к хранилищам контента, что и CMIS, но не предусматривает широкого применения в среде Internet и других гетерогенных распределенных вычислительных средах. JCR — Java-стандарт, а CMIS использует Web-сервисы, которые не являются специфичными для какого-либо языка.

WebDAV. Этот стандарт — расширение протокола HTTP 1.1 для поддержки создания документов в Сети. В этом качестве он предлагает базовый уровень взаимодействия с контентом для Web-ресурсов. WebDAV определяет переносимую модель с простыми свойствами и поддерживает наборы ресурсов с помощью иерархии имен, подобной файловым системам. Эта простая модель дает гибкость, необходимую различным неструктурированным приложениям (наподобие публикации в Web), но не обеспечивает удовлетворительной дисциплины поддержки корпоративных приложений.

Atom. Протокол Atom Publishing Protocol (APP), построенный на базе HTTP, предназначен для публикаций и пополнения Web-ресурсов. Хотя APP становится все более распространенным интерфейсом для доступа к ресурсам Web, он не предоставляет достаточных возможностей моделирования для управления контентом предприятия. Эту задачу должен выполнить CMIS, используя APP в качестве одного из протоколов, с помощью которого приложения смогут доставлять контент.

1 Основы стандарта CMIS

Стандарт CMIS определяет четыре различных типа объектов: документы, папки, политики и отношения.

Связи между этими типами определяются следующим образом (рис. 2):

1. Каждый объект в хранилище CMIS имеет уникальный идентификатор объекта, который может быть использован как ссылка на объект, и множество именованных свойств для хранения данных различных типов (например, строки, числа и т.п.).
2. Документ может иметь несколько различных версий и может содержать несколько потоков данных (2b), чтобы представлять один и тот же контент в разных форматах.
3. Папки — это объекты, которые могут содержать другие объекты, то есть документы и другие папки. Если реализация хранилища поддерживает мультизагрузку, документы могут быть

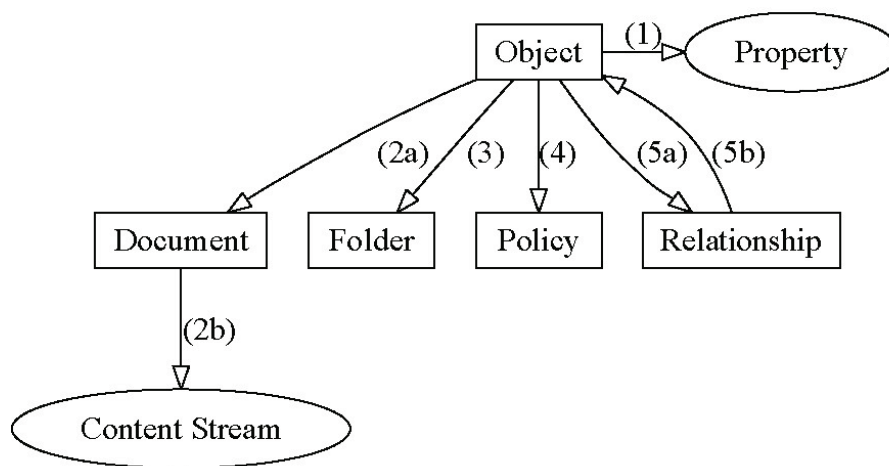


Рисунок 2. Модель данных CMIS

загружаться одновременно в более чем одну папку.

4. Политики – это объекты, определяющие варианты возможных действий, которые могут быть применены к объектам.
5. Отношения – это связи между двумя объектами (документы или папки) в конкретном хранилище.

Базовые сервисы CMIS:

- создание, получение, обновление, удаление объекта;
- поиск по объектам и ссылкам (в том числе поиск документов, ссылающихся на данных);
- сервисы для подключения политик управления к объектам, отключения политик от объектов.

Поиск строится на базе языка запросов подмножества SQL92, с некоторыми ограничениями и расширениями. Например, расширения включают возможность полнотекстового поиска (по контенту), поиск в многозначных реквизитах (реквизитах в форме списков), ограничение поиска по папкам и деревьям папок.

2 Использование стандарта CMIS при разработке информационных систем

Первый вопрос, на который необходимо ответить при реализации поддержки стандарта CMIS в информационной системе – что реализовать: хранилище (сервер контента) или клиент для доступа к контенту других хранилищ. В том случае, если мы решили реализовывать сервер, то мы обязаны реализовать как связывание на основе Web-сервисов, так и RESTful-связывание на основе протокола AtomPub (Atom Publishing Protocol) – всё в соответствии со стандартом. И это помимо реализации функционала обработки и хранения данных. Поэтому реализация CMIS-хранилища довольно сложная задача, однако если разработчик желает сделать по-настоящему универсальную систему управления контентом или систему электронного документооборота, то использование стандарта CMIS для построения хранилища данных обязательно.

Реализация CMIS-клиента гораздо проще. Все сводится к реализации вызова и обработки результатов нескольких десятков функций (можно привести пример функций из так называемой «шпаргалки» (cheatsheet) с сайта Alfresco – [3]). К тому же при реализации CMIS-клиента можно опираться на существующие разработки, например, [4].

При реализации стандарта CMIS в своей информационной системе можно ориентироваться, главным образом, на два известных программных продукта, которые поддерживают CMIS – Alfresco и Microsoft Sharepoint Server 2010:

Alfresco – тиражируемая интегрированная система управления контентом для организаций разработки одноимённой британской компании. Используется для управления документами, записями, веб-публикацией, групповой работой в организации, [5]. Поддержка CMIS реализована

начиная с версии «3.3». На сайте компании имеется большое количество материалов и документации по CMIS.

Microsoft Sharepoint Server может быть использован для создания сайтов, предоставляющих пользователям возможность для совместной работы, [6]. Поддержка CMIS реализована начиная с версии «2010».

Приведем примеры вызова нескольких функций CMIS-клиента (без привязки к конкретному языку программирования):

- *getFolderTree* – получить список документов, обязателен аргумент repositoryId – идентификатор хранилища;
- *createDocument* – создание документа, указывается хранилище и папка;
- *createFolder* – создание папки, указывается хранилище и родительская папка;
- *getObject* – получение документа;
- *getContentStream* – получение контента документа, необходимо указать MIME-тип.

3 Элементы стандарта CMIS в экспериментальной системе электронного документооборота QDocuments

Разработка системы электронного документооборота QDocuments ведется автором, начиная с 2010 года. Система является экспериментальной и не используется на практике. Она создавалась, чтобы промоделировать работу распределенной системы электронного документооборота (СЭД), использующей ряд новых технологий. Это агентно-ориентированный подход к проектированию СЭД, реализация глобальной системы сообщений и, конечно, методологии интеграции и взаимодействия информационных систем, одной из которых является стандарт CMIS.

При разработке системы QDocuments не ставилась цель реализовать CMIS-сервер. В QDocuments реализовано несколько способов доступа к документам, одним из которых и является стандарт CMIS. Создан модуль системы QDocuments, который реализует CMIS-клиента. Данный модуль в настоящее время проходит тестирование (главным образом – обеспечение работы всех функции CMIS-клиента при взаимодействии с Alfresco).

Выводы

Проведен анализ возможного использования стандарта Content Management Interoperability Services при проектировании информационных систем. Показано, что стандарт CMIS - это актуальная технология взаимодействия и обмена данными в информационных системах. Рассмотрены особенности стандарта CMIS при создании системы электронного документооборота.

Литература

- [1] Полубояров В.В. Введение в технологии создания Интернет-узлов. Электронный ресурс. Режим доступа: http://www.intuit.ru/department/internet/inwwwtech/7/inwwwtech_7.html
- [2] Content Management Interoperability Services (CMIS). Электронный ресурс. Режим доступа: <http://xml.coverpages.org/cmisis.html>.
- [3] Content Management Interoperability Services Cheat Sheet. Электронный ресурс. Режим доступа: cmis.alfresco.com/cmisis-cheatsheet.pdf
- [4] Джей Браун Библиотека API-интерфейса для Python, соответствующая спецификации CMIS: Часть 1. Введение в cmislib. Электронный ресурс. Режим доступа: <http://www.ibm.com/developerworks/ru/library/x-cmis2/index.html>
- [5] Alfresco (ECM-система). Материал из Википедии – свободной энциклопедии. Электронный ресурс. Режим доступа: [http://ru.wikipedia.org/wiki/Alfresco_\(ECM-система\)](http://ru.wikipedia.org/wiki/Alfresco_(ECM-система))
- [6] SharePoint. Материал из Википедии – свободной энциклопедии. Электронный ресурс. Режим доступа: <http://ru.wikipedia.org/wiki/SharePoint>