A New Approach to Implement Enterprise Content Management System Using RSS and Folksonomy

Jianlin Wu[1] and Guocong Yan[2]

[1]Beijing Key Laboratory of Intelligent Communcations Software and Multimedia, School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100876, P.R.China jlwu@bupt.edu.cn
[2]Telecommunications Software Engineering Group, School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100876, P.R.China yanguocong@gmail.com

Abstract. Nowadays, 85% of the enterprise information is stored as unstructured data. And ECM (Enterprise Content Management) provides enterprises with a platform to house these unstructured contents and deliver them in a proper way. However, content sharing, one of the key aspects of ECM, is still not efficient and effective enough in most of the ECM systems. In this paper, we propose a new enterprise content management approach to make it easy to share and deliver enterprise content. The approach mainly takes advantage of RSS (Really Simple Syndication), a content syndication format that is a popular technology in the Web2.0 world, to solve the problem of enterprise content sharing and delivery in an ECM system. Also, the approach adopts the Folksonomy concept, a user generated taxonomy emerged in Web2.0 world, in an ECM system to classify the enterprise contents in a proper way so that the ECM system users can find the contents they want more accurately and efficiently. By bringing the Web2.0 technology and philosophy to enterprise, we make ECM system more efficient and effective, especially in content sharing, searching and delivering, which will help to achieve the goal of information on demand in enterprise.

Keywords: *Enterprise content management, Information management, Web2.0, RSS, Folksonomy*

1. INTRODUCTION

As enterprise content management (ECM) is becoming more and more important to the enterprise information management, ECM system is being given more and more attention to. However, since a large portion of the information stored in the ECM system is unstructured content, it is not easy to share these kind of content as those

structured data does. People have been proposing different methods to improve the efficiency and effectiveness of content sharing and delivery, but it is still not good enough to satisfy the need in the enterprise.

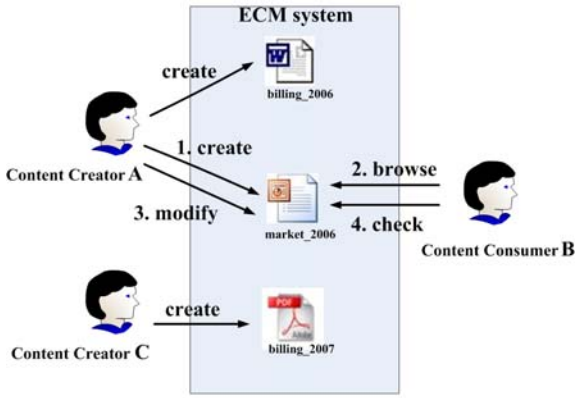

Figure 1. Traditional ECM scenario

In this paper, by introducing the web 2.0[1] technology and concept into the enterprise field, we propose an approach to implement enterprise content management using RSS technology and Folksonomy philosophy in order to solve the following issues that lead to the inefficiency and ineffectiveness of sharing and delivering enterprise contents[2].

1. It is difficult to trace some certain topics or some content creators that the content consumers are interested in. When changes on the existed contents are made, the content consumers who concern the changes are not notified.
2. Though some ECM systems provide traditional taxonomy method to classify the contents in which users can label the predetermined tags to the to-be-created content, it is still not accurate and flexible enough since it only provides predetermined terms.

In ECM system based on RSS and Folksonomy, user can define the tags that he thinks will match the to-be-created contents the most during the creation of the content. Due to the use of RSS technology, besides being able to trace the modification of the existed contents, content consumer is also able to trace the content on the topics or the authors he is interested in. In the first way, the content is labeled on the tags that is most matched to the intention of the content, which will help content consumer find the wanted contents more accurately. In the second way, it will improve the efficiency of content sharing and delivery.

1.1 Traditional Scenario

In the traditional ECM system, there are some certain issues in the content management. Figure 1 introduces the use case in a traditional ECM system. As seen in the figure, the cycle of the content maybe as follows:

1. A content creator A produces content and since the classification mechanism of the ECM system is taxonomy, the content creator has to choose from the predetermined tags, which maybe limited for him to express the content;

2. Some content consumer B navigates the content in the ECM system and finds the content that A has created through either by searching the title of the content or by browsing the hierarchy structure of the content, and then consumes the content;

3. When content creator A modifies the content afterward, content consumer B will not be notified that the content has been modified.

4. In case content consumer B wants to trace the content he has consumed, he has to check the state of the content from time to time, whether the content has been modified or not.

Obviously, it is not efficient for content consumers in such a scenario. Actually, presume content consumer B is interested in all content creator A's contents (maybe A is the project leader of B), or B is attractive to the contents focus on "billing" (maybe B is work in the billing department of the company). In both cases, B will not be notified when A creates a new content or some other content creator C produces a content related to "billing". Thus, in a traditional ECM system, content consumer has to browse the content through the hierarchy structure of the metadata and has no way to trace some certain topic or some content creator that the content consumers are interested in. And as a result, the former leads to the ineffectiveness of the ECM system and the latter leads to the inefficiency.

2. SOLUTION APPROACH

To solve the issues mentioned in section 1.1, we introduce two technologies to the ECM system: RSS and Folksonomy, both of which are emerged in Web2.0 era and are so popular nowadays.

2.1 Really Simple Syndication

The RSS format [3], an acronym for Really Simple Syndication, is an XML dialect that is best understood by first examining its relation to earlier web-based push technologies, and then its new ability to enable widespread content syndication on the internet, including time-stamped personal weblogs [4]. A RSS file example is as follow:

```
<rss version="2.0">
<channel>
    <title>My blog</title>
    <link>http://tsegblog/myblog</link>
        <item>
            <title>This is the second post</title>
            <link>http://tsegblog/myblog?id=2</link>
            <description>
```

```
            Finish the job today.
            </description>
            <pubDate>Fri, 12 May 2007 19:00:00 EDT</pubDate>
        </item>
        <item>
            <title>This is the first post</title>
            <link>http://tsegblog/myblog?id=1</link>
            <description>
            Very happy to start blogging.
            </description>
            <pubDate>Thu, 11 May 2007 15:00:00 EDT</pubDate>
        </item>
</channel>
</rss>
```

However, RSS is not just for weblogs and news. It has expanded to a variety usage nowadays. Pretty much anything that can be broken down into discrete items can be syndicated via RSS: the most popular song list, the result return by a web search engine like yahoo.com, even the stock price at a certain moment. Once information about each item is in RSS format, an RSS-aware program called RSS aggregator can check the RSS for changes and react to the changes in an appropriate way [5]. For example, once you subscribe RSS in a RSS aggregator, it can help you keep up with all your favorite weblogs by checking their RSS feeds and displaying new items from each of them.

2.2 Folksonomy

Folksonomy [6], which is a combination of "folk" and "taxonomy", is a new classification approach which is different from the past well known taxonomy classification. The term, given by Thomas Vander Wal in a discussion on an information architecture mailing list when referring to the organic system of organization developing in websites Delicious (http://del.icio.us) and Flickr (www.flickr.com), is a newly emergent philosophy in Web2.0 era.

The main difference between Folksonomy and traditional classification taxonomy is that: though both classification approaches are comprised of terms, there is no hierarchy, and no directly specified parent-child or sibling relationships between the terms in Folksonomy, while there are multiple kinds of explicit relationships between terms in taxonomy. The folksonomies are simply the set of user generated terms, which are called "tags", rather than a predetermined set of classification terms in taxonomy.

Take web page classification for example, in the case of using taxonomy, someone constructs the classification tree at first, and then when end user is going to classify the pages, he has to classify the pages based on the already established tree. An example of taxonomy tree is Yahoo Directory. On the other hand, in Folksonomy, end user put tags to each page freely and subjectively, only if he thinks the tag is

suitable to describe the web page. Anyone can choose any word as tag, and can put more than one tag to one page. Obviously, Folksonomy is a bottom up approach while taxonomy is a top down one. Since the freedom is given to the creator, Folksonomy classification can reflect users' actual interest more accurately [7].

As the benefit aforementioned, we will introduce folksonomy in the ECM system, hoping to make use of the convenient that tags will bring to us during managing the enterprise contents.

2.3 Improved Scenario

As the advantages of the above two technologies, we develop an ECM system combining the two together to solve the content sharing and delivery issues, so as to improve the accuracy and efficiency of the Enterprise Content Management.

The improved scenario using our ECM system is as follows in detail (as shown in Figure 2):

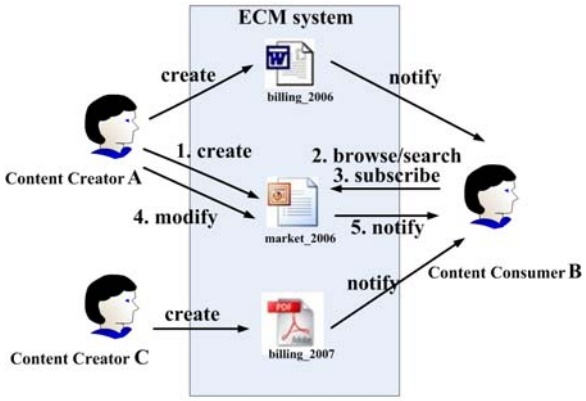

Figure 2. Improved ECM scenario

1. Like the content creation step in the ECM system mentioned in Section 1, content creator A creates a content in our ECM system. Instead of using the predetermined tags, A can use any tags to describe the content as freely as he wants. So, the content will be labeled with the tags that A think will be more useful for the future content consumption.

2. When content consumer B navigates in the ECM system looking for contents, besides what he can do in the traditional ECM system, he can also search through the tags. For example, he is looking for some materials on "billing"; he can search the contents that have been labeled on "billing". Then all the "billing"-related contents will return to him for his further consumption.

Then, the following steps are different from the traditional ECM system because of the introduction of RSS and Folksonomy in our ECM system.

3. As our ECM system implements RSS, after the above steps, content consumer can subscribe to three object level so he can trace the related contents afterwards. First, he can subscribe to the content he found. By this way, he will be notified when the content is modified sometime in the future. Second, he can subscribe to the tag so he can trace a specific topic. For example, if content consume B is interested in all the "billing" related contents, he can subscribe to the tag "billing". Whenever there is a new content labeled on "billing", or there is a modification occurred in the existed "billing" related content, he will be notified. The last object level that content consumer can subscribe to is content creator level. By doing this, the content consumer is able to trace someone's content production. When the content creator that B has subscribed to creates a new content or modified an existed content, B will be notified. As a RSS aggregator is able to consume the RSS operation such as subscription and unsubscription, content consumer can launch these operation in the RSS aggregator integrated in our ECM system. Besides, the notification also happens in the RSS aggregator. So every time the content consumers log in the RSS aggregator, he will be notified whether there are changes in the contents he is interested in.

4. Also, the content user can unsubscribe all the three object level mentioned above while he is no longer interested in the related contents. This way is more convenient than the email metaphor approach [8], in which content consumer can only receive the contents whether he likes or not passively.

All the above, introducing RSS and Folksonomy to ECM system completes the content life cycle, from content creation to consumption, making it a closed loop, promoting the productivity of an ECM system.

3. ARCHITECTURE AND IMPLEMENTATION

3.1 System Architecture

Figure 3 shows implementation architecture of the ECM system present in the paper. As the popular MVC architecture, the system is divided into three layers: presentation layer, core layer, and data layer.

First, the presentation layer represents the user interface of the system, which provides a way for user to interact with the system. As presentation of the Portal, there are two main parts of the layer. Content operation part provides an interface for user to launch the content operation, including content creation, management, navigation and search. The other part in this layer is RSS aggregator, in which user can launch RSS related operation such as subscription and unsubscription. Besides, the notification of the creation and modification of content is also launched in it.

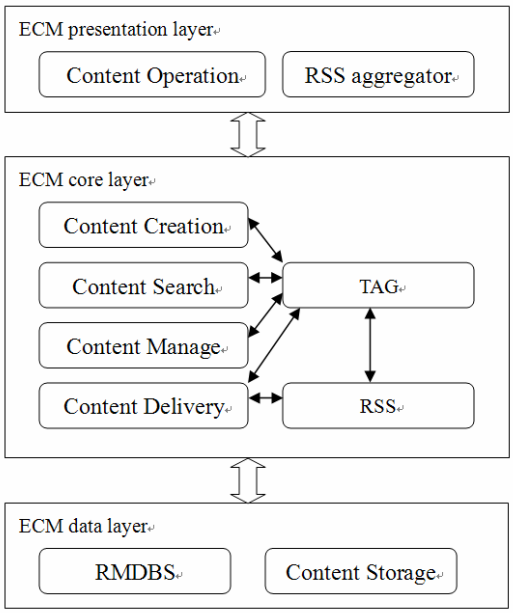

Figure 3. Implementation Architecture of the ECM System

The second layer of the system, the core layer, is the main part of the ECM system. As we can see in the figure, the right hand side's four parts of the core layer are all content-related components including content creation, content search, content management and content delivery. These are all basic components of a traditional ECM system. In our ECM system, there are two additional components, TAG component and RSS component. According to these two additional components, the new ECM system is able to provide the functions mentioned in the section 2.3. Since all the contents in our system will be labeled on some tags, all the content-related components interact with the TAG component, which is responsible for the management of the tags, while RSS component is just responsible for the content delivery.

The data layer is in charge of the data resource, including RDBMS and Content Storage. The former component is used to store the relational information of the system, while the latter is used to store various kinds of contents.

3.2 RSS implementation

The blow two RSS files are samples generating by the RSS component mentioned in Figure 3 in section 3.1, which represent two of three object levels subscription mention in section 2.3. The RSS on the left represents the subscription to the content tag with "billing", while RSS on the right represents the subscription to the author "Johnson", both of which are indicated in the elements <title> and <link>.

Also, the <link> element represents the RSS file's URL. The contents are divided into items, just as news items in a news RSS file. There are several elements in the content item. The element <title> describes the name of the content file, <link> represents the URL of the content, <description> represents the annotation of the content and <pubDate> indicates the creation date of the content. The content object level subscription is almost the same except the <pubDate> element in every item is different, so as to differentiate the change of the content.

```
<rss version="2.0">
<channel>
<title>billing</title>
<link>http://ecm.tseg.org/tags/contents/billing.rss</link>
<item>
<title>billing_2006.xls</title>
<link>http://ecm.tseg.org/content?id=2</link>
<description>
Some description about the content goes here.
</description>
<pubDate> Mon, 11 Dec 2006 15:21:30 EDT </pubDate>
</item>
<item>
<title>biling_2007.xls</title>
<link>http://ecm.tesg.org/content?id=5</link>
<description>
Some description about the content goes here
</description>
<pubDate> Fri, 11 May 2007 16:10:20 EDT </pubDate>
</item>
</channel>
</rss>
```

```
<rss version="2.0">
<channel>
<title>Johnson</title>
<link>http://ecm.tseg.org/tags/authors/Johnson.rss</link>
<item>
<title>billing_2006.xls</title>
<link>
http://ecm.tseg.org/content?id=2</link>
<description>
Some description about the content goes here.
</description>
<pubDate> Tue, 12 Dec 2007 16:30:12 EDT </pubDate>
</item>
<item>
<title>marketing_2006.ppt</title>
<link>
http://ecm.tesg.org/content?id=85</link>
<description>
Some description about the content goes here
</description>
<pubDate> Wed, 13 Dec 2006 15:340:24 EDT </pubDate>
</item>
</channel>
</rss>
```

Once a content creator publishes a new content, all the RSS file correspondent to the tags he labeled will be added a new <item> element which contains the <title>, <link>, <description>, <pubDate> elements. And afterward, when a content consumer logs into the ECM system, the system will check all the RSS files that he subscribed to and look for new published contents. If there are new contents, the system will notice the user in the RSS aggregator telling user that there are new contents that he is interested in. The same steps occur when there are some modifications in the existed enterprise contents.

4. SUMMARY AND FUTURE WORK

In this paper, we proposed a content management approach that combines RSS and Folksonomy. And the ECM system we introduced an experimental application already used in a middle scale department. Our proposal prevents inefficiency of sharing and delivering enterprise contents.

1. By allowing users attach any tags that he think will match the most instead of predetermined tags to the created content, the contents are classified more accurately and when navigating by user, the return results will be more meaning focused.
2. The content consumer can trace some certain topic or some content creator that he is interested in by subscribing to the topics or authors. As a result, once there is modification or addition of related contents, the subscriber will be notified. In this way, the efficiency of content sharing and delivery is improved and also give the right to the content consumer to consume the content in more effective and positive way.

In the future, we will keep on taking advantage of more web2.0 concept such as rating, commenting as well as group sharing in the ECM system, hoping to help improve the content delivery accuracy, also the current version demands more privilege control on the content sharing, which is also an important aspect in enterprise application.

REFERENCES

1. T.O'reilly, *What Is Web 2.0*, O'Reilly (2005). http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (Accessed July 13, 2007).
2. T. Paivarinta and B.E. Munkvold, Enterprise Content Management: An Integrated Perspective on Information Management, in *Proc. of the 38th Annual Hawaii International Conference (HICSS) (*Jan. 03-06, 2005), pp.96-96.
3. D. Winer, *RSS Specification2.0*, RSS Advisory Board (2007). http://www.rssboard.org/rss-specification (Accessed July 13, 2007).
4. D. Bourges-Waldegg and C. Hoertnagl, Combination of RSS Newsfeeds and Forms for Driving Web-based Workflow, in *Proc. of e-Business Engineering, ICEBE, IEEE International Conference (*Oct 18-21. 2005), pp.142-149.
5. M. Pilgrim. *What is RSS?* (2002). http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html (Accessed July 13, 2007).
6. J. Golbeck, B. Parsia, and J. Hendler, *Folksonomies-Cooperative Classification and Communication Through Shared Metadata* (2004). http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (Accessed July 13, 2007).

7. S. Niwa, T. Doi, and S. Honiden, Web Page Recommender System based on Folksonomy Mining for ITNG ’06 Submissions, in *Proc. of Information Technology: New Generations, ITNG, Third International Conference* (April 10-12, 2006), pp.388-393.
8. K. Nakatsuka and T. Ishida, Content management for inter-organizational projects using e-mail metaphor, in *Proc. of Applications and the Internet (SAINT), International Symposium* (Jan. 23-27, 2006), p.4.