Applying Lessons Learned from a Case Study of Metadata in Web Publishing to Internal Corporate Intranet Document Management

Chris Helwig
26905 North 22nd Ave.
Phoenix, AZ 85085
University of Wisconsin – Milwaukee
Ph.D. Student in Computer Science
623-242-7648
Chris.Helwig@gmail.com
Advisor Prof. Ichiro Suzuki

ABSTRACT
Why is it that searches of the Internet using Google often return the webpage sought on the first attempt while searches of corporate intranets often take several searches to find desired documents? Part of the reason lies in the metadata that exists on the Internet -- information in tags or words contained in hyperlinks describing the content of information contained in webpages and documents posted on the Web. [1]

Corporate intranet website managers can apply the techniques of web publishing to their document management strategies in order to make internal company intranet searches more effective. This poster examines the techniques of one such web publisher, Articlesbase, and suggests how such techniques can be applied by corporate intranet document managers.

Categories and Subject Descriptors
H.5.4 Hypertext/Hypermedia.

Keywords
Metadata, Hypertext, Web Publishing, Information Retrieval.

1. DISCUSSION
Articlesbase, http://www.articlesbase.com/faq.php, "is a free article directory where you can submit and find articles. You can publish your articles for free or find content for your website, ezine, or newsletter." Articlesbase provides an excellent example of how the use of metadata can increase the number of hits an article posted on the Internet receives.

I had written ten articles I wanted to post on the Internet. Many were from speeches or five-minute presentations on various topics which I had previously presented at the local Toastmasters club. I first posted them on the Internet using the space provided by my Internet service provider. A hit counter installed on my website showed that my articles had received virtually no hits. I then published the same articles on the Articlesbase website and the number of hits increased dramatically.

Part of the process for submitting an article to Articlesbase involves answering a series of questions about your article. At the article submission webpage you are first asked to select a category for your article. You are then asked for the article title and for a short summary of the article and for keywords associated with your article.

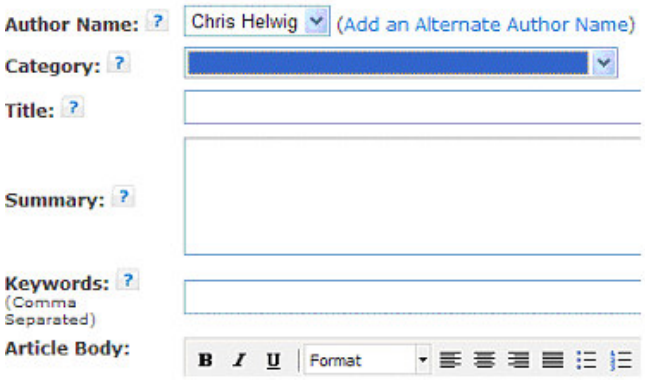

Figure 1. Screen shot of form asking author to select a category.

When the article is posted on the Web several tags have been added to the HTML that provide more information about the article. For example, the following two meta tags were added to one of my articles:

<meta name="description" content="An Automated System for Pulling Economic Data and Graphically Displaying it on the Web." />

<meta name="keywords" content="Xml, Economic Data" />

These two meta tags contain the keywords entered at the time the article was submitted as well as the description of the article. This information makes it easier for search engines to find the article for search engine users using these keywords or words contained in the description.

Category information also appears in HTML tags. The category information appears not in a meta tag but rather embedded within an HREF link path to the article:

href=http://www.articlesbase.com/information-technology-articles/an-automated-system-for-pulling-economic-data-and-graphically-displaying-it-on-the-web-506829.html

Including category information, in this case the words "information technology," is another way of embedding a form of metadata into the HTML in a manner that aids search engines in locating the document on the Internet.

Figure2. Screen Shot of categories selected for articles.

Title	Category
An Essay on the White Whale	Fiction
Using Sas to Analyze System Performance Metrics	Programming
Current Issues in Environmental Science	Environment
How to Help a Friend Quit Smoking	Quit Smoking
Joining and Fully Utilizing a Gym	Fitness
Philosophy 101	Metaphysics
Intro to Ice Skating	Hockey
Patent Law 101	Patents
An Automated System for Pulling Economic Data and Graphically Displaying it on the Web	Information Technology
Sas Programming Course - Proposal	Programming

Metrics from Articlesbase indicate that with the metadata, the articles are receiving numerous hits.

Figure 3. Screen shot of statistics on article views.

Title	Views
An Essay on the White Whale	43
Using Sas to Analyze System Performance Metrics	124
Current Issues in Environmental Science	135
How to Help a Friend Quit Smoking	196
Joining and Fully Utilizing a Gym	38
Philosophy 101	14
Intro to Ice Skating	221
Patent Law 101	87
An Automated System for Pulling Economic Data and Graphically Displaying it on the Web	60
Sas Programming Course - Proposal	80
Totals	**998**

Often when documents are posted on company intranets, no such metadata is generated. No meta tags are used. And there may be no hyperlinks set up containing words related to the content of the documents.

So what can company librarians and intranet site managers learn from Web publishing? First they can try to set up an automatic process much like the Articlesbase website with an automatic web form used to extract metadata about the article that can be used to automatically generate HTML containing metadata. Second, they can identify a topic or category that identifies the content of the document. The Articlesbase website uses a drop-down list of topics that authors can select from. Third, they can leverage the authors of the document and have them identify the keywords needed for a particular article. Thus the person with the most knowledge of the article is the one selecting the keywords. And fourth, they can utilize hypertext markup language which has built-in support for categorizing and placing metadata within tags that make it easier to locate such documents. Finally they can collect metrics on the number of hits documents received with and without the metadata. Metrics will indicate the effectiveness of the metadata.

2. REFERENCES

[1] Schabes, Yves, Enterprise Search and Automatic Metadata Generation, presented at the M2008 Datamining Conference.