

Folksonomy Optimization in Content Management

Nathan S. Wilson

Taxonomist
 Project Performance Corporation

October 2011

Introduction

Your organization has information and your users need to find that information. Do you know the answers to these questions?

- How much time do your users spend searching for information?
- How often do those searches fail to find the information your users need?
- How much time do your users waste on those unsuccessful searches?

According to a study conducted by International Data Corporation (IDC) and other organizations:¹

- Users spend from 15 to 35 percent of their time searching for information
- They find what they are looking for 50 percent of the time or less, and only 21 percent of users said they found the information they needed 85 to 100 percent of the time
- Forty percent of users reported that they cannot find the information they need to do their jobs on their organization's intranets

To improve users' ability to retrieve information, content management systems can allow users to tag content using both terms that the organization has created (taxonomy) and terms the users create themselves (folksonomy).

Folksonomy is a very loose tagging schema that is developed organically by the users; whereas *taxonomy* is a very rigid classification schema often developed by a subject matter expert. Despite the difference in form, both serve the same function -- to increase the effectiveness of information retrieval thereby increasing the effectiveness of your organization.

The Problem

Taxonomy benefits the organization, but not always the individual user. Folksonomy benefits the individual user, but not always the organization. Taxonomy benefits the organization by utilizing a subject matter expert and providing a structured and hierarchal view of the organization's information. Folksonomy benefits the user by drawing on his/her knowledge of the information. Every user applies his/her own meaning, understanding, and vocabulary to the content. Subject matter experts are often not users, and users are not always subject matter experts.

Unlike taxonomy, folksonomy is not governed by controlled vocabulary standards; there is no structure, nor any conceptual relationships among the terms. As such, the terms used may have multiple meanings and, because the folksonomy relies on the user's own idiolect and vocabulary, the terms used may also be jargon, slang, or abbreviations.

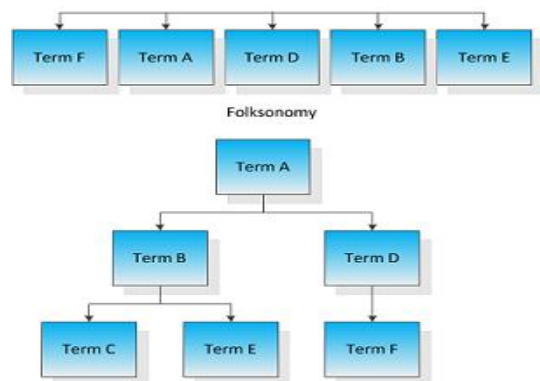


Figure 1: Unlike taxonomy, folksonomy is not governed by controlled vocabulary standards; there is no structure, nor any conceptual relationships among the terms

¹K.S. Taylor. "The brief reign of the knowledge worker," 1998. Cited by Sue Feldman in: "The high cost of not finding information." 13:3 KM World (March 2004) <http://www.kmworld.com/articles/readarticle.aspx?articleid=9534>



How does folksonomy benefit the individual user and not the organization?

User Alice could tag content that discusses the North American Whooping Crane with the following terms: tall, white, and crane. She knows what those terms mean to her, she understands them and how they relate to the content. User Bob, however, does not. User Bob could search for content tagged with those same terms, but if his understanding of those terms or the meaning he assigns to them is different from Alice's, then Bob's search will be less effective. For example, Bob could have been searching for information regarding tower cranes used in construction.

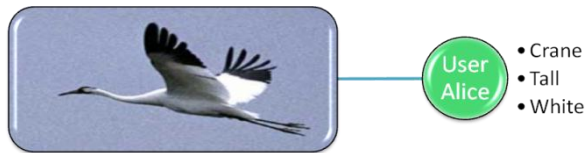


Figure 2: What Alice Tagged



Figure 3: What Bob Wanted

Another problem can arise when multiple users tag the same content using similar but different terms. For example, to describe a North American Whooping Crane, users Alice, Bob, and Charlie could each tag the same content using slightly different terms. User Alice tags the content with "crane," User Bob tags the content with "Aves," and User Charlie tags the content with "bird."

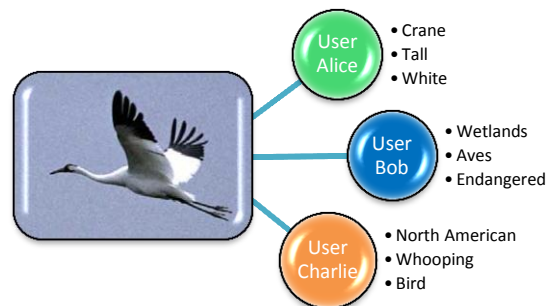


Figure 4: Different Tags for the Same Content

If User Charlie did not know that User Bob uses the term "Aves" instead of "bird," User Charlie could miss relevant information that was tagged by User Bob.

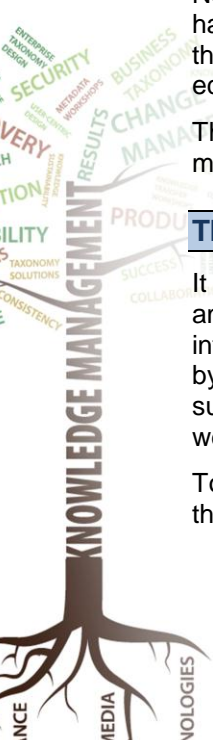
None of these tags is wrong; they are just different words used to describe the same content. This happens because different users and groups approach content from different perspectives. However, if the users do not have a clear agreement on the tagging system, the tags themselves can become eclectic and will hamper information retrieval instead of improving it.

The question remains, how do we take folksonomy, which is designed to benefit the individual user, and make it beneficial for the organization?

The Solution

It may seem that folksonomy cannot be beneficial to the organization. Terms in a folksonomy may be ambiguous and there is no control over terms that are synonymous. These factors can hinder effective information retrieval. For folksonomy to benefit the organization, it must align with the taxonomy used by the content management system. There are several content management systems on the market, such as SharePoint, Drupal, and Documentum, which use taxonomy along with folksonomy. So how do we bring the folksonomy into alignment with the taxonomy?

To align the taxonomy and folksonomy, Project Performance Corporation (PPC) recommends leveraging the folksonomy and the taxonomy in concert with each other. This is accomplished by managing both the



folksonomy terms and the taxonomy terms. Because many content management systems provide tools to manage both effectively, aligning taxonomy and folksonomy terms is not an issue of technology, but of methodology. The most effective methodologies are those that utilize the dynamism and flexibility of folksonomy, while providing the consistency and broad usability of taxonomy. By building a methodology for continued analysis and improvement of the folksonomy and taxonomy into your business processes, you can harness the power of both and improve search and information retrieval for your organization.

Let your users continue to tag content as they see fit. As terms are added, new ones will be included in the folksonomy. This should allow for a more organic growth of the folksonomy. After a period of time, a taxonomy editor should perform an analysis of the folksonomy.¹ This analysis should involve looking at the frequency with which tags are used, along with any organization-specific criteria.

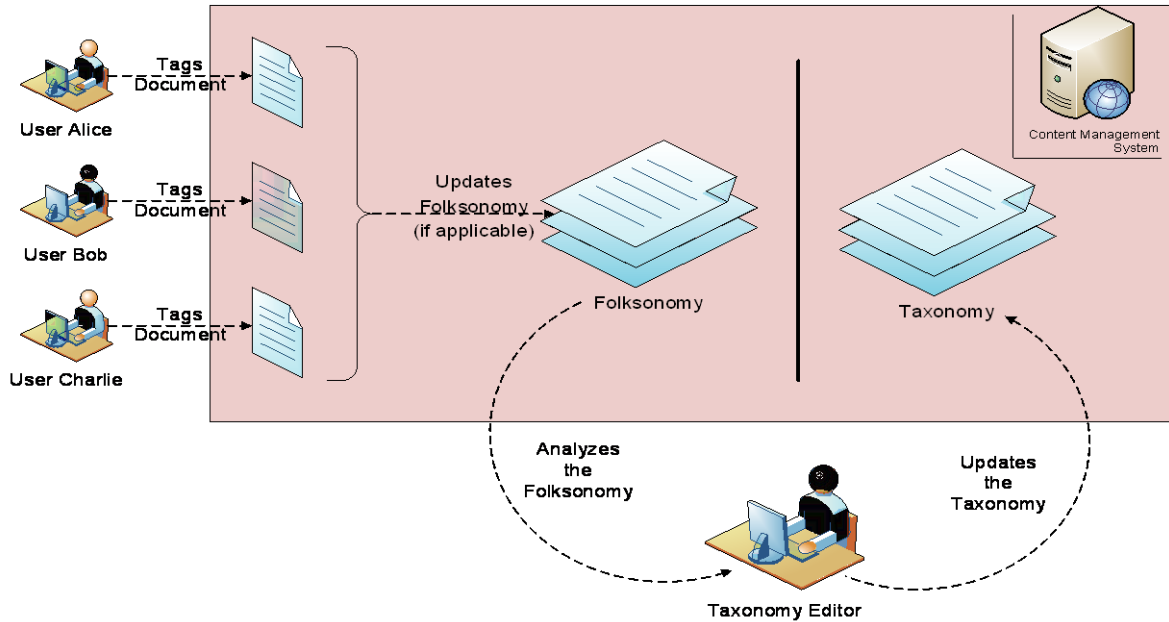


Figure 5: Folksonomy Optimization Process

The goal is to improve the taxonomy by harnessing the collective knowledge of your users. Tags used with high frequency, or that meet the specific organizational requirements, become candidates for inclusion in the taxonomy. Tags with low frequency and those that do not meet the organization's needs should still be kept, but should remain in the folksonomy. This way, the folksonomy still benefits the individual users by keeping their tags.

Once candidate terms have been decided upon, the taxonomy editor can update the taxonomy with the new terms, as needed. These terms can be included as new terms or as synonyms to existing terms in the taxonomy. This methodology and process allows your organization to utilize the understanding of your data by its users and aligns it with the expertise that comes from a subject matter expert.

Conclusion

Your organization is unique, and so is your content and information. There is no one-size-fits-all solution to the optimal usage of both folksonomy and taxonomy in content management. Yet, the goal is the same -- make your content quicker and easier to find. If your organization has content, allow PPC to show you how we can work together to meet your organization's unique information retrieval needs. By incorporating a methodology and practice that will leverage the knowledge of your users with the expertise of a subject matter expert, your organization can successfully implement folksonomy optimization in content management.

¹ PPC recommends that a taxonomy governance boards meet at least quarterly. Similarly, the folksonomy should also be reviewed at a minimum of once per quarter.

About Project Performance Corporation

Project Performance Corporation is part of the AEA group, a 1,200-person, multi-disciplinary team of information technology and knowledge management professionals, project management experts, scientists, technologists, and regulatory specialists. With more than 20 years of experience in information management and enterprise services implementation, we specialize in delivering effective, reliable answers to the most complex challenges. At the heart of it all are our people - innovative thinkers, well versed in business processes and drivers, with unequalled focus on delivering quality products and services.

About the Author

Mr. Nathan S. Wilson is a Senior Associate in Project Performance Corporation's Knowledge Management practice. Mr. Wilson has extensive experience implementing taxonomy strategies as part of manual and automatic categorization schemas. In addition to working with commercial businesses to improve their content management needs, he has worked with library and archival organizations to better optimize their information retrieval methods and processes.

For more information contact:

Nathan S. Wilson
Taxonomist
Project Performance Corporation
1760 Old Meadow Rd., McLean, Virginia 22102
p: 703.748.7254
e: nathan.wilson@ppc.com

