

**FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES**

**OUR FOLKSONOMY
SUBMITTED FOR LIS5703 INFORMATION ORGANIZATION
PROFESSOR: DR. MICHELLE KAZMER**

By

JAMIE WITMAN

Graduate Student – LIS Master's Program

Our Folksonomy:

Before, we can analyze the folksonomy the class created for our Contribute Four and Five assignments we must first look at the existing structure that a folksonomy of any kind fits into. In the increasing technological environment, the cataloging and indexing of sources has become a daunting task that requires patience, skill, and most importantly time. It is through vocabulary control that databases and other resources are able to provide a consistent and “controlled” list of metadata for particular sources. As Taylor defines it, “A controlled vocabulary is a list or database of subject terms in which all terms or phrases representing a concept are brought together” (2009). This means, that databases use specific words and word phrases to categorize and define an exact subject area. Furthermore, controlled vocabularies are broken down into three different and distinct categories: Subject heading lists, thesauri, and Ontologies. Each use different methods for recalling metadata for particular resources and are generally used for different purposes. For instance, Subject heading lists are mostly used in libraries, while thesauri are mostly used for indexing and cataloging and intend to be used by a specific audience. Furthermore, Ontologies seem to exist in a more abstract context. As defined by Taylor, “An ontology is an attempt to define the ‘essence’ of a situation, domain or conceptual framework” (2009). Due to this, Ontologies use what is called Natural Language Processing to organize and define the metadata used in its records. This concept of using natural language to organize metadata removes the idea of a “controlled” vocabulary and replaces it with an idea of collaboration among laypeople to determine what is most useful in regards to recalling metadata for particular sources. Furthermore, some LIS professionals see ontologies as a substitute for the more traditional thesauri and subject heading lists due to freedom provided by its natural structure (Hilera, et al. 2010). This is due to the participation by the public sector to incorporate their own set of information into the metadata used by these sources. For these reasons, it is clear that the folksonomy created by the class can only fall under the umbrella of Ontologies.

Through the concept of Ontologies, we find a place where not only the trained and certified indexers and catalogers can provide the terms and phrases used to organize metadata but frequent internet users and volunteers can contribute their natural language concepts to provide a more in depth and well-rounded set of metadata. This is how a folksonomy is created. These users are able to input keywords about a resource and create tags which determine where the item is placed in the organization of the metadata. As Morrison points out, these

folksonomies are derived from the traditional taxonomies controlled by experts however they do not exist in a hierarchical fashion (2008). Instead, non-expert users can provide the tags to create a broader more accessible to the public resource. Furthermore, as Kim points out, this introduction of ontologies has provided an active network for collaboration (2009). It truly is a place for users of all types of resources can contribute and interact.

This is how we created our own folksonomy for class. As non-experts we located resources through an array of databases and used specific “tags” that as individuals we decided were the most descriptive and accurate for each source we found. Again as Morrison points out this allows for a more comprehensive way to search for resources on a particular resource (2008). Our folksonomy seemed to revolve around the same few tags: tagging, ontologies, social tagging, folksonomy, and controlled vocabulary; with “tagging” and “folksonomy” as the top two. Overall, it seems like the tags fell into three types of categories: subject, purpose, and tasks. Furthermore, the folksonomy created by the class seems to be fairly accurate and does not contain any tags that are inappropriate for the resource they are describing.

However, there are issues that can arise with user created folksonomies. For instance, because the tags are generated by voluntary users, there is no control for what tags are used (Hilera, et al. 2010). This can cause issues as semantics such as homonyms and homographs are not organized properly as well multiple word phrases may be difficult to search for as well. It seems that any folksonomy created by the public would result in a basic lack of vocabulary control. For instance, another issue that can arise with these types of folksonomies are what is known as “ambiguous labelling” (Font 2013). This occurs when sources are tagged using too personal text words or too common and synonymous words that are not specific enough to create an accurate description. While this is a fairly common problem with folksonomies, as a class it seems as though we were able to maintain at least a regulated group of terms. Furthermore, as MacGregor points out that due to ambiguous tags, precision in results is extremely difficult (2006). He advises that since folksonomies essentially negate the controlled vocabulary aspect of ontologies it is difficult to yield accurate results unless the tags are very specific. The one issue we do seem to have as a class relates to using related and synonymous terms. Due to being allowed to add our own descriptors to each resource there are quite a few entries that contain similar tags that would be synonymous with each other, such as “tagging” and “social tagging”. Both refer to the same act of tagging resources.

In comparison to other folksonomies there is one thing in particular that I noticed immediately. When looking at the list that the class prepared for the Contribute 3 assignment I noticed that many of the sites did not necessarily have preferred search terms that needed to be used in order to yield the results the user was looking for. For instance, the Sports Authority website will pull up multiple items when searching for the term “basketball”. They will automatically pull up items for the team sport of basketball such as the actual equipment needed and they will also retrieve results in the shoe department and clothing department. This is similar to our own folksonomy as we do not have a set of preferred terms that must be used in order to bring up results for sources related to ontologies, tagging, or metadata for instance. Furthermore, this is the same for the website fictionpress.com. In this website allows for stories to be searched by any word that is in the title. This can be extremely problematic as there could be multiple titles that contain the same string of text words and will therefore be retrieved during the search which can lead the user to have to search through extraneous results.

While our folksonomy does not use preferred search terms, there are some that do require that type of searching. For instance, the CINAHL (Cumulative Index of Nursing and Allied Health Literature) does require particular text words to be used in order to achieve accurate and successful search results. As an example, when searching for articles, case studies, or information relating to diseases such as cancer, there will not be any results returned as “cancer” is not considered a valid search term. In order to find results relating to this particular subject, you need to search with the term “neoplasm”. This is the only way that the CINAHL will yield articles, journals, and documents relating to cancer.

Overall, creating our own folksonomy was an excellent tool to learning how metadata is organized with vocabulary control. Creating our own folksonomy has shown why it is so important to establish vocabulary control as we have seen that with synonyms and ambiguous tags it is difficult to find the results that you are searching for. As Fidel points out, while vocabulary control can be labor-intensive, without some type of designated text words to search by there is no way to yield precise and accurate results (1992).

References:

1. Blackboard Class Website: Contribute Three Assignment.
2. Fidel, R. (1992). Who needs controlled vocabulary? *Special libraries*. 83(1).
3. Font, F. (2013). Folksonomy based tag recommendation for collaborative tagging systems. *International journal on semantic web and information systems*. 9(2). DOI:
<http://dx.doi.org.proxy.lib.fsu.edu/10.4018/jswis.2013040101>
4. Hilera, J. (2010). An evolutive process to convert glossaries into ontologies. *Information technology and libraries*. <http://www.ala.org.proxy.lib.fsu.edu/>
5. Kim, H. (2009). Representing and sharing folksonomies. *Journal of information science*. 36(57).
6. Macgregor, G. (2006) Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*. 55(5), 291 – 300.
7. Morrison, J. (2008). Tagging and searching: search retrieval effectiveness of folksonomies on the world wide web. *Information Processing and Management*. 44(4), 1562 – 1579.
8. Sports Authority Website: <http://www.sportsauthority.com/home/index.jsp>
9. Taylor, A. (2009). The organization of information.