Bringing a More Accurate User's Perspective into Web Navigation: Facet Analysis of Folksonomy Tags

Yunseon Choi

Graduate School of Library and Information Science, University of Illinois at Urbana Champaign

501 E. Daniel Street, MC-493

Champaign, IL 61820-6211, USA

(voice): 1-217-333-3280

ychoi10@illinois.edu

ABSTRACT

Faceted navigation is useful for finding information on the Web, but facets created by professionals do not represent users' preferences and understandings of concepts. To make faceted navigation more suitable for users and fulfill users' needs, a user's perspective should be explicitly represented in the design of browsing interface. This study performed facet analysis on folksonomy tags to identify user-centric facets.

Categories and Subject Descriptors

H. 3. [**Information Storage and Retrieval**]: Content Analysis and Indexing, Linguistic processing, Thesauruses

General Terms

Design, Human Factors, Languages, Theory

Keywords

Faceted classification, Facet analysis, Faceted navigation, Facets, Folksonomy, Tags, Web directories, Web navigation, Web browsing, Web searching interface.

1. INTRODUCTION

Effective searching and navigation of web resources is at the forefront of issues related to the area of information organization. Faceted classification has become widely used to organize resources on the Web. Growing interests in the application of faceted classification on the design of websites have led to successfully providing users with findability of resources. However, the faceted structures of faceted classification were constructed by professionals, not by users. Although professionally developed facet structures are consistent and systematic, sometimes it is not easy for users to understand the concepts of facets and their relationships. Those facets are mainly based on controlled vocabulary and the priority of facets is not represented by users' preferences, leading to difficulty in finding appropriate facets to their needs. To make faceted navigation more suitable for users, users' perspectives should be explicitly represented in designing a browsing interface.

Folksonomy expresses users' preferences and needs since it allows users to add their own tags based on their interests. Several researchers have studied folksonomy and its usefulness for classification or retrieval [1][2][3][4], but no research has considered the use of its tags as "facets" in designing a web-searching interface. Despite criticism that folksonomy tags are ambiguous and uncontrolled terminology [5][6], in effect, this very problem could play a significant role in reflecting real users' views and their vocabulary. The purpose of this research is thus to bring a more accurate user's perspective into the design of the faceted navigation.

2. BACKGROUND
2.1 Faceted Classification and the Web

2.1.1 Faceted Classification

Faceted classification, which is based on the Colon Classification scheme developed by Ranganathan [7], is an "*analytico-synthetic*" scheme deriving from two processes: *analysis* (the process of breaking down subjects into their elemental concepts) and *synthesis* (the process of recombining those concepts into subject strings). Taylor [8] defines "facets" as: "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties, or characteristics of a class or specific subject." Ranganathan [7] originally postulated five fundamental categories: Personality, Matter, Energy, Space, and Time (see Table 1).

Table 1. Ranganathan's five fundamental categories

Five fundamental categories	
Personality	-can be understood as the primary facet. century, decade
Matter	-material, property, method
Energy	-action
Space	-continents, countries, counties
Time	-time period, the least difficulty in identification

2.1.2 Faceted Classification for Organizing Web Resources

There have been numerous studies regarding the application of faceted classification on organizing web resources. Priss and Jacob [9] discuss the adaptability and flexibility of a faceted thesaurus for a better website design. La Barre [10] has studied websites using faceted classification. Additionally, Flamenco project has investigated the usefulness of hierarchical faceted interface to help users' web searching [11].

However, the faceted structures of faceted classification were constructed by professionals, not by users. Hence, the concepts and relationships of facets cannot be easily grasped, so users have difficulty for finding appropriate facets. Although faceted structures created by professionals are consistent and systematic, sometimes users might get lost in browsing the faceted structures because those facets are mostly based on controlled vocabulary which users are not familiar with. The other problem with those faceted structures is that the most interesting objects to users are not represented in top-level categories in faceted hierarchy[1], i.e., the priority of facets is not based on users' preferences. Therefore, to make the faceted web navigation more suitable for users, users' perspectives should be explicitly represented in facets.

2.2 Folksonomy as User-centric Resource Organization

The term, "folksonomy" was coined by Thomas Vander Wal in 2004. He describes "folksonomy" as "user-created bottom-up categorical structure development with an emergent thesaurus" [14]. Flickr, Del.icio.us, and LibraryThing are popular folksonomy sites. The main characteristic of folksonomy is to allow users to create and browse their own tags as well as tags created by other users.

3. METHODOLOGY

This research employed the method of examining folksonomy tags to identify user-centric facets reflecting a user's point of view. An exciting and important aspect of this research is to provide a new angle for understanding folksonomy tags by considering them as "facets." Tags express properties of a category, and this characteristic of tags corresponds to that of "facets," in that a facet describes "properties or characteristics of a class or specific subject" [8]. For example, tags "school" or "science" are facets that represent the properties of a category for kids. Aura [15] also notes that a tag is labeled by "its characteristic attributes of features, not by the categories to which it belongs."

To identify user-centric facets, it is helpful to investigate how a professional's point of view is different from a user's point of view in terms of describing a domain. For the former, category

---

[1] Facets can be either "*flat* (containing a single level of values) or *hierarchical* (containing multiple levels of values in an ancestor-descendant structure)." [12] Also, it is noticeable that faceted navigation systems allow users to navigate resources *hierarchically* from a super category to its sub-categories [13].

labels of in web directories were examined, and for the latter, tags from a folksonomy were analyzed. Web directories are simple taxonomies to organize web resources in a hierarchy. Taxonomy is "a collection of controlled vocabulary terms organized into a hierarchical structure" [16]. Web directories create the taxonomy of categories named by indexers' languages, i.e., controlled vocabulary.

Category labels used in web directories can be regarded as potential "facets", in that each label represents the properties of a unique domain. Glassel [17] and Hearst [18] also note that the Yahoo directory uses "faceted metadata" in their top level directory. The Yahoo directory combines the Regional category with other hierarchical facets. For example, finding one starting with the Education category will be guided to browse the hierarchy under the Regional category, and as a result, the users will get the information. The category paths below indicate how the University of California, Berkley can be located in both Education and Regional categories [18]:

- Education > College and University > Colleges and Universities >United States > U > University of California > Campuses > Berkeley
- Regional > U.S. States > California > Cities >Berkeley > Education > College and University > Public > UC Berkeley

For the comparison of web directory categories and folksonomy tags, facet categories from Aitchison et al. [19] were used, because they have been extended to include different subject areas, and they are more explanatory than Ranganathan's categories [19].

3.1 Analysis of Categories in Web Directories

For a professional's perspective, I collected category labels from two popular web directories, i.e., the Open Directory Project (ODP) and Yahoo. I selected one domain, "kids." The characteristic of domain "kids" is that its user population includes parents as well as kids. ODP has a category for kids named "Kids and Teens" (http://www.dmoz.org/Kids_and_Teens/) in second-level categories, and Yahoo manages a separate directory for kids, Yahoo!Kids (http://kids.yahoo.com/directory). I classified all category labels by Aitchison et al.'s facets [19] which consist of five main categories (i.e., Entities/things/objects, Attributes, Actions/activities, Space/place/location/environment, and Time) and their sub-categories (see Table 2).

Table 2. Analysis of labels of categories from two web directories

Aitchison et al.'s categories	ODP: Kids and Teen category	Yahoo!Kids directory
Entities/things/objects		
(By characteristics)		
Abstract entities (e.g., topics)	Arts, Directories, Entertainment, Health, News, School Time, Science	Around the world, Arts, Entertainments, School Bell
Naturally occurring entities	Nature, Environment	-

Living entities, organisms	-	-
Artefacts (e.g., man-made)	-	-
Materials	-	-
Parts/components	-	-
Whole entities/complex entities	Your family	-
(By function)		
- Agents (performers of action)		
Individuals, personnel, organisations involved	People and Society	-
Equipment /apparatus for operation	Computers	Computers and Online
- Patients (recipients of action)	-	-
- End-products of process or operation	-	-
Attributes		
Properties/qualities, states/conditions (e.g., age, gender)	Pre-school, Teen-life	-
Actions/activities		
Processes/functions (internal processes, intransitive actions, operations)	Games, Sports and Hobbies	Games, Sports and Recreation
Space/place/ location/ environment	International	-
Time	-	-

3.2 Identification of User-centric Facets Based on Folksonomy Tags

To identify user-centric facets, I collected folksonomy tags. I chose one folksonomy site, Del.icio.us.com, and a total of 394 tags among "Related Tags" about the domain "kids" were randomly collected and examined. Finally, 18 main tags were selected and arranged by Aitchison et al.'s facets [19]. Table 3 shows the comparison of web directory categories and folksonomy tags.

Table 3. Comparison of web directory categories and folksonomy tags

Aitchison et al.'s categories	**Kids categories from directories**	**Kids-related tags from folksonomy**
Entities/things/objects		
(By characteristics)		
Abstract entities (e.g., topics)	Arts, Around the world, Directories, Entertainment,	Books, Health, Parenting, References,

	Health, News, School Bell, School Time, Science,	Science, School, Shopping, Tutorial
Naturally occurring entities	Nature, Environment	-
Whole entities/complex entities	Your family	-
(By function)		
- Agents (performers of action)		
Individuals, personnel, organisations involved	People and Society	-
Equipment/ apparatus for operation	Computers and Online	Computer
Attributes		
Properties/ qualities, states/ conditions (e.g., age, gender)	Pre-school, Teen-life	Baby, Toddler
Actions/activities		
Processes/ functions (internal processes, intransitive actions, operations	Games, Hobbies, Sports, Recreation	21st century skills, Activities, Digital literacy, e-Learning, Party, Writing&Reading
Space/place/ location/ environment	International	-

4. DISCUSSION & FUTURE RESEARCH

Table 2 demonstrates similarity with slight variation in labeling categories between two different web directories, in terms of the same facets. Some labels were exactly the same in both directories. For example, "Arts," "Computer," and "Entertainment" for the Entities/things/objects facet and "Game" and "Sports" for the Action/activities facet were identically named. Other labels were slightly different for the same facets. For instance, both "School Time" and "School Bell" labels described the same property, i.e., "school subjects" including math, languages, science, and social studies. However, "School Time" and "School Bell" labels required looking into all their sub-categories to grasp the exact meaning due to their metaphorical expression. Unlike these labels, the folksonomy tag was named just "School" (see Table 3). It can be explained that folksonomy tags do not need to be unnecessarily long. Also, it can be implied that ambiguous or metaphorical terms as category labels might be redundant from users' perspectives, so rather simple but clear terms would be better for users to understand facets.

Table 3 indicates that there was a large difference between web directory categories and folksonomy tags regarding most of facet labels, e.g., Entities/things/objects, Attributes, and Action/activities facets. The folksonomy tags presented have two significant implications in the design of a user-centric faceted interface. First, those tags can be used to decide the priority of facets in the structure. For example, it is suggested that tags "Books," "Parenting," "Shopping," "Tutorial" (for Entities/things/objects facet) and "Party," "Activities" (for Processes/functions facet) could be used as main facets (e.g., top-level categories) directly reflecting users' preferences. Especially, "Parenting" tag proves that users of kids-category are not only kids but also their parents, and confirms that folksonomy tags correctly explain user population. Second, tags "21st century skills," "Digital literacy," and "e-Learning" illustrate users' tendency toward up-to-date terminology by implying the growth of interests in fast-changing technology.

Future research will be conducted with additional websites using faceted navigation, and be complemented by statistical analysis based the number of tags. Furthermore, the semantic and syntactic relationships among tags can be identified to create a more rigorous facet structure.

5. CONCLUSION

The faceted navigation has been effective for organizing resources in websites, but their faceted structures have been mainly constructed by professionals, not by users. To make the web navigation suitable for users, users' views should be reflected in facets. The main contribution of this research is bringing a more accurate user's perspective into the design of a faceted browsing interface. Its unique contribution is the examination of folksonomy tags as user-centric facets.

6. REFERENCES

[1] Dye, J. 2006. Folksonomy: A game of high-tech (and high-stakes) tag. *E-Content Magazine, 29*(3), 38-43.

[2] Gruber, T. 2007 Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web & Information Systems* 3(2). Retrieved April 14, 2008, from http://tomgruber.org/writing/ontology-of-folksonomy.htm

[3] Mathes, A. 2005. *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*: Retrieved from, http://www.adammathes.com/academic/computer-mediated- communication/folksonomies.html

[4] Peterson, E. 2006. Beneath the metadata: Some philosophical problems with folksonomy. *D-Lib Magazine* 12(11).

[5] Etches-Johnson, A. 2006. The Brave New World of Social Bookmarking:Everything You Always Wanted to Know but Were Too Afraid to Ask, *Feliciter*, Issue #2, 2006, Canadian Library Association

[6] Shirky, C. 2005. "Ontology Is Over-Rated: Categories, Links and Tags", from http://www.sims.monash.edu.au/subjects/ims2603/resources/week7/7.1.pdf

[7] Ranganathan, S. R. 1967. *A descriptive account of the colon classification*. Bombay: New York: Asia Publishing House. Reprinted by Sarada Ranganathan Endowment for Library Science.

[8] Taylor, Arlene G. 1992. Introduction to Cataloging and Classification. 8th ed. Englewood, Colorado: Libraries Unlimited, 1992.

[9] Priss, Uta and Jacob, Elin K. 1999. "Utilizing Faceted Structures for Information Systems Design." Proceedings of the 62st Annual Meeting of ASIS, 1999, p. 203-212.

[10] La Barre, K. 2004. "Adventures in faceted classification: a brave new world or a world of confusion?", in McIlwaine, I.C. (Eds), *Knowledge Organization and the Global Information Society, Proceedings of the 8th International Conference of the International Society for Knowledge Organization, University College London, 13-16 July, 2004*, Ergon, Würzburg, Advances in Knowledge Organization, Vol. Vol. 9 pp.79-84.

[11] Hearst, M. A. 2006. Design Recommendations for Hierarchical Faceted Search Interfaces. ACM SIGIR Workshop on Faceted Search, August 2006.

[12] Yee, P., K. Swearingen, K. Li, and M. Hearst. 2003. Faceted metadata for image search and browsing. In *Proceedings of the 2003 Conference on Human Factors in Computing Systems (CHI'03)* (401–408). New York, NY: ACM Press.

[13] Howard, olga. 2007. Web facet analysis, IA summit 2007, from http://www.slideshare.net/olgahow/web-facet-analysis

[14] Vander Wal, Thomas. 2007. Folksonomy, from http://www.vanderwal.net/folksonomy.html

[15] Aura, 2007. Communications & Strategies, no. 65, 1st quarter 2007, p 69) p. 67-89.

[16] National Information Standards Organization (NISO), 2005. Guidelines for the construction of controlled vocabularies, ANSI/NISO z39.19-2005

[17] Glassel, Aimee. 1998. Was Ranganathan a Yahoo!? , from http://scout.wisc.edu/Projects/PastProjects/toolkit/enduser/archive/1998/euc-9803.html

[18] Hearst, M. A. 2003. Using Words to Search a Thousand Images: Hierarchical Faceted Metadata in Search & Browsing. Seminar on People, Computers, and Design, Standford University.

[19] Aitchison, J., Gilchrist, A., and Bawden, D. 2000. Thesaurus construction and use: practical manual. Fitzroy Dearborn, London