HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead

Cameron Marlow[1], Mor Naaman[1], danah boyd[1,2], Marc Davis[1,2]

[1]Yahoo! Research Berkeley
1950 University Avenue, Suite 200
Berkeley, CA 94704-1024

[2]UC Berkeley School of Information
102 South Hall
Berkeley, CA 94720-4600

{cameronm, mor, danah, marcd}@yahoo-inc.com

ABSTRACT

In recent years, *tagging systems* have become increasingly popular. These systems enable users to add keywords (i.e., "tags") to Internet resources (e.g., web pages, images, videos) without relying on a controlled vocabulary. Tagging systems have the potential to improve search, spam detection, reputation systems, and personal organization while introducing new modalities of social communication and opportunities for data mining. This potential is largely due to the social structure that underlies many of the current systems.

Despite the rapid expansion of applications that support tagging of resources, tagging systems are still not well studied or understood. In this paper, we provide a short description of the academic related work to date. We offer a model of tagging systems, specifically in the context of web-based systems, to help us illustrate the possible benefits of these tools. Since many such systems already exist, we provide a taxonomy of tagging systems to help inform their analysis and design, and thus enable researchers to frame and compare evidence for the sustainability of such systems. We also provide a simple taxonomy of incentives and contribution models to inform potential evaluative frameworks. While this work does not present comprehensive empirical results, we present a preliminary study of the photo-sharing and tagging system Flickr to demonstrate our model and explore some of the issues in one sample system. This analysis helps us outline and motivate possible future directions of research in tagging systems.

Categories and Subject Descriptors

H.1.1 [**Information Systems**]: Models and Principles – *Systems and Information Theory*.

General Terms

Algorithms, Design, Human Factors.

Keywords

Tagging systems, taxonomy, folksonomy, tagsonomy, Flickr, categorization, classification, social networks, social software, models, incentives, research.

1. INTRODUCTION

Web-based *tagging systems* such as Del.icio.us, Technorati and Flickr allow participants to annotate a particular *resource*, such as a web page, a blog post, an image, a physical location, or just about any imaginable object with a freely chosen set of keywords ("tags"). In this paper, we aim to articulate a framework for studies of such systems.

One approach to tagging has emerged in "social bookmarking" tools where the act of tagging a resource is similar to categorizing personal bookmarks. In this model, tags allow users to store and collect resources and retrieve them using the tags applied. Similar keyword-based systems have existed in web browsers, photo repository applications, and other collection management systems for many years; however, these tools have recently increased in popularity as elements of social interaction have been introduced, connecting individual bookmarking activities to a rich network of shared tags, resources, and users.

Social tagging systems, as we refer to them, allow users to share their tags for particular resources. In addition, each tag serves as a link to additional resources tagged the same way by others. Because of their lack of predefined taxonomic structure, social tagging systems rely on shared and emergent social structures and behaviors, as well as related conceptual and linguistic structures of the user community. Based on this observation, the popular tags in social tagging systems have recently been termed *folksonomy* [22], a folk taxonomy of important and emerging concepts within the user group.

Social tagging systems may afford multiple added benefits. For instance, a shared pool of tagged resources enhances the metadata for all users, potentially distributing the workload for metadata creation amongst many contributors. These systems may offer a way to overcome the Vocabulary Problem – first articulated by George Furnas et al in [8] – where different users use different terms to describe the same things (or actions). This disagreement in vocabulary can lead to missed information or inefficient user interactions. The taxonomy of tagging systems articulated in this paper, and the results of our preliminary experiments on the relationship between tag overlap and social connection, both point to the possibility that thoughtful sociotechnical design of tagging systems may uncover ways to overcome the Vocabulary Problem without requiring either the rigidity and steep learning curve of tightly controlled vocabularies, or the computational complexity and relatively low success of purely automatic approaches to term disambiguation.

Figure 1 shows a conceptual model for social tagging systems. In this model, users assign tags to a specific resource; tags are represented as typed edges connecting users and resources. Resources may be also be connected to each other (e.g., as links between web pages) and users may be associated by a social network, or sets of affiliations (e.g., users that work for the same company).