Understanding the Motivation behind Tagging

Christian Körner
Knowledge Management Institute
Graz University of Technology
Inffeldgasse 21A 8010 Graz
+43 316 873 9298

christian.koerner@tugraz.at

ABSTRACT

Tagging is the process of annotating resources with keywords or terms in order to enable better organization, retrieval and sharing of the resource. Today, a large number of diverse applications exist that allow users to tag their documents. Del.icio.us, for example, enables users to organize their bookmarks and tag them accordingly. Flickr allows users to tag pictures and YouTube enables content creators to assign tags to videos in order to facilitate retrieval. While these systems have increasingly become an object of research for web scientists, little is known about these systems, for example how and why they are used, how behavior on a micro level propagates to phenomenon on the macro level, or we can simulate the dynamics of tagging system's evolution. My research interest lies in understanding and modeling the motivation of users, in particular why and how they are using these systems. A related topic of interest is the modeling of the user behavior with simulators. Modeling and evaluating the habits of users opens doors for deeper empirical studies, such as simulation. This way it might also be possible to study relations and interactions between the behavior of single users and the behavior of larger tagging communities.

Current research mainly utilizes questionnaire-based user studies to learn more about the different motivations of users for tagging (such as [2]). [1] introduce a tagging model that does incorporate background knowledge and previous tag assignments made by users. With the help of this model the authors are able to successfully reproduce selected characteristics found in tagging systems.

I have conducted a broad review of the current literature, which shows that at least two vastly different types of tagging behavior can be found in tagging systems: First, we can identify users who use tags mainly to organize their resources so that they can be found more easily later. We can refer to this type of tagger *categorizer*. These users want to organize their resources with the help of tags and establish their own kind of vocabulary for the annotation process based on their mental model. An example is the tag "MacOSX". A typical categorizer would always stick to the same tag instead of introducing new synonym tags such as "apple" or "mac". Because of the fact that these tags are normally very close to the mental model of a user the tags are facilitators for navigation and browsing.

Another type of tagger can be referred to as a *describer*. This type of user utilizes tags to support the retrieval process by describing the resources they annotate. Describers can be found in tagging systems such as YouTube where content providers tag their resources so that their submitted videos can be found easier. This way a typical describer tries to describe the resource he annotates as much as possible. Therefore tags of describers are often used only once and contain lots of synonyms. In an exemplary tagging vocabulary of such user a lot of tags like "MacOSX", "apple" and "mac" can be found. In addition the vocabulary of a describer is much larger than that of a categorizer who has his own stable vocabulary. Because the content of the tags is very close to the content of the annotated resources, these tags support the process of searching.

The goal of my research is the study and identification of these two tagging types on social tagging platforms. Therefore I need to develop metrics and simulators that are able to describe and illustrate the observed behavior.

A prerequisite for addressing these goals is the identification and acquisition of relevant tagging datasets and the identification of collections of users who can serve as extreme examples of the identified types.

In the current status of my research, the following data sets were used to conduct exemplary experiments:

- Del.icio.us – A social bookmarking platform that allows the tagging of URLs.
- Bibsonomy – A web application that helps users organizing publications and resources.
- Flickr – One of the biggest sites that enables users to store and build their photo collection.
- ESP game – An online game that tries to overcome the problem of the lack of picture metadata by using human annotators.

The dataset of the ESP game was used to establish the notion of a "perfect describer" because players of this game need to find the right keywords for pictures they are shown. To identify representatives of "perfect organizers", the "set" structure of a Flickr user was used. Because one photo is typically only stored in one set this characteristic mimics the behavior of a categorizer. In other datasets, we would expect that "real behavior" would lie between the two extremes.

In the current status of the work some metrics for separating categorizers from describers have been identified. Such characteristics include:

- The number of tags that were only used once by a user. This is an indicator that the given person is a describer.

- The growth of the tagging vocabulary. The faster the vocabulary increases the more likely it is that the person is a describer. Figure 1 shows the trends of ten tagging vocabularies. The red line (ESP game) depicts a perfect describer.
- The entropy of tags. A categorizer tends to get entropy that is as low as possible because he tries to "encode" his resources in a good and balanced way.
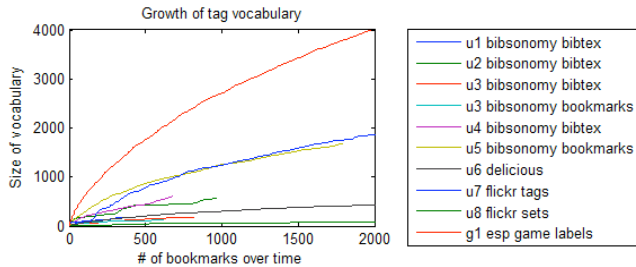
Figure 1 - Growth of Tagging Vocabulary

One important aspect of my research is to develop a model that produces a score describing how likely a given user is a describer or categorizer based on a user's individual tagging history. This indicator could range from -1 to +1, where -1 displays an absolute describer and +1 indicates a total categorizer. The next step is to begin work on simulators that might enable to imitate the behavior of taggers in social tagging platforms. In the future, my work might enable researchers interested in social tagging systems to simulate whole social tagging platforms based on individual user characteristics with this approach.

Advancing the state of the art of research in this area is important for a number of reasons: Recommender systems that are aware of these characteristics might have an advantage over tag recommendation engines that focus on a simple attribute such as e.g. "five top used tags of a given resource". The anticipated contribution of this work is to introduce a model for tagging behavior that is applicable to a broad range of tagging systems independent of the resources that are being tagged (images, textual documents, etc). This model should allow getting hold on to what extent a certain user can be understood as a describer or a categorizer. In addition, it might help to set up simulators that emulate the dynamics of social tagging systems on a macro scale.

1. SELECTED REFERENCES

[1] Dellschaft, K. & Staab, S. (2008), An epistemic dynamic model for tagging systems, *in* 'HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia', ACM, New York, NY, USA, pp. 71--80.

[2] Heckner, M.; Heilemann, M. & Wolff, C. (2009), Personal Information Management vs. Resource Sharing: Towards a Model of Information Behaviour in Social Tagging Systems, *in* 'Int'l AAAI Conference on Weblogs and Social Media (ICWSM)'.

2. ADDITIONAL INFORMATION

Academic Advisor: Markus Strohmaier

Christian Körner is a graduate student.