Chapter III
Analyzing Communal Tag Relationships for Enhanced Navigation and User Modeling

Edwin Simpson
HP Labs, UK

Mark H. Butler
HP Labs, UK

ABSTRACT

The increasing amount of available information has created a demand for better, more automated methods of finding and organizing different types of information resources. This chapter investigates methods for enabling improved navigation, user modeling, and personalization using collaboratively generated tags. The authors discuss the advantages and limitations of tags, and describe how relationships between tags can be used to discover latent structures that can automatically organize a collection of tags owned by a community. They give a hierarchical clustering algorithm for extracting latent structure and explain methods for determining tag specificity, then use visualization to examine latent structures. Finally the authors discuss future trends including using latent tag structures to create user models.

INTRODUCTION

Most current methods of accessing information do not incorporate a user’s personal perspective or current task, even though such context significantly affects their information needs. As a result, users are often overwhelmed by the number of information resources that potentially match their requirements. This problem can be reduced by giving users a clearer overview of the available resources, or by the system making better use of data about the user. Specifically, if a system is able to model a user’s perspective, current tasks and diverse information resources, it can push relevant
information to users and improve navigation by automatically filtering and organizing resources for the user (Belkin, 1992).

In order to do this, a system requires a model that describes information resources, users and their context. An ideal model should be applicable across heterogeneous information sources, capture the user’s interests, their interactions with information, and describe their current tasks.

One starting point for such a model could come from the simple and flexible descriptors used in information retrieval systems called tags. Tags are in widespread use by systems such as photo, bookmark and video management applications on the Web, and a collection of tags used by a particular system is known as a folksonomy. However, because these folksonomies lack the explicit structure of taxonomies, they suffer from problems with synonyms, polysemy, depth of specificity and scalability. Research has shown that there are various hidden structures in folksonomies that can help overcome these problems (Mika, 2005; Diederich & Iofciu, 2006). This chapter explores how we can find relationships between tags, and hence expose the latent structure in folksonomies by extracting topic clusters and hierarchies. We then describe how these structures could be used to improve navigational interfaces and model user interests.

This chapter aims to inform the reader of various techniques that use tags to automatically organize resources and model the interests of users in a community. Readers will be able to apply this set of techniques to help derive greater value from large communities and repositories of information, by improving the way users explore and access resources.

In the second section, Background, we explain current uses of tagging and alternative methods for organizing information. We also define some key terms and concepts.

Then, in the third section, Extracting Latent Structure from Folksonomies, we describe the latent structure present in folksonomies and explain tag similarity methods that may be used to detect this structure. We present some different methods of representing latent structure, focusing on one method, clustering, in more detail. We then consider how we can use latent structure to represent general concepts in the folksonomy.

In the fourth section, Visualization of Latent Structures, we describe how to visualize the resulting latent structure and analyze the results of applying our clustering method to two datasets.

In the fifth section, Future Trends, we suggest the development of dynamic and hierarchical visualizations and discuss the possibility of modeling user interests and current tasks using information derived from folksonomies.

**BACKGROUND**

The Web is notable for its increasing abundance of information and the freedom it allows for new innovation, particularly the development of new methods to connect users with information and allow them to organize resources. These methods augment more established approaches such as free-text search (Gudivada et al., 1997; Page et al., 1998), directories and taxonomies (Garshol, 2004), explicit communication between users and subscriptions to news feeds or mailing lists. In this section we describe folksonomies and consider both the advantages and limitations of this approach. We then compare tagging with thesauri and controlled vocabularies, and consider how these techniques can be combined with tagging.

**Tagging**

As already mentioned, one idea for improving information access which is now in widespread use is *tagging*. Tagging is a process where users add arbitrary labels, known as *tags*, to information resources.