Chapter 15
Social Tagging: Properties and Applications

Yong Yu
Shanghai Jiao Tong University, China

Rui Li
Shanghai Jiao Tong University, China

Shenghua Bao
Shanghai Jiao Tong University, China

Ben Fei
IBM China Research Lab, China

Zhong Su
IBM China Research Lab, China

ABSTRACT

Recently, collaborative tagging Web sites such as Del.icio.us and Flickr have achieved great success. This chapter is concerned with the problem of social tagging analysis and mining. More specifically, we discuss five properties of social tagging and their applications: 1) keyword property, which means social annotations serve as human selected keywords for Web resources; 2) semantic property, which indicates semantic relations among tags and Web resources; 3) hierarchical property, which means that hierarchical structure can be derived from the flat social tagging space; 4) quality property, which means that Web resources' qualities are varied and can be quantified using social tagging; 5) distribution property, which indicates the distribution of frequencies of social tags usually converges to a power-law distribution. These properties are the most principle characteristics, which have been popularly discussed and explored in many applications. As a case study, we show how to improve the social resource browsing by applying the five properties of social tags.

INTRODUCTION

With the rapid development of new technologies, both ordinary users and service providers are experiencing the coming wave of the next-generation Web. As a representative, tagging based services have achieved a significant success. Services like Del.icio.us (http://del.icio.us), Flickr (http://www.flickr.com), and Technorati (http://Technorati.com),
Social Tagging

enable users to annotate and categorize Web resources, e.g., Web pages, photos and blogs, with freely chosen words. Taking the famous social bookmarking service, Del.icio.us, as an example, the service allows users to collect and annotate Web pages with one-word descriptors, which are also known as social tags or social annotations (in this chapter, we use the terms “annotation” and “tag” interchangeably). The social annotations assigned to bookmarks can help users organize their collected Web pages. Social annotations are a little bit like keywords or categories, but they are chosen by the users, and they do not form a hierarchy. Users can assign as many tags to a bookmark as they like and rename or delete the tags later. So, tagging can be much easier and more flexible than fitting users’ information into predefined categories or folders. In 2004, Thomas Vander Wal named these services “Folksonomy”, which came from the terms “folk” and “taxonomy” (Smith, 2004).

Social annotations from tagging based Web sites are increasing at an incredible speed. Millions of Web users with different backgrounds are using these services to annotate their favorite Web resources. For example, Del.icio.us has more than 1 million registered users soon after its third birthday, and the number of Del.icio.us users has increased by more than 200% in the past nine months (Bao et al, 2007). Mathes (2004) attributes the success of these services to the following reasons:

1) Low Barriers to Entry: The freely chosen keywords enable users - not just professionals - to participate in the system immediately without any training or prior knowledge. Additionally, annotating Web resources is easy in terms of time, effort and cognitive costs.

2) Feedback and Asymmetric Communication: Feedback is immediate, which leads to a form of asymmetrical communication between users through metadata. The users of a system are negotiating the meanings of the terms through their individual choices of tags to describe documents for themselves.

3) Individual and Community Aspects: Individuals have an incentive to tag their materials with terms that will help them organize their collections in a way that they can find these items later. The services are designed to share materials. Users can contribute to the system and other users by sharing the tags and associated resources.

4) Unanticipated Uses: While most tags developed at Flickr and Del.icio.us have a concrete focus on subject categorization, there are tags being used in some unexpected, interesting ways that reflect communication and ad-group formation facilitated through metadata.

The large amount social annotations are not only simple tags for organizing contents but also are useful in sharing information within a social environment. The power of social annotations is that aggregation of information provided by group of users form a social distributed environment. In this chapter, we will summarize current research on social tagging services and discuss how to aggregate the knowledge from social annotations. Specifically, we present a detailed analysis of five main properties of social annotations:

1) Keywords property: tags, used to organize personal collected Web resources, often describe and summarize the content and usage of Web resources perfectly.

2) Semantic property: semantics of each tag can be generated from the associations between tags and Web resources, since similar tags are usually used to annotate similar resources.

3) Hierarchical property: although tags are flat keywords associated with corresponding resources for personal use, the hierarchical