Chapter 7

Exploring Semantic Tagging with Tilkut

Sari Vainikainen
VTT Technical Research Centre of Finland, Finland

Pirjo Näkki
VTT Technical Research Centre of Finland, Finland

Asta Bäck
VTT Technical Research Centre of Finland, Finland

ABSTRACT

Social bookmarking is one of the earliest examples of social media services. In bookmarking services there are two main approaches for adding metadata: user-generated freely chosen tags and keywords based on taxonomies or semantic ontology. This chapter presents a social bookmarking service Tilkut that combines the benefits of both of these approaches. Tilkut utilizes both freely defined tags and semantic tag suggestions based on predefined ontology. This chapter describes two different versions of the service and user experiences from a small scale user study and long-term test use in real context. Work related knowledge sharing was selected as a primary use case for the second version. The results from the first user studies were used as the starting point when developing the second version of Tilkut. A survey and workshop were organised to get more information of the requirements for enterprise use. In this chapter, we explain our approaches to adding semantics to social bookmarking, present the experiences, and discuss future research directions.

INTRODUCTION

There are several online bookmarking and clipping applications available that support storage, sharing and retrieval of web links. Delicious.com, Connotea and Clipmarks are examples of such social bookmarking applications. In these services bookmarks and image clips from web pages are typically organized using tags – user-defined keywords describing the content of the bookmark.

Tagging is an easy way to add metadata to web content compared to predefined and formal taxonomies. The main advantage of tags, the free and bottom-up emergence of vocabularies, can
Exploring Semantic Tagging with Tilkut

also become a disadvantage. An active tagger is quickly faced with a large and messy tagcloud, which does not support finding content so well any more. Our goal was to offer extra support for users to manage their tags and to find relevant content.

Tagging can be enhanced semantically by utilising ontology. The purpose of increased semantics is to support managing and utilising tags and tagged content without losing the ease and power of user-defined keywords.

We explored the usefulness of semantic tagging by developing a social bookmarking prototype and testing it with potential users. In this paper we present the results from the user tests and discuss the possibilities for combining user-generated tags and well-structured ontology. Semantic bookmarking could be especially beneficial in work-related knowledge sharing, and therefore have explored the opportunities for adapting the Tilkut application for that purpose.

BACKGROUND

Challenges with Tagging

A lot of the success of tags and tagging can be attributed to the freedom of being able to use any word as a tag. Tagging is typically flat: all tags are at the same level and any number (or a high number) of tags can be applied to a resource. This has some drawbacks for utilising them even for the users themselves and more so for applications that aim at utilising this information automatically.

Well-known and frequent challenges with tags are that people use different words to describe the same thing, or a word has several different meanings (polysemy). People may also describe things at various levels of detail – an expert in a subject will use more detailed and specific words, whereas others use more general words. Also different forms of the same word (sinigers, plurals, typos) exist. (Golder, 2006)

In addition to differences in vocabularies there are also differences between people in how they tag and why they tag. Also, applications have different restrictions and support to tags, which naturally affects the user behaviour.

There are several research papers (Golder, 2006; Maala, 2007; Marlow, 2006; Xu 2006) that report studies of the type of tags people use. In these papers the work has been based mostly on Delicious or Flickr tags. In Delicious, the following tag categories have been identified: topic, type of referenced resource, proper name (person, company, product, event, and location), subjective tags (adjectives, ratings), self reference, toDo tags and time (Golder, 2006). In Flickr photo tags categories include place, time, event, name, action and camera (Maala, 2007). The results of these studies were used when defining the tag categories for our prototype.

When the aim is to utilise tags, different types of tags give different opportunities. Topics (like travel, semanticweb, cat, cars) can be used for analyzing users’ interests as well as characteristics of the tagged resource. Proper names can be used as an indication of interests as well, particularly when additional information related to them can be found on the web. In our approach, we developed methods for automatic analysis of tag categories and methods for adding semantics to different type of tags. The aim is to use this additional metadata for finding and combining similar resources.

Adding Semantics to Tags

Semantics can be utilised at two different levels. The first level is the tag level. Semantic knowledge bases like Geonames, KOKO, Freebase and DBpedia that offer application programming interfaces (APIs) and widgets to developers have become available, and they can be used to add semantics to tags. There are two alternatives to adding semantics to tags. The first alternative is to help users to use tags that have a semantic meaning is a semantic knowledge base. With the