

Chapter XXV

On the Effectiveness of Social Tagging for Resource Discovery

Dion Hoe-Lian Goh

Nanyang Technological University, Singapore

Khasfariyati Razikin

Nanyang Technological University, Singapore

Alton Y. K. Chua

Nanyang Technological University, Singapore

Chei Sian Lee

Nanyang Technological University, Singapore

Schubert Foo

Nanyang Technological University, Singapore

Jin-Cheon Na

Nanyang Technological University, Singapore

Yin-Leng Theng

Nanyang Technological University, Singapore

ABSTRACT

Social tagging is the process of assigning and sharing among users freely selected terms of resources. This approach enables users to annotate/describe resources, and also allows users to locate new resources through the collective intelligence of other users. Social tagging offers a new avenue for resource discovery as compared to taxonomies and subject directories created by experts. This chapter investigates the effectiveness of tags as resource descriptors and is achieved using text categorization via support vector machines (SVM). Two text categorization experiments were done for this research, and tags and Web pages from del.icio.us were used. The first study concentrated on the use of terms as its features while the second used both terms and its tags as part of its feature set. The experiments yielded a macroaveraged precision, recall, and F-measure scores of 52.66%, 54.86%, and 52.05%, respectively. In terms of microaveraged values, the experiments obtained 64.76% for precision, 54.40% for recall, and 59.14% for F-measure. The results suggest that the tags were not always reliable indicators of the resource contents. At the same time, the results from the terms-only experiment were better compared to the experiment with both terms and tags. Implications of our work and opportunities for future work are also discussed.

INTRODUCTION

The increasing popularity of Web 2.0-based applications has empowered users to create, publish, and share resources on the Web. Such user-generated content may include text (e.g., blogs, wikis), multimedia (e.g., YouTube), and even organization/navigational structures providing personalized access to Web content. The latter includes social bookmarking/tagging systems such as del.icio.us and Connotea.

Social tagging systems allow users to annotate links to useful Web resources by assigning keywords (tags) and possibly other metadata, facilitating their future access (Macgregor & McCulloch, 2006). These tags may further be shared by other users of the social tagging system, in effect creating a community where users can create and share tags pointing to useful Web resources. Put differently, tags function both as content organizers and discoverers. Users create and assign tags to a useful resource they come across so that it would be easy for them to retrieve that resource at a later date. At the same time, other users can use one or more of these tags created to find the resource. The same tags may also be used to discover other related and relevant resources. In addition, through tags, a user can potentially locate like-minded users who hold interests in similarly-themed resources, leading to the creation of social networks (Marlow, Naaman, Boyd, & Davis, 2006).

Social tagging provides an alternative means of organizing resources when compared with conventional methods of categorization based on taxonomies, controlled vocabularies, faceted classification, and ontologies. Conventional methods require experts with domain knowledge and this often translates to a high cost of implementing such systems. They are also bound strictly by rules to ensure their classification schemes remain consistent (Morville, 2005). As the system becomes larger, the rules tend to be more complicated, leading to possible maintenance and accessibil-

ity issues. In contrast, the classification scheme in social tagging systems is deregulated. Instead of relying on (a few) experts, they are supported by a (possibly large) community of users. At the same time, tags are “flat,” lacking a predefined taxonomic structure, and their use relies on shared, emergent social structures and behaviors, as well as a common conceptual and linguistic understanding within the community (Marlow et al., 2006). Tags are therefore also known as “folksonomies,” short for “folk taxonomies,” suggesting that they are created by lay users, as opposed to domain experts or information professionals such as librarians, and may in fact be more effective in describing the resource

While social tagging systems have become popular, it is not known if tags created by ordinary users (as opposed to experts) are useful for the discovery of information. A few studies have investigated the use of tags as resource descriptors. Examples include comparing the use of tags against author-assigned index terms in academic papers (Kipp, 2006; Lin, Beaudoin, Bui, & Desai, 2006), examining the ability of tags to classify blogs using text categorization methods (Sun, Suryanto, & Liu, 2007), and investigating the ability of del.icio.us tags to classify Web resources in a small scale study (Razikin, Goh, Cheong, & Ow, 2007). However, to the best of our knowledge, no large scale work has been conducted with del.icio.us, one of the earliest and more popular social tagging sites. The site has a diverse set of tags and Web resources, and its main function is to store, organize, and share bookmarks among a community of users.

The goal of this present chapter is to investigate if tags are useful in helping users to access relevant Web resources. Specifically, we obtain Web pages and their associated tags from del.icio.us and study whether the tags are good descriptors of these resources. Here, we adopt from techniques in text categorization (Sebastiani, 2002) and argue that an effective tag is one in which a classifier is able to assign documents to with high precision

8 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/effectiveness-social-tagging-resource-discovery/19888

Related Content

Open Access to Knowledge and Challenges in Digital Libraries

Yemisi Oluremi Oladapo (2018). *Handbook of Research on Managing Intellectual Property in Digital Libraries* (pp. 260-291).

www.irma-international.org/chapter/open-access-to-knowledge-and-challenges-in-digital-libraries/188553

Software Process Asset Libraries Using Knowledge Repositories

Leonardo Bermón-Angarita, Antonio Amescua-Seco, Maria Isabel Sánchez-Segura and Javier García-Guzmán (2009). *Handbook of Research on Digital Libraries: Design, Development, and Impact* (pp. 465-475).

www.irma-international.org/chapter/software-process-asset-libraries-using/19911

Automation Process in Nigerian Academic Libraries

Janet Onomeh Ubogu (2012). *International Journal of Digital Library Systems* (pp. 40-46).

www.irma-international.org/article/automation-process-in-nigerian-academic-libraries/83501

Developing a Digital Library of Historical Records in Traditional Mongolian Script

Biligsai Khan Batjargal, Garmaabazar Khaltarkhuu, Fuminori Kimura and Akira Maeda (2012). *International Journal of Digital Library Systems* (pp. 33-52).

www.irma-international.org/article/developing-digital-library-historical-records/68816

Availability of Necessary Electronic Infrastructure to Support Open Access Initiative to Literature in Academic Libraries in Delta State, Nigeria

Joseph Chukwusa (2014). *International Journal of Digital Library Systems* (pp. 1-9).

www.irma-international.org/article/availability-of-necessary-electronic-infrastructure-to-support-open-access-initiative-to-literature-in-academic-libraries-in-delta-state-nigeria/105107