Chapter 11

A URI is Worth a Thousand Tags:
From Tagging to Linked Data with MOAT

Alexandre Passant  
National University of Ireland, Ireland

Philippe Laublet  
Université Paris-Sorbonne, France

John G. Breslin  
National University of Ireland, Ireland

Stefan Decker  
National University of Ireland, Ireland

ABSTRACT

Although tagging is a widely accepted practice on the Social Web, it raises various issues like tags ambiguity and heterogeneity, as well as the lack of organization between tags. We believe that Semantic Web technologies can help solve many of these issues, especially considering the use of formal resources from the Web of Data in support of existing tagging systems and practices. In this article, we present the MOAT—Meaning Of A Tag—ontology and framework, which aims to achieve this goal. We will detail some motivations and benefits of the approach, both in an Enterprise 2.0 ecosystem and on the Web. As we will detail, our proposal is twofold: It helps solve the problems mentioned previously, and weaves user-generated content into the Web of Data, making it more efficiently interoperable and retrievable.

INTRODUCTION

The Social Web, or Web 2.0 (O’Reilly, 2005), has become an important trend during the last few years. While end-users of the Web were previously considered as being only consumers of content, the paradigms that the Social Web introduced has led them to become producers as well. For instance, blogs allow anyone to publish and share...
their thoughts on the Web whereas wikis are used to collaboratively build consensual information within a community. In the meantime, social network services have allowed people to define acquaintance networks and to keep in touch with each other on the Web. Moreover, apart from providing a means to create discussions and to define or manage social networks, an important feature of social Web sites is the ability to share content with one’s peers. On many social Web sites, this data can be shared either with whoever is subscribed to (or just browsing) the Web site or else it can be shared within a restricted community. Also, not only textual content can be shared, but various types of media or other content objects: pictures (Flickr), videos (YouTube), slides (Slideshare), trips (Dopplr), and so forth. To make this content more easily discoverable, most of these websites allow users to add free-form keywords, or tags, that act like subjects or categories for anything they wish to share. For example, this article could be tagged with “semanticweb” and “socialweb” on a scientific bibliography management system such as Bibsonomy or Connotea.

Although tags can be generally considered as a type of metadata, since they provide additional information about a tagged item, it is important to keep in mind that they are user-driven. Indeed, while a blog engine may automatically assign a creation date to any blog post or a photo sharing service could use embedded EXIF information to display the aperture of a camera, tags are added voluntarily by users themselves. To that extent, they clearly reflect the needs and the will of the user who assigns the tags. In this way, tags focus on what a user considers as important regarding the way he or she wants to share and present information. The main advantage of tagging for end users is that one can use the keywords that fit exactly with his or her needs and they do not have to learn a pre-defined vocabulary scheme (such as a taxonomy). Tags and tagging actions lead to what is generally called a folksonomy (VanderWal, 2007), an open and user-driven classification scheme that evolves during time thanks to the tagging actions of the community itself, contrary to pre-defined and authoritative classification directories, which are generally fixed.

Yet, in spite of its advantages when annotating content items, tagging leads to various issues regarding information retrieval, which makes the task of retrieving tagged content sometimes quite costly. Mathes (2004) estimates that a “folksonomy represents simultaneously some of the best and worst in the organization of information.” Indeed, even if dedicated algorithms like FolkRank (Hotho, Jäschke, Schmitz, & Stumme, 2006) and clustering techniques can be used to improve retrieval of tagged-content—in spite of the shortcomings we will discuss later—tag-dedicated search engines are generally simply based on plain-text strings, that is, a user types a tag and gets only the content that has been tagged with that particular keyword. Therefore, this can lead to various issues, since such an engine only considers a set of characters that it cannot interpret which consequently introduces some noise and silence issues.

In the Semantic Web domain, the Web of Data is considered a more pragmatic vision of the Semantic Web, focused mainly on exposing data in RDF and interlinking it, that is, providing Linked Data on the Web, rather than on using formal ontologies and inference principles that form the complete Semantic Web vision. Interlinking user-generated content with URIs of well-known and unambiguous resources from the Semantic Web would help to solve the aforementioned issues, as user-generated content would be then interlinked with well-defined and unambiguous identifiers. Moreover, it offers a way to weave such content into the Semantic Web, hence considering Web 2.0 and the Web of Data not as disjoint domains but as being beneficial to each other.

In this article, we describe the MOAT framework that aims to provide an intuitive and lightweight way to bridge this gap between free-tagging and Linked Data, in what we consider a twofold approach with strong benefits for both the