Data & Knowledge Engineering 79–80 (2012) 62–86

Contents lists available at SciVerse ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

A multi-layer framework for personalized social tag-based applications ☆

Barbara Carminati [a], Elena Ferrari [a], Andrea Perego [b,*]

[a] Dipartimento di Scienze Teoriche ed Applicate, Università degli Studi dell'Insubria, Varese, Italy
[b] European Commission — Joint Research Centre, Ispra, Italy

ARTICLE INFO

Article history:
Received 31 May 2010
Received in revised form 25 June 2012
Accepted 25 June 2012
Available online 4 July 2012

Keywords:
Web-based information systems
Social Web
Semantic Web
Social tagging
Personalization

ABSTRACT

Recent years have seen an increasing diffusion of online communities giving their members the ability of specifying and sharing metadata concerning online resources. Such practice, also known as *social* or *collaborative tagging*, has the purpose of collecting and sharing opinions about Web resources and simplifying their retrieval. In this paper, we go one step further and show how tags can have more enhanced applications to be exploited for customizing Web content fruition. More precisely, we propose a multi-layer framework where data collected by social tagging communities are complemented with additional services. Such services provide users the ability of expressing their dis/agreement with existing tags, denoting the members they trust based on their characteristics and relationships, or specifying policies on which "quality" assessment of resources should be returned. Besides providing the formal specification of the proposed framework, we illustrate two case studies we have implemented and the experiments we have carried out in order to verify the feasibility of our approach.

1. Introduction

In recent years, the widespread adoption of Web 2.0 related technologies has greatly facilitated user collaboration and knowledge sharing, bringing several benefits to the field of Web metadata generation and management. Notable examples of Web 2.0 technologies applied to Web metadata are those related to online communities, whose members have the ability of specifying and sharing metadata (referred to as *tags*). These are the cases, for example, of Delicious (http://delicious.com), RawSugar (http://rawsugar.com), Flickr (http://flickr.com), and Last.fm (http://last.fm). Such practice, also known as *social* or *collaborative tagging* [1,2], has the purpose of collecting and sharing opinions about Web resources, and simplifying resource retrieval by organizing them according to a tag-based browsing criterion.

The huge availability of social tagging systems has pushed the development of several applications exploiting these metadata, hereafter called *social tag-based applications*. Recommender systems [3] are notable examples of social-tag based applications. Here, users of online communities share resources that they consider relevant, and express personal opinions on them with the purpose of making resource retrieval easier.

In general, social tag-based applications gather metadata associated with resources, elaborate them and exploit the obtained results to trigger some actions (for example, resource recommendation, classification or filtering). Up to now, research on social tag-based applications has mainly focused on personalized recommendations of tags (e.g., [4–6]) or resources (e.g., [7–10]), which make use of techniques derived from the data mining area to predict which tags/resources might be relevant, and how much.

However, we believe that social tag-based applications can be further improved following several directions. First of all, existing social tag-based applications do not consider much the issue of metadata trustworthiness. As metadata evaluation might trigger crucial actions (for example, those related to resource filtering or access control), it is important to have an accurate estimation

☆ The work reported in this paper is partially funded by the European Community under the QUATRO Plus project (SIP-2006-211001) and by the Italian Ministry of University, Education and Research under the ANONIMO project (PRIN-2007F9437X_004).

* Corresponding author at: European Commission DG JRC, Via E. Fermi, 2749–TP 262, 21027 Ispra, Italy. Tel.: +39 0332786423; fax: +39 0332786325.
E-mail addresses: barbara.carminati@uninsubria.it (B. Carminati), elena.ferrari@uninsubria.it (E. Ferrari), andrea.perego@jrc.ec.europa.eu (A. Perego).

0169-023X/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.datak.2012.06.002

of metadata trustworthiness, so to exclude untrustworthy metadata. As such, a *mechanism able to assess the trustworthiness of social metadata* is needed. More precisely, we think that collaborative environments and Semantic Web technologies can help. In fact, the availability of online communities consisting of thousands of users would help not only in increasing the number of labeled/tagged resources, but also in assessing their trustworthiness [11].

So far, this issue has been addressed by providing a measure/definition of trust based on some statistics on tags' frequency (see, e.g., [9]). As a naïve example, trustworthiness of a tag associated with a resource could be defined based on the percentage of tags providing identical descriptions for that resource. However, in designing mechanisms to assess social metadata trustworthiness, we think that it is fundamental to take into account also the *explicit* users' opinions on the metadata itself. Indeed, if users could express their *agreement/disagreement with the descriptions* provided by metadata, this would further help in determining metadata trustworthiness.

Further, we believe that several scenarios exist where taking into account user preferences during metadata selection could bring to more meaningful resource descriptions. This is inspired by our normal behaviors in real life. Let us consider, for example, the case of a person, say Kate, who is looking for some science-fiction books for a teenager. In real life, Kate might ask recommendations and suggestions only to those of her friends that are considered expert in science-fiction books. Also, Kate could restrict her selection only to recommendations of those experts that are teens. Applying the same approach in social tag-based applications implies *to make users able to denote who he/she considers trustworthy in describing resources*, that is, who are the users whose metadata have to be elaborated during resource evaluation. In contrast, existing social tag-based applications process resources by elaborating the whole set of metadata associated with it (see for instance [12]).

Further, the tags associated with a resource and their trust values can be used by users to decide how a given resource has to be processed, that is, they can be used to *personalize application behaviors according to user preferences*. For instance, a user might prefer that a recommender system recommends only those books whose metadata state that their content is related to science-fiction with trust value at least equal to 80%.

To cope with the above-discussed requirements, in this paper we propose a framework to support *personalized social tag-based applications based on trust policies and user preferences*. Trust policies allow one to identify trusted users – i.e., users whose metadata have to be considered – according to a variety of criteria (i.e., users' profiles, users' relationships, specific topics). In contrast, user preferences allow users to specify one or more conditions on resources' descriptors and corresponding trust values, and to state which "quality" assessment must be returned (e.g., "the resource is safe for children") and, possibly, which action has to be performed (e.g., recommend, filter, classify) in case at least one of the specified conditions is satisfied.

More precisely, we propose a *multi-layer framework*, where each layer is designed as a black box, providing basic services to the upper layers. The framework supports a *data layer*, gathering metadata from social tagging systems, a *rule layer*, that enforces trust policies and user preferences, and an *application layer*, which elaborates the metadata, filtered according to trust policies, by returning the notices and actions stated by user preferences. As it will be discussed in Section 2, the literature offers several proposals for the *data layer* and the *application layer*, but, to the best of our knowledge, nothing equivalent to our rule layer. More precisely, the main difference of our proposal with regard to existing personalization approaches is that they address a specific issue only, namely, tag/resource recommendation, whereas our framework is designed to be as flexible as possible with regard to the purposes social tags are used for.

The novelty of this framework is in the supported features and in its modular architecture, thanks to which it is possible to tailorize, or even disable, one or more components depending on the different contexts and requirements. We would like to note, however, that, although in this framework trust computation plays an important role, it is just one of the components of our framework. Also, it is not our purpose to propose a new trust system. Actually, our framework does not rely upon a specific method for trust computation, but it is designed to support different methods depending on the considered application scenario.

The work reported in this paper is an extension of [13], where we proposed a system for collaborative resource labeling and label rating, showing how this can be exploited for Web access personalization. In this paper, we significantly extend [13] in that we make it independent from which purpose and by what end user applications it is used. This is achieved by the introduction of the application layer in architecture proposed in [13], obtaining thus a multi-layer framework. With respect to [13], in this paper we have verified the feasibility of the multi-layer approach by designing and developing a prototype system implementing the proposed framework, and addressing a real world scenario. More precisely, we have tested the ability of our framework to (a) reuse datasets of existing Web-based communities, (b) enhance them by providing support to trust policies and user preferences, and (c) return information which can be exploited by end user applications for a variety of purposes. For this purpose, we have used the dataset provided by the Delicious online community (i.e., tags and users' relationships) to implement two distinct case studies, namely, personalized Web search and Web access personalization.

We would like to note that, in this paper, we focus on the technical feasibility of the framework we propose, in order to demonstrate that it is technically possible to enhance existing social media by providing personalization features currently not supported, and by exploiting the potential of user-generated content, as social tags are. Nonetheless, the paper includes a preliminary evaluation of the usability and effectiveness of our framework, which gave us an important feedback on the issues at stake, which we plan to address in future work.

The rest of the paper is organized as follows. Related work is discussed in Section 2. Section 3 provides an introduction to the proposed multi-layer framework. Data layer is described in Section 4,whereas Section 5 presents the rule layer. We prove the feasibility of the proposed framework by showing two case studies in Section 6. Section 7 concludes the paper and outlines future research directions. Finally, Appendix A contains the main algorithms underlying our framework.