



Exploitation of semantic relationships and hierarchical data structures to support a user in his annotation and browsing activities in folksonomies

Pasquale De Meo, Giovanni Quattrone, Domenico Ursino*

DIMET, Università Mediterranea di Reggio Calabria, Via Graziella, Località Feo di Vito, 89122 Reggio Calabria, Italy

ARTICLE INFO

Article history:

Received 2 May 2008

Received in revised form

10 February 2009

Accepted 16 February 2009

Recommended by: F. Korn

Keywords:

Folksonomies

Social tagging

Tag similarity

Social annotation

ABSTRACT

In this paper we present a new approach to supporting users to annotate and browse resources referred by a folksonomy. Our approach is characterized by the following novelties: (i) it proposes a probabilistic technique to quickly and accurately determine the similarity and the generalization degrees of two tags; (ii) it proposes two hierarchical structures and two related algorithms to arrange groups of semantically related tags in a hierarchy; this allows users to visualize tags of their interests according to desired semantic granularities and, then, helps them to find those tags best expressing their information needs. In this paper we first illustrate the technical characteristics of our approach; then we describe various experiments allowing its performance to be tested; finally, we compare it with other related approaches already proposed in the literature.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Motivations

The term *folksonomy* is currently used to indicate a support data structure that allows human users to classify and categorize various kinds of resources (e.g., URLs, photos, videos, scientific papers, and so on) by means of plain keywords, also known as *tags* [6,20,31]. A folksonomy consists of a set of URLs (used to identify the resources referred by it), a set of tags (used to label these resources) and a set of users (who produce and label these resources) [26].

The number of folksonomies on the Web is rapidly increasing since 2004. In the meantime, also the number of resources referred by each folksonomy, as well as the number of users who exploit folksonomies, are rapidly growing [6]. Popular examples of folksonomies are Flickr

[3] (which allows users to annotate their photos), del.icio.us [2] (which allows users to store and share their Web bookmarks), and Bibsonomy [1] (which allows users to share bibliographic data on scientific papers).

Actually, folksonomies are gaining a wider and wider popularity not only on the Web but also in large organizations and businesses (see, for instance [15,33]).

Some of the main reasons underlying the pervasive diffusion of folksonomies are the following:

- In traditional knowledge management systems, classification activity is performed by a human expert or a pool of human experts. As the size and the variety of available information increase, both the costs and the time required to carry out classification in this way increase too. Specifically, we can observe that the rate at which new resources are referred by a folksonomy is very high; for instance, authors of [25] report that, in June 2007, del.icio.us users posted about 120,000 URLs per day; the same study estimated that the number of URLs

* Corresponding author. Tel.: +39 0965 875235; fax: +39 0965 875220.
E-mail addresses: demeo@unirc.it (P. De Meo), quattrone@unirc.it (G. Quattrone), ursino@unirc.it, domenico.ursino@gmail.com (D. Ursino).

stored in del.icio.us in 2007 was about 115 millions. Clearly, in presence of this rate, the amount of time needed by human experts to view, analyse and catalogue these URLs would be huge. Moreover, the pool of human experts should be huge and, consequently, the classification costs would be prohibitive. In addition, resources referred by a folksonomy encompass a large number of disparate domains: for instance, we observed that, in del.icio.us, 524,109 URLs were labelled with the tag “Environment” whereas 617,927 were labelled with the tag “Database”; as a consequence, in this system, there are very important and frequent topics (which, therefore, cannot be neglected) that are, at the same time, very different and possibly related each other. A human expert (or a pool of human experts) would be required to own a very large vocabulary spanning many specialized domains. Moreover, since new topics rapidly emerge, the experts are required to swiftly acquire new knowledge.

An analogous problem arises for large-scale businesses; in fact, also in this context the number and the variety of available resources are high; therefore, also in this scenario, performing and maintaining a classification of available resources is a very complex task which cannot be delegated to a human expert (or a pool of human experts).

In both the two contexts described above it appears necessary a totally different form of coordination in which each user classifies his resources and users' classifications are made available to other users on the basis of their needs/requirements.

- Many authors [6,25] conjecture that the usage of folksonomies (in particular, those regarding Web bookmarks) can enhance the performance of traditional Web search engines. As an example, the authors of [6] observe that the set of tags used to label a Web page can be more effective than a classical information retrieval technique (like TF/IDF) to summarize its content. As a consequence, tags can play the role of metadata and can be effectively used to compute the degree of match between a query and a Web page.
- Folksonomies are useful to highlight hidden social ties among users. For instance, the authors of [22] propose an approach to organize tags in a hierarchical structure and to associate a specific topic with each tag. After this, given two users u_1 and u_2 and a topic T , this approach considers the set $I(u_1, T)$ (resp., $I(u_2, T)$), consisting of the set of resources that u_1 (resp., u_2) labelled with the tags associated with T , and computes the degree of overlap between $I(u_1, T)$ and $I(u_2, T)$; if this degree is higher than a specific threshold it is possible to conclude that there exists a social tie between u_1 and u_2 . Therefore, the exploitation of folksonomies provides a notion of users' social tie more effective than that generally proposed in commercial systems; in fact, these systems assume that there exists a social tie between two users only if this fact is explicitly claimed by them; as a consequence, with these systems, there would not exist a social tie

between two users who do not know each other.

- In large businesses folksonomies are useful to identify communities of users with shared (or complimentary) interests as well as experts of a certain topic. For instance, in the social bookmark system DOGEAR [33] a firm expert can insert a tag t and can identify all URLs tagged with t ; after this, he can retrieve the list of all firm experts who exploited t and can contact them to receive help or additional material on a subject related to t . Moreover, he can select one (or more) of the firm experts and can browse his (their) lists of bookmarks to access new resources, to strengthen his skills or acquire new ones.

However, as clearly pointed out in [20,41], despite these advantages, folksonomies suffer from some, quite crippling, disadvantages that can be summarized in: (i) ambiguity, (ii) usage of synonymous tags, and (iii) discrepancy on granularity of tags. In order to concretely illustrate these disadvantages we consider a real-life folksonomy dealing with Databases and Information Systems; it will be the reference folksonomy throughout the paper. It covers a wide spectrum of topics, like the design/implementation of a Database/Information System, the usage of Information Systems on the Web, the usage of object-oriented programming languages in Databases/Information Systems, and so on.

Ambiguity refers to the fact that some terms may have multiple meanings. The basic example of ambiguity are homonyms; we say that two terms are homonymous if they have the same name but different meanings. For instance, in our reference folksonomy, the tag “Generalization” could be exploited to label both slides about the inheritance among classes in an object-oriented programming language and slides about the generalization relationship in E/R diagrams. As a consequence, the answer to a query consisting of the term “Generalization” performed by a user interested in E/R diagrams would include also a reference to the slides about the inheritance among classes in an object-oriented language.

Usage of synonymous tags in a folksonomy means that different users could exploit different (yet synonymous) tags to label/query the same type of resources (or, even, the same resource). For instance, in our reference folksonomy, some resources about an E/R modelling tool could be labelled with the tag “Data Modelling” by some users, whereas other resources about the same tool could be labelled with the tag “Database Design” by other users. As a consequence, a user who ignores the tag “Database Design” and submits a query containing only the tag “Data Modelling” would not receive some relevant resources as answer to his query. Generally speaking, a user could receive a complete set of answers to his query only if his vocabulary is rich enough to allow him to encompass a tag with the whole set of its synonyms. This hypothesis is clearly unrealistic because users' vocabulary is often quite limited.

Discrepancy on granularity of tags could arise because a resource could be reasonably described by various tags, ranging from terms having a broad meaning to terms

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات