Recommending Smart Tags in a Social Bookmarking System

Pierpaolo Basile, Domenico Gendarmi, Filippo Lanubile, Giovanni Semeraro

University of Bari,
Dipartimento di Informatica,
Via Orabona, 4, 70126 - Bari, Italy
{basilepp,gendarmi,lanubile,semeraro}@di.uniba.it

Abstract. Collaborative tagging systems are harnessing the power of online communities, making the task of knowledge contribution more attractive to a broader audience of Web users. In particular, social bookmarking systems have shifted the organization of bookmarks from an individual activity performed on a personal desktop to a collective endeavor over the Web. In such a context, suggestive tagging has proved to be helpful in consolidating the usage of tags, leading to a quick convergence to a folksonomy.

In a social bookmarking system, users' annotations can be regarded as a reliable indicator of interests and preferences. A recommender system is able to learn user interests and preferences during the interaction in order to construct a user profile.

In this paper, we propose a smart tag recommender able to learn from past user interaction as well as the content of the resources to annotate. The aim of the system is to support users of current social bookmarking systems by providing a list of new meaningful tags. The proposed system is based on ITem Recommender, a content-based recommender previously used in a Digital Library scenario.

Keywords: collaborative tagging, folksonomy, recommender system, semantic web, user profile, suggestive tagging, social bookmaking

1 Introduction

Since Tim Berners-Lee's inceptive Semantic Web vision [2], online communities have taken an active role in the task of knowledge contribution on the Web. Users are no longer passive information consumers, but active participants working in close collaboration to create new content and share it, using the Web as the underlying platform. The phenomenon of Web 2.0[1] has led to the development of several tools which have succeeded in making this task more attractive to a broader audience.

[1] Tim O'Reilly: What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html, 2005.

Powerful tools for lightweight metadata creation, such as collaborative tagging systems, harness the power of virtual communities and have been shown effective in gathering quickly large amounts of information directly generated by users.

Collaborative tagging systems, also known as folksonomies [8], allow people to organize a set of resources, annotating them with tags via a web-based interface. Unlike top-down centralized approaches, folksonomies have revealed a noteworthy ability in adhering to the personal way of thinking [7]. The opportunity of using free tags with no restrictions allows users to express their own perspective on the annotated resource. Therefore, these annotations can become a reliable indicator of interests and preferences of active participants in such systems.

On the other hand, recommender systems [5] are able to learn user interests during the interaction in order to construct (and update) a user profile that can be later exploited for information filtering. A recommender system can be improved by the sheer size of the content available on the Web and the diverse expectations of its user base. Web applications need to combine all available knowledge in order to provide personalized and user-friendly services. Over the years, personalized Web applications and services have been developed, which exploit Web Mining technologies to discover shallow patterns hidden within masses of transactional, navigational, and content-structural data. In addition, knowledge-based recommender systems are able to exploit domain knowledge by integrating domain ontologies.

We think that combining the strengths of Web Mining with the benefit of deeper semantic and the attractiveness of collaborative tagging systems can be a first step to bridge the gap between Semantic Web and Web 2.0.

In this paper we propose an approach to improve an existing recommender system with the purpose of exploiting the information about users' interests provided in form of tags by del.icio.us[2], the most popular social bookmarking system. Our aim is to support users of current collaborative tagging systems by providing tag recommendations based on both the annotations already performed and the content to annotate. The contribution is twofold: A semantic suggesting feature in a social bookmarking system can foster the tag convergence, useful for example to limit the synonymy issue; furthermore, suggesting meaningful tags to a user according to the interests stored in her profile can significantly improve the user experience, augmenting the number of active participants in the collaborative system.

The remainder of the paper is structured as follows. Section 2 presents background information about tag recommendation in social bookmarking systems. An illustrative user scenario motivating our approach is provided in Section 3, while Section 4 describes how we plan to extend our recommender system. Finally, Section 5 draws conclusions and points out some challenges we are going to address in the near future.

2 Related Work

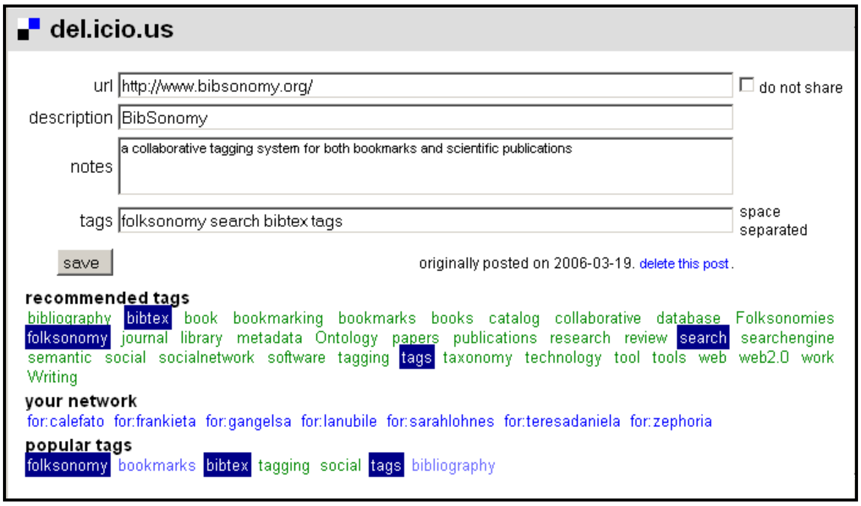
Previous studies on bookmarks use showed that main motivations for creating bookmarks are based on personal interests and quality of the content, high frequency

[2] http://del.icio.us

of current use, as well as a sense of potential reuse [1]. The most familiar approach to store markers for re-finding information on the Web has been through the use of personal bookmarks, supported by almost all browsers. In the last few years, social bookmarking systems have shifted the organization of bookmarks from an individual activity performed on a personal desktop to a collective endeavor over the Web.

Although bookmark collections are personal, the opportunity of accessing to such personal collections from any Web-connected machine (together with the use of free multiple tags, helpful in overcoming the limitation of the traditional hierarchically organized folders) have led to a wide spread of these social systems. Even though contributions are motivated by the private need to easily organize personal items, they also aggregate at a higher level via a collaborative tagging endeavor, that allows the shaping of social networks [13]. Furthermore, some tagging support features, such as suggestive tagging [11], have proved to be helpful in improving the user experience as well as fostering an emerging consensus on the meaning of the terms rising up in the folksonomy [6].

Among the different social bookmarking systems, del.icio.us, one of the earliest and most popular ones, is the only application that illustrates some remarkable suggestive tagging features. When a user saves a bookmark in del.icio.us, she can manually enter as many tags as she would like, but she can also be supported by a list of suggested tags (Figure 1). Popular tags are what other people have tagged this page as, and recommended tags are a combination of tags user has already used and tags that other people have used.



url http://www.bibsonomy.org/ □ do not share

description BibSonomy

notes a collaborative tagging system for both bookmarks and scientific publications

tags folksonomy search bibtex tags

originally posted on 2006-03-19. delete this post.

recommended tags
bibliography bibtex book bookmarking bookmarks books catalog collaborative database Folksonomies folksonomy journal library metadata Ontology papers publications research review search searchengine semantic social socialnetwork software tagging tags taxonomy technology tool tools web web2.0 work Writing

your network
for:calefato for:frankieta for:gangelsa for:lanubile for:sarahlohnes for:teresadaniela for:zephoria

popular tags
folksonomy bookmarks bibtex tagging social tags bibliography

Figure 1. Saving a bookmark in del.icio.us

Rather than recommendations based on some underlying analysis, this kind of suggestions can be regarded as a selection of tags in the sense that the system has to choose a small number of tags to display among the sheer size of terms already associated to an item. According to the tag selection approach, Sen et al. [16]

investigate how different algorithms for selecting tags to display, influence users' personal vocabularies while annotating movies in a movie recommendation system.

On the other hand, as an evidence of the lack of social bookmarking systems that exploit actual tag recommendations (as far as we know), there is few work on such a topic published via the scholarly literature.

Xu et al. [17] define a set of general criteria for a good tagging system to identify the most appropriate tags, while eliminating noise and spam. These criteria, identified through a study of tag usage by real users in My Web 2.0, cover desirable properties of a good tagging system, including high coverage of multiple facets to ensure good recall, least effort to reduce the cost involved in browsing, and high popularity to ensure tag quality. The authors then propose a collaborative tag suggestion algorithm that adopts those criteria to recommend appropriate tags.

Hotho et al. [9] propose an adaptation of both a data mining and information retrieval approach to detect emergent semantics within a collaborative tagging system. The first adaptation lies in reducing the three-dimensional folksonomy to a two-dimensional formal context in order to apply association rule mining techniques. Discovered association rules can be then exploited in a recommender system which supports the user in choosing useful tags. The latter is an adaptation of the PageRank algorithm [3] to the tripartite hypergraph structure of a folksonomy. The algorithm, named FolkRank, incorporates the idea that a node is important if there are many edges from other nodes pointing to it and if those nodes are important themselves, and applies the same principle to the tripartite graph of the folksonomy. The FolkRank algorithm is then used to rank users, tags, resources by their importance. Authors suggest that such rankings can be exploited to generate recommendations for each user about new potential resources of interest, related tags and other users possibly interested on analogous topics.

3 Motivating Scenario

We consider del.icio.us as reference system because of the huge number of registered users and the richness of suggestive tagging. In our scenario, John is a novice user, who has just registered into the system and has no stored bookmarks yet. When John is going to save his first bookmark, the current system suggests popular tags, i.e., terms heavily used by other users to annotate the same resource. A recommender system cannot suggest anything, until the user provides enough information to generate a profile delineating personal interests. However, the use of del.icio.us as an underlying platform makes it possible to support John with popular suggested tags, until the recommender becomes able to actually learn John's interests on the strength of his personal bookmarks and tags.

After John has been using del.icio.us for a while, he has progressively built a large bookmark collection, as well as a rich vocabulary of personal tags that can be exploited by the Smart Tag Recommender system.

When John wishes to save a new bookmark in his personal space, he has a chance to reuse some tags previously used, but he might also enter new tags according to the subject of the resource he is going to annotate. This time the Smart Tag

Recommender can analyze the content of the resource selected by John in order to obtain a collection of concepts describing the bookmark. The output of the content analysis can be then used to retrieve similar bookmarks already annotated by John and find out which tags John has previously used to store such references.

According to the concepts extracted by the analyzer and the tags associated to existing similar bookmarks in John's user profile, the Smart Tag Recommender can now suggest meaningful tags for the resource John wishes to store. The Smart Tag Recommender is not intended to replace the existing del.icio.us recommender, since it provides a new layer of recommendation based on personal profiles and not on popularity.

4 Recommender Architecture

The proposed scenario can be supported by a service that relies on a content-based recommender system, such as ITem Recommender (ITR) [10]. Indeed, this system is able to induce a profile of the user by learning from the content of documents she annotated with a feedback according to her preferences. The induced user profile is a structured representation of user interests which is then exploited to decide whether a new document fits in with the user's preferences. In our case, we consider the problem of learning user profiles as a binary text categorization task [14]: Each document has to be classified as interesting or not with respect to the user preferences. Therefore, the set of categories is $C = \{c_+, c_-\}$, where c_+ is the positive class (*user-likes*) and c_- the negative one (*user-dislikes*). ITR uses a Naïve Bayes method to text categorization; in this way the learned probabilistic model is used to classify a document d_i by selecting the class with the highest probability. As a working model for the Naïve Bayes classifier, we use the multinomial event model [12] to estimate the *a posteriori* probability, $P(c_j|d_i)$ of document d_i belonging to class c_j.

In order to capture the semantics of the user interests, learning is performed on documents that have been previously analyzed by advanced Natural Language Processing (NLP) techniques (implemented in the Content Analyzer module in Figure 2) able to discover relevant concepts representing the content of the documents. The key step in this process is Word Sense Disambiguation (WSD), which is the task of assigning a word with the most appropriate meaning, by taking into account the context (a set of words that precede and follow the word to be disambiguated) in which the word appears. To sum up, documents are represented by concepts instead of keywords, as in the classical vector space model [4]. In order to recognize correctly the meaning of the words, the WSD procedure relies on the WordNet lexical database, in which the set of all possible meanings for each word is maintained. Moreover, the WSD procedure will be integrated with an Entity Recognizer module in order to identify Named Entities that do not occur in WordNet. More details on the ITR system and the WSD procedure are reported in [15].

The ITR system can be easily adapted to the scenario of del.icio.us tags recommending. Indeed, ITR can be used to build a user profile able to support the user in the task of annotating resources by suggesting tags on the basis of previously tagged documents. Given $T = \{t_1, t_2, \ldots, t_n\}$, the set of all tags employed by the user in her "tagging history", the idea is to include a set of *n* binary classifiers, each classifier c_k corresponding to tag t_k, in the user profile. Any new document *d* is then matched against the user profile so that each classifier c_k in the profile can predict whether *d* should be annotated with t_k. The final outcome of the matching process is the set of tags recommended by the classifiers in the user profile.

The set of documents used to train ITR is the set of all the documents previously annotated by the user. Each training document tagged with t_k is considered as a positive example for c_k, while the set of negative examples for c_k is represented by all documents that have not been tagged with t_k.

Figure 2 shows the conceptual architecture of the Smart Tag Recommender system. Full rows indicate the learning step, while dotted rows indicate the classification step.:

a) *Learning step:* An annotated documents is processed by the *Content Analyzer* in order to obtain the Bag-Of-Synsets (BOS) model of the document. To this purpose, NLP techniques, including WSD, are exploited. After that, for each tag t_k the *Profile Extractor* builds the corresponding classifier c_k, that will be part of the *User Profile*.

b) *Classification step:* A new document *(New Doc*) is processed by the *Content Analyzer*, then the *Recommender* uses *User Profile* to select the most appropriated tags for the document. Specifically, for each tag t_k *New Doc* is classified using the corresponding classifiers c_k. The output of this process is the list of recommended tags.

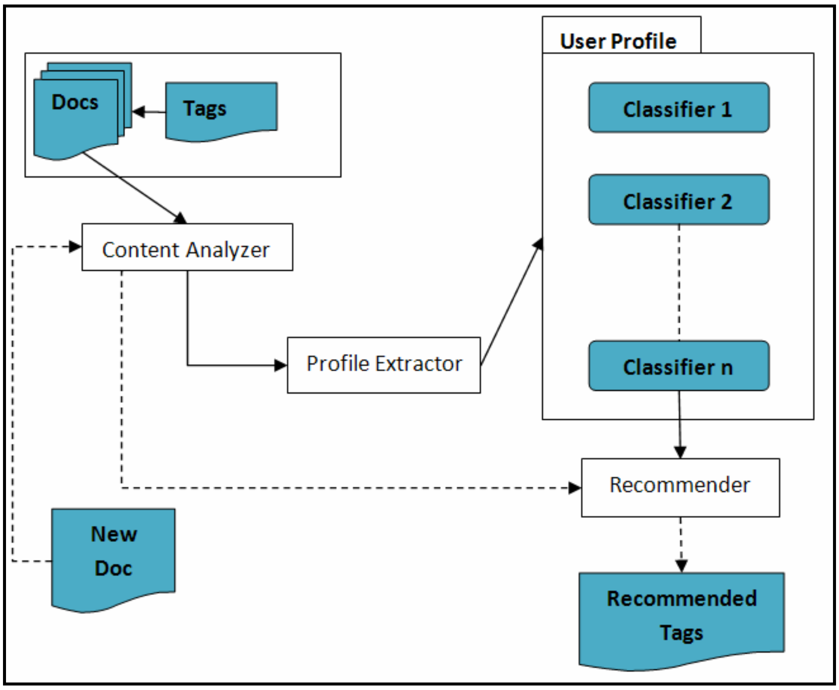


Figure 2. Smart Tag Recommender architecture

5 Conclusions and Future Work

Web 2.0 applications provide chance to semantically exploit the sheer size of user-generated content. Tags in a social bookmarking system can reveal users' interests and preferences. However, current systems suggest a lot of irrelevant tags, either on the basis of personal recent use or because of their popularity among the community. Our aim is to combine the strengths of Semantic Web and Web 2.0 in order to provide better personalized tag recommendations.

In this paper, we have described a strategy to design an intelligent recommender system which is able to learn from both past user interaction and the content of the resources to annotate. The system is based on an existing content-based recommender, that has been previously used in a Digital Library scenario. The main idea is presented in the context of del.icio.us, the most popular social bookmarking system. As future work, we plan to complete the development of the new recommender system and perform an experimental evaluation within del.icio.us, having the basic suggested tagging feature as a control group.

References

1. Abrams, D., Baecker, R., Chignell, M.: Information archiving with bookmarks: personal web space construction and organization. Proceedings of the SIGCHI conference on Human factors in computing systems, (1998), 41–48
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001).
3. Brin, S., Page, L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30 (1–7) (1998), 107–117
4. Bruce, K.B., Cardelli, L., Pierce, B.C.: Comparing Object Encodings. In: Abadi, M., Ito, T. (eds.): Theoretical Aspects of Computer Software. Lecture Notes in Computer Science, Vol. 1281. Springer-Verlag, Berlin Heidelberg New York (1997), 415–438
5. Burke, R.: Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12 (4) (2002), 331–370
6. Gendarmi D., Abbattista F., Lanubile F.: Fostering knowledge evolution through community-based participation. Proceedings of the 1st Workshop on Social and Collaborative Construction of Structured Knowledge at WWW'07, (2007)
7. Gendarmi D., Lanubile F.: Community-driven ontology evolution based on folksonomies. OTM Workshops, LNCS, Vol. 4277. Springer-Verlag, (2006), 181–188
8. Golder, S., Huberman, B.: Usage patterns of collaborative tagging systems. Journal of Information Science, 32(2) (2006), 198-208
9. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Emergent semantics in Bibsonomy. Proceedings of Workshop on Applications of Semantic Technologies, Informatik 2006. Lecture Notes in Informatics, (2006)
10. Lops, P., Degemmis, M., Semeraro, G.: Improving social filtering techniques through Wordnet-based user profiles. Proceedings of 11th International Conference on User Modeling, (2007)
11. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. Proceedings of the Seventeenth Conference on Hypertext and Hypermedia. (2006), 31–40

12. McCallum, A., Nigam, K.: A comparison of event models for naive bayes text classification. Proceedings of the AAAI/ICML-98 Workshop on Learning for Text Categorization, (1998), 41–48
13. Mika, P.: Ontologies are us: A unified model of social networks and semantics. Proceedings of the 4th International Semantic Web Conference, LNCS, Vol. 3729. Springer-Verlag, (2005) 522-536
14. Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys, 34(1), (2002)
15. Semeraro, G., Degemmis, M., Lops, P., Basile, P.: Combining learning and word sense disambiguation for intelligent user profiling. Proceedings of twentieth International Joint Conference on Artificial Intelligence, (2007), 2856–2861
16. Sen, S., Lam, S. K., Rashid, A., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., Riedl, J.: Tagging, communities, vocabulary, evolution. Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, (2006), 181-190
17. Xu, Z., Fu, Y., Mao, J., Su, D.: Towards the Semantic Web: Collaborative Tag Suggestions. Proceedings of Collaborative Web Tagging Workshop at 15th International World Wide Web Conference (2006).