FolksAnnotation: A Semantic Metadata Tool for Annotating Learning Resources Using Folksonomies and Domain Ontologies

Hend S. Al-Khalifa and Hugh C. Davis
Learning Technology Research Group, ECS, The University of Southampton, Southampton, UK
{hsak04r/hcd}@ecs.soton.ac.uk

Abstract

There are many resources on the Web which are suitable for educational purposes. Unfortunately the task of identifying suitable resources for a particular educational purpose is difficult as they have not typically been annotated with educational metadata. However, many resources have now been annotated in an unstructured manner within contemporary social bookmaking services. This paper describes a novel tool called 'FolksAnnotation' that creates annotations with educational semantics from the del.icio.us bookmarking service, guided by appropriate domain ontologies.

1. Introduction

Metadata standards such as Dublin Core (DC) and IEEE-LOM[1] have been widely used in e-learning applications. However, with the advent of the Semantic Web, e-learning applications are beginning to evolve their metadata representation from these standards by adding semantic structure or by converting entirely to semantic representations. One reason behind this evolution is that semantic metadata are resilient; they can be further annotated with more metadata without being confined to a fixed template.

Nonetheless, in the era of so-called 'Web 2.0[TM]', a new family of social applications is currently emerging. Wikis, blogs and social bookmarking services, to name a few, are all signatures of the Web 2.0 phenomenon. The contents of these applications, especially social bookmarking services have been annotated in an unstructured manner using folksonomies. These folksonomic annotations are also considered a form of metadata.

Furthermore, web resources stored in social bookmarking services have potential for educational use. In order to realize this potential, we need to add an extra layer of semantic metadata to the web resources stored in social bookmarking services; this can be done via adding pedagogical values to the web resources so that they can be used in an educational context.

In this paper, we will focus on how to benefit from social bookmarking services (in particular del.icio.us[2]), as vehicles to share and add semantic metadata to bookmarked resources. Thus, the research question we are tackling is: *how folksonomies can support the semantic annotation of web resources from an educational perspective*?

The organization of the paper is as follows: Section 2 introduces folksonomies and social bookmarking services. In section 3, we review some recent research on folksonomies and social bookmarking services. In section 4, we outline the three ontologies we have used for the annotation process. Then in section 5, we describe the system architecture of our prototype tool. Finally, sections 6 and 7 respectively, report on preliminary evaluation of the tool and conclude the paper with a summary and ongoing research.

2. Folksonomy and Social Bookmarking Services

The growing popularity of folksonomies and social bookmarking services has changed how people interact with the Web. Many people have used social bookmarking services to bookmark web resources they feel most interesting to them, and folksonomies were used in these services to represent knowledge about the bookmarked resource. Next a brief overview of the two named concepts will be given.

---

[1] The Learning Object Metadata (LOM) was created by the Learning Technology Standards Committee (LTSC) of the IEEE, http://ltsc.ieee.org/wg12/

[2] A social software web service for storing and sharing web bookmarks. http:// del.icio.us

2.1 Folksonomies

The word folksonomy is a blend of the two words 'Folks' and 'Taxonomy'. It was first coined by the information architect Thomas Vander Wal in August of 2004. Folksonomy as Thomas [1] defines is:

"… the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. *The tagging is done in a social environment (shared and open to others). The act of tagging is done by the person consuming the information.*"

From a categorization perspective, folksonomy and taxonomy can be placed at the two opposite ends of categorization spectrum. The major difference between folksonomies and taxonomies are discussed thoroughly in [2] and [3].

Taxonomy is a top-down approach. It is a simple kind of ontology that provides hierarchical and domain specific vocabulary which describes the elements of a domain and their hierarchal relationship. Moreover, they are created by domain experts and librarians, and require an authoritative source.

In contrast, folksonomy is a bottom-up approach. It does not hold a specific vocabulary nor does it have an explicit hierarchy. It is the result of peoples' own vocabulary, thus, it has no limit (it is open ended), and tags are not stable nor comprehensive. Moreover, folksonomies are generated by people who have spent their time exploring and interacting with the tagged resource [4].

2.2 Social Bookmarking Service

Social bookmarking services are server-side web applications; where people can use these services to save their favorite links for later retrieval. Each bookmarked URL is accompanied by a line of text describing it and a set of tags (aka folksonomies) assigned by people who bookmarked the resource (as shown in Figure 1).

Figure 1. Excerpt from the del.icio.us service showing the tags (Blogs, internet, ... ,cool) for the URL of the article by Jonathan J. Harris, the last bookmarker (pacoc, 3mins ago) and the number of people who bookmarked this URL (1494 other people)

A plethora of bookmarking services such as Furl[3], Spurl[4] and del.icio.us exists; however, del.icio.us is considered one of the largest social bookmarking services on the Web. Since its introduction in December 2003, it has gained popularity over time and there have been more than 90,000 registered users using the service and over a million unique tagged bookmarks [5, 6]. Visitors and users of the del.icio.us service can browse the bookmarked URLs by user, by keywords (aka tags or folksonomies) or by a combination of both techniques. By browsing others bookmarks, people can learn how other people tag their resources; thus, increasing their awareness of the different usage of the tags. In addition, any user can create an inbox for other users' bookmarks, by subscribing to the other user's del.icio.us pages. Also, users can subscribe to RSS feeds for a particular tag, group of tags or other users.

3. State of the Art - Related Work

To the best of our knowledge, no scholarly publications have addressed the use of folksonomies in the process of semantically annotating learning resources. However, there is a lot of recent research dealing with folksonomies. Among them are overviews of social bookmarking tools with special emphasis on folksonomies as provided by [7]. On the other hand, other research papers have discusses the strengths and weaknesses of folksonomies as in [8], [9], [3] and [10].

Another genre of research has experimented with folksonomy systems. For instance, Mika [11] has carried out a study to construct a community-based ontology using del.icio.us as a data source. He created two lightweight ontologies out of folksonomies; one is the actor-concept (user-concept) ontology and the other is the concept-instance ontology. The goal of his experiment was to show that ontologies can be built using the context of the community in which they are created (i.e. del.icio.us community). By the same token, Tom Guber is working on a system called 'TagOntology' to build ontologies out of folksonomies, and in his paper entitled "Ontology of Folksonomy: A Mash-up of Apples and Oranges" he cast some light on some design considerations needed to be taken into account when constructing ontologies from tags [12]. In addition, Ohmukai et al. [13] proposed a social bookmark system, called 'socialware', using several representations of personal network and metadata to construct a community-based ontology. The personal network was constructed using Friend-Of-A-Friend (FOAF), Rich Site Summary (RSS), and simple Resource Description Framework Schema (RDFS),

---

[3] http://www.furl.net/

[4] http://www.spurl.net/

while folksonomies were used as the metadata. Their system allows users to browse friends' bookmarks on his/her personal network, and map their own tag onto more than one tag from different friends, so that they are linked by the user. This technique will allow for efficient recommendation for tags because it is derived from personal interest and trust. They also used their social bookmark system 'socialware' to design an RDF-based metadata framework to support open and distributed models.

Golder and Huberman, from HP Labs, [14] have analyzed the structure of collaborative tagging (aka folksonomies) to discover the regularities in user activity, tag frequencies, the kind of tags used and bursts of popularity in bookmarked URLs in the del.icio.us system. They also developed a dynamic model that predicts the stable patterns in collaborative tagging and relates them to shared knowledge. Their results show that a significant amount of tagging is done for personal use rather than public benefit. However, even if the information is tagged for personal use other users can benefit from it. They also state that del.icio.us, for most users, functions as a recommendation system even without explicitly providing recommendation.

In MIT labs, an experiment was carried out by Liu et al. [15] to generate a taste fabric of social networks. Folksonomies were used in the experiment to weave the taste fabric. Their idea was based on philosophical and sociological theories of taste and identity to weave a semantic fabric of taste. They mined 100,000 social network profiles, segmented them into interest categories and then normalized the folksonomies in the segments and mapped them into a formal ontology of identity and interest descriptors. Their work has inspired us in the idea of using folksonomies in the process of semantic annotation.

Finally, Hotho et al. [16] have presented a new search algorithm for folksonomies, called 'FolkRank', which exploits the structure of the folksonomy. Their proposed algorithm is used to support the retrieval of resources in the del.icio.us social bookmarking services by ranking the popularity of tags. They demonstrated their findings on a large-scale dataset (around 250k bookmarked resources) and showed that their algorithm yielded a set of related users and resources for a given tag. Therefore, 'FolkRank' can be used to generate recommendations within a folksonomy system.

From the previous discussion the reader can observe that most research on folksonomies is either user-centric e.g. [11] and [13] or tag-centric e.g. [12], [15] and [16]. Little has been addressed towards the URL-centric perspective, which our system tackles. By URL-

centric we mean that we are constructing knowledge about a specific URL from the tags associated with it.

4. Ontologies For resources representation

The main goal for building ontologies in e-Learning systems is to represent the semantics of the learning resource; such semantics are stored in repositories, so that they can be reused, shared and queried by users (e.g. teachers and students).

Moreover, ontologies form the base for generating semantic metadata, and in our proposed system, they are used to add more fine grained semantics to web resources.

4.1 Ontological Modeling of the three Ontologies

The three ontologies in our system were chosen based on observed patterns in peoples' tags in the del.icio.us bookmarking service for our domain of interest (for this case study our domain of interest will be teaching 'CSS' in a 'web design' course). In other words, the three ontologies we have modeled are the domain of 'web design', the subject of 'CSS' and the learning resources 'types'. Therefore, the three ontologies were designed with a web design and development course in mind, and they were built using Protégé[5] ontology editor and saved in OWL DL format.

A) **Web Design Domain Ontology**: The domain ontology represents an abstract level of the domain of 'web design' and the relation of the concepts in that domain. The rational of using domain ontology is to place the CSS subject ontology in the context of its domain.

B) **CSS Subject Ontology**: The CSS ontology gives a fine grained listing of the concepts used in the subject of CSS. The concepts were derived from assorted websites that classify the subject of CSS.

C) **Resource Type Ontology**: The resource type ontology models resource types that go beyond the scope of the common-set provided by IEEE-LOM. The rationale behind using a different vocabulary set is that different learning resources can come in a variety of forms. To give an example, suppose a learning resource was of type 'editor', a software tool used to create or modify files of a particular type. A possible use of the resource will be to use it as an additional resource in the context of a programming course. This type of resource and others have not been mentioned in the IEEE-LOM

---

[5] http://protege.stanford.edu/

resource type set; therefore, new vocabulary needs to be modeled to represent the new resources emerging in people's vocabulary.

4.2 The Semantic Metadata

Learning resources are usually described using standards such as Dublin Core and IEEE LOM with their RDF bindings. The semantic metadata used in this system builds on these standards and add more fine grained semantics to web resources. In other words, parts of the generated semantic metadata elements were derived from the IEEE-LOM standard. These elements include the web resource title, description, URL and keywords. The rest of the generated semantic metadata elements were derived from our ontologies. These include fine grained elements that are dedicated for the subject of 'CSS'.

5. The FolksAnnotation System Architecture

Two components play a major role in our system, namely: Folksonomies and Ontologies. Folksonomies are used to find the corresponding concepts in the ontology, and ontologies are used as the schema which the generated metadata will adhere to.

However, before using folksonomies as guides in the semantic annotation process, we have carried out an experiment to measure the semantic value of folksonomies compared to automatic keyword extraction techniques. The aim of our experiment was to show that even if folksonomies are used as index keywords; automatic keyword extraction can not simply replace the collective knowledge of people depicted in the form of folksonomies. Therefore, the experiment followed two approaches: automatically, by measuring the percentage of overlap between the folksonomy set and a machine extracted keywords set for a given document; and subjectively, by asking a human indexer to evaluate the quality of the generated keywords from both systems. The results of the experiment showed that folksonomies hold more semantic value than keywords extracted using machines. For a full detail of the experiment, the reader is referred to [17].

The system we have implemented consists of two stages: 1) tags extraction and normalization and 2) semantic metadata creation, as shown in Figure 2.

In the first stage, all tags assigned to a web resource in the del.icio.us service are extracted and then normalized using several techniques. First, tags are converted to lower case so that string manipulation (e.g. comparison) can be applied to them easily.

Secondly, non-English characters are dropped; this step is to insure that only English tags are present when doing the semantic annotation process. Thirdly, tags are stemmed (e.g. convert plural to singular) then similar tags are grouped (e.g. inclusion of substrings). Finally, general concept tags (e.g. programming, web, etc) in our domain of interest are eliminated. The process of normalization is done automatically and it is potentially useful to clean up the noise in people's tags.
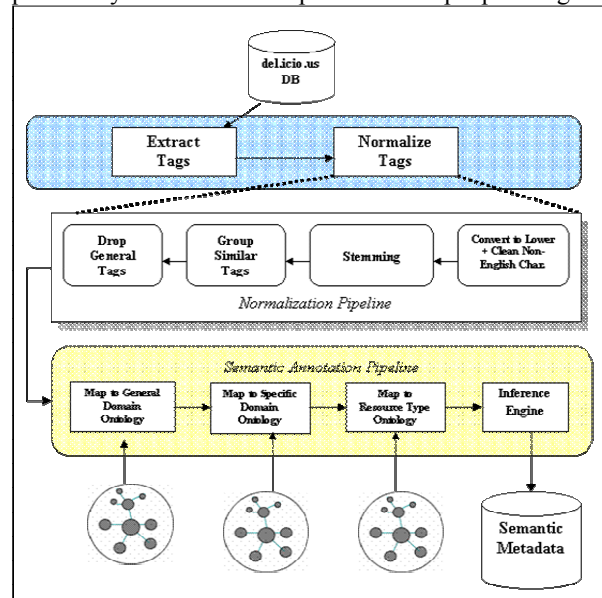

Figure 2. System Architecture of the 'FolksAnnotation' Tool

The normalized list is then ready to be used in the semantic annotation pipeline process, where each normalized folksonomy tag will be mapped to the different ontologies concepts.

The Semantic Annotation pipeline process is the backbone process that generates semantic metadata using the three proposed ontologies. The process attempts to match normalized folksonomy terms from the bookmarked resource against terms in the ontology (i.e. the ontologies work as a controlled vocabulary) and only selects those terms that appear in the ontology.

After assigning semantic descriptors to the web resource, the inference engine is responsible of associating pedagogical semantics (i.e. difficulty level and instructional level) to the annotated web resource. These two values are generated from a set of reasoning rules when enough information is available in the basic semantic descriptors. To give an example, suppose a web resource within the subject of 'CSS' was tagged with a folksonomy value of 'font'. When this tag is passed through the reasoning rules it will trigger the rule that states (if a web resource has a tag value of

'font' then its difficulty level will be 'easy' and its instructional level will be 'basic').

After finishing the annotation process, each item of the generated semantic metadata is saved in a database (e.g. a triple store).

6. Preliminary Evaluation

To evaluate the performance of the generated semantic metadata, we have embarked an evaluation procedure adopted from [18], where they compared *keywords* to *semantic topic* search. However, in our system we have compared the performance of *folksonomy* search against *semantic topic* search to see which search results in more relevant results.

CSS topics were entered in the folksonomy search field and the number of returned records was compared with the number of the records returned by the same topics when mapped to the CSS ontology. Topics covered by the CSS ontology include: *BoxModel, Layout, Navigation, Positioning and Typography*.

The implemented semantic search benefits from the relation between topics in the CSS ontology to retrieve relevant resources, in this case the '*related_to*' relation which links between related concepts. For instance, when someone searches for the topic 'positioning', all resources that have as their subject the word 'positioning' plus all related resources will be retrieved. Table 1 shows the result obtained when searching for the positioning topics in the CSS ontology.

Table 1. The relevance result between Folksonomy search and topic search

CSS Topic	Positioning	
	Folksonomy Search	**Topic Search**
Number of records found	3	4
Number of records relevant to topic	3/4	4/4

7. Conclusion and Ongoing Work

Our aim in this research was to show that semantic metadata can be generated using folksonomies guided by domain ontologies. And to some extent we tried to show that part of our claim is valid by reporting on the results of the preliminary evaluation.

However, despite the preliminary evaluation results which showed that semantic search of folksonomies is more powerful than searching by folksonomies alone, further rigorous evaluation procedure is planned. This

will include the measurement of the validity, quality and performance of the generated semantic metadata.

8. References

[1] Folksonomy definition and wikipedia. Vander Wal, T. Accessed on April 29, 2006.Available on line http://www.vanderwal.net/random/category.php?cat=153

[2] Ontology is Overrated: Categories, Links, and Tags. Shirky, C. Accessed on March 27, 2006.Available on line http://shirky.com/writings/ontology_overrated.html

[3] Quintarelli, E. Folksonomies: power to the people. in ISKO Italy-UniMIB meeting. 2005. Milan, Itlay.

[4] Folksonomy. Wikipedia. Accessed on March 26, 2006.Available on line http://en.wikipedia.org/wiki/Folksonomy

[5] Sieck, S., connotea and citeulike: "folksonomies" emerge within scholarly communities, E. Insight, Editor. 2005

[6] Menchen, E. Feedback, Motivation and Collectivity in a Social Bookmarking System. in Kairosnews Computers and Writing Online Conference. 2005.

[7] Hammond, T., T. Hannay, B. Lund, and J. Scott, Social Bookmarking Tools (I): A General Review. D-Lib Magazine, 2005. 11(4).

[8] Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Mathes, A. Accessed on February 28, 2006.Available on line http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

[9] The Hive Mind: Folksonomies and User-Based Tagging. Kroski, E. Accessed on January 14, 2006.Available on line http://infotangle.blogsome.com/category/folksonomies

[10] Guy, M. and E. Tonkin, Folksonomies: Tidying up Tags? D-Lib Magazine, 2006. V 12(1).

[11] Mika, P. Ontologies are us: A unified model of social networks and semantics. in In Proceedings of the Fourth International Semantic Web Conference (ISWC 2005). 2005. Galway, Ireland: Lecture Notes in Computer Science.

[12] Gruber, T., Ontology of Folksonomy: A Mash-up of Apples and Oranges. AIS SIGSEMIS Bulletin, 2005. 2(3&4).

[13] Ohmukai, I., M. Hamasaki, and H. Takeda. A Proposal of Community-based Folksonomy with RDF Metadata. in The 4th International Semantic Web Conference (ISWC2005). 2005. Galway, Ireland.

[14] Golder, S.A. and B.A. Huberman, The Structure of Collaborative Tagging Systems. Journal of Information Science, 2006.

[15] Liu, H., P. Maes, and G. Davenport, Unraveling the taste fabric of social networks. International Journal on Semantic Web and Information Systems, 2006. 2(1): p. 42-71.

[16] Hotho, A., R. Jäschke, C. Schmitz, and G. Stumme. Information Retrieval in Folksonomies: Search and Ranking. in Proceedings of the 3rd European Semantic Web Conference. 2006. Budva, Montenegro: LNCS, Springer.

[17] Al-Khalifa, H.S. and H.C. Davis. FOLKSONOMIES VERSUS AUTOMATIC KEYWORD EXTRACTION: AN EMPIRICAL STUDY. In Proceedings of IADIS Web Applications and Research 2006 (WAR2006). 2006. Online

[18] Li, J. Z., Gasevic, D., Nesbit, J. C., & Richards, G. (2005). Ontology Mappings Enable Interoperation of Knowledge Domain Taxonomies. Paper presented at the 2nd LORNET international annual conference, November 16-18, 2005, Vancouver, Canada