Collaborative Tagging Approaches for Ontological Metadata in Adaptive E-Learning Systems

Scott Bateman, Christopher Brooks and Gord McCalla

Lab for Advanced Research in Intelligent Educational Systems (ARIES Lab)
Department of Computer Science, University of Saskatchewan
Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, cab938@mail.usask.ca, mccalla@cs.usask.ca

Abstract. One of the main approaches for creating metadata for learning resources in adaptive e-learning systems has been through the use of semantic web ontologies. This approach is limiting because it doesn't usually address a requirement for the support of annotators or the requirement for significant effort by annotators in learning ontological metadata domains and technologies. This paper proposes a method of addressing these shortfalls, by incorporating techniques used on the hugely popular collaborative tagging websites, such as del.icio.us[1]. By extending on a natural language ontology, we aim to achieve simplicity in metadata authoring while maintaining the expressiveness of a lightweight ontology. The goal of the approach is to facilitate metadata creation such that new metadata creators (such as students) may become involved with creating machine consumable metadata about learning objects, for use in adaptive e-learning systems.

1 Introduction

The key to the Semantic Web is the creation of machine consumable knowledge [1]. The main approach being taken is to use ontologies to describe concepts, the properties of concepts, and the relationships between concepts. Among the e-learning community much focus has been placed on investigating how ontologies might facilitate the creation of adaptive e-learning systems. The approach has been to use instances of concept descriptions to annotate resources (often learning objects), as seen in [2–4]. However, ontologies are not without their drawbacks. Materialized, these drawbacks not only limit the amount and quality of ontological metadata created but also who can be involved in its creation and therefore the overall usefulness of the approach.

This paper will explore the idea of merging the use of ontologies with collaborative tagging through a new approach we have dubbed *CommonFolks*. This approach should help to reduce the effort required in the production of useful

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

metadata, while maintaining the expressiveness inherent in lightweight ontologies, thus opening the door to a better quality of metadata and authoring by those not traditionally involved in metadata creation.

The next section will continue to explore ontologies and collaborative tagging as alternative approaches to metadata creation and use, and identify some corresponding strengths and weaknesses. Section three will then detail how CommonFolks uses aspects of both approaches, along with the WordNet [5] English language ontology, to create a simple yet expressive metadata framework. Further, anecdotal observations of an initial prototype are given, which found the approach more straight forward for inexperienced metadata authors when compared with another metadata creation tool. Finally, future directions for the further development of the CommonFolks framework will be discussed.

2 Representing Knowledge: Comments on Taxonomies, Ontologies, and Collaborative Tagging

Often collaborative tags are referred to as folksonomies, as was originally intended by Thomas Vander Wal who coined the term. He combined the words "folk" and "taxonomy", and was referring to the now popular collaborative classification scheme. Taxonomies, are hierarchical structures of classifications for concepts, where the child concepts are a subclass of their parents [6]. Since these systems do not contain hierarchies we will use the term: collaborative tagging.

Collaborative tagging techniques can be seen on websites like del.icio.us, Flickr[2], and CiteULike[3], where people publicly annotate resources with keywords that describe those resources (called tags). Each tag is both an annotation to describe the resource and a vote for the annotation being suggested to others. Tags have no intrinsic relationship with one another, and cannot be easily arranged into hierarchies of meaning. Nonetheless, they have become a powerful tool for metadata creation, and have been principle in exploiting the abilities of the casual web surfer to contribute to the larger web community (a main focus in the "Web 2.0" movement) [7].

Ontologies extend taxonomies by further describing the relationships between any concepts, effectively turning a hierarchy into a directed graph. Gruber defined an ontology, as "... an explicit specification of a conceptualization" [8], which was refined by Borst as "...a formal specification of a shared conceptualization" [9]. Though ontologies give the foundations for interchangeable knowledge representation, we argue that this does not necessarily entail their conceptualizations being shared. Ontology terms can be easily misinterpreted for many reasons, such as synonymous concept definitions, misinterpretation of concepts or different applications in end use. While it is true that a representation of conceptualization may be shared, it is incorrect in saying that the conceptualization itself is "shared". Rather, for a conceptualization to be truly shared,

[2] http://flickr.com
[3] http://citeulike.org

we must also fully represent how it is to be used. Our anecdotal experience is that collaborative taggers tend to create metadata with their own end uses in mind. So, collaborative tagging by its very nature lends itself to allow for sharing conceptualizations with a representation of purpose.

2.1 Problems with Current Learning Object Metadata

The IEEE's Learning Object Metadata (LOM) [10], is the most widely used specification for learning object metadata. To fit into our discussion, we will define the LOM as, *a taxonomy of terms, some of which are constrained by preset vocabularies.* The LOM is a big and well designed specification, and has in the neighborhood of 80 terms, but it is at the same time too broad and too specific to be applied to all possible annotations. As described in [11], the LOM requires a lot of information, but at the same time there may be information not adequately represented. There is significant effort used to create and define such a structure, and the training of instructors to provide good metadata when learning objects are being created has been an ongoing issue. Even the creators of the standard, the IMS Global Learning Consortium, note that the desire does not exist in practice to support such a detailed structure:

> "Many vendors expressed little or no interest in developing products that were required to support a set of meta-data with over 80 elements [and the] burden to support 80+ meta-data elements on the first iteration of a product is too great for most vendors to choose to bear." [12]

Further, we have seen in less detailed metadata, that non-expert annotators desire help from experts while creating their metadata, and this is especially true of relational types of data that are inherent in ontologies [13]. We can extend, our discussion of taxonomic metadata, therefore, to the case of ontologies. The difficulties associated with metadata can also be problems for ontologies. Compounding the difficulty of using ontological metadata is the requirement that the annotator has an in-depth knowledge of the use of tools and technologies. Thus there exists a need to have ontological metadata more easily created and easily used, while making help available throughout the authoring process.

2.2 Enter Collaborative Tagging

Though not nearly as expressive for machine knowledge as ontologies, collaborative tags are more easily applied to resources because of their simplicity. Their lack of expression can be seen in more complex domains - for instance, some del.icio.us users have implemented a hierarchy of tags (effectively taxonomies) by prepending categorical symbols to tags [14]. This ad hoc extension to collaborative tags still allows ambiguous semantics (e.g. is the child tag a more narrow term from the parent, or are elements within the child a subset of the parent?) and begins to become less useful for other users who are not using a similar strategy. Further, collaborative tags have no property or predicate relationship explicitly implied by the tags. This is okay for humans who are good at determining which keyword implies which relation, but a computer would have great difficulty doing the same thing. For instance, tagging the URL for “Adaptive Hypermedia 2006 Conference” webpage, http://ah2006.org, with “semantic”, “web”, “Barry” and “Smyth”. Humans seeing these tags can determine fairly easily what the tags mean, especially if they know anything about AH2006. They imply that the conference has some involvement with the Semantic Web technologies, and that Barry Smyth is involved in its organization. We can see that computers may have a great deal of trouble determining what a human can see relatively easily. Even so, the second relation may be a bit of stretch even for a human, because of other possibilities for the implied predicate (ie. “presenter at”). Further, the computer may not be able to group the tags together effectively, “semantic” and “web”, should really be treated as the compound term, “semantic web”. These issues limit the approach greatly, not just for computers since they create understanding problems for humans too.

With a lack of a hierarchy, tags associated with resources can grow to unruly proportions, even for the tags of a single user. However, the lack of relational information and the ease of simply using keywords without design allows users to annotate more and worry less - if they miss an appropriate tag it is likely that someone else in the community will get it. The universal applicability of this approach arises despite what the majority in a tagging community would deem erroneous tags or meta noise. Rather, it is this normally useless information which allows the metadata to cater to minorities as well. Shirky, on his well read posting, argues for the organic classification system of tags, over a the rigidly dictated ontological techniques, “... del.icio.us has no idea what the tags mean. The tag overlap is in the system, but the tag semantics are in the users. This is not a way to inject linguistic meaning into the machine” [15]. His point hits directly at the main problem with collaborative tagging: the semantics of the tags are only useful for human consumption and computational entities can provide only limited reasoning in a collaborative tagging system. This is in some ways a catch-22: collaborative tagging harnesses the power of the community and has been shown effective in creating large amounts of metadata quickly, but this metadata is limited to human consumption only, which can be quickly overwhelmed. Ontologies (in the semantic web sense), on the other hand, were designed specifically to provide rich machine decidable semantic representations. These semantics are found within the subject-property-object relations that have been defined by the W3C’s core ontological RDF technologies. This technique of knowledge representation allows for more intelligent searching and reasoning over resources, as can be seen in [16].

The idea of combining community knowledge sharing techniques with writing RDF statements has resulted in the award winning Confoto[4].Confoto, like Flickr creates annotations of pictures; however, it uses RDF statements instead of keywords. Unfortunately, the application falls short of reaching the ease of use of Flickr. Further, there is no way to extend the knowledge represented in

[4] http://www.confoto.org

the application, as its annotations are based on a closed set of ontological terms, and there is no authoring support for metadata creators.

The examination of ontologies and collaborative tags closes with a brief examination of the process for creating metadata with each approach. Steps in a standard approach might include:

1. learning languages for writing ontologies (RDF, RDFS, and possibly OWL)
2. designing an ontology for the domain or finding and customizing an existing domain ontology
3. creating metadata with instances that fit the ontology
4. assess the ontologies usefulness for some application(s)
5. return to the design step when the ontology doesn't meet all needs

The de facto standard for creating metadata with collaborative tagging is much simpler:

1. write appropriate tags (keywords) taking into account community recommendations, until you think you have enough

3 CommonFolks

3.1 The Approach

With CommonFolks, we wish to define a method for anyone to easily annotate learning objects with metadata, while providing enough semantics to map metadata and thus meet the needs of adaptive e-learning systems. The name CommonFolks borrows from the etymology of folksonomy - we want "commonly" defined and used metadata, designed by people without special training ("folks"). Using the same techniques to annotate resources with keyword tags, annotations can be made with subject-predicate-object triples as seen with Confoto.

Collaborative tags are simple because they are based on natural language. Therefore, our only prerequisite knowledge for their creation is knowledge of English. Granted, while these keywords sometimes become symbolic compound statements [14], they nonetheless have natural language connotations for their authors. We take the advice offered by Smith, et al. in the OWL Guide,

> "We want simple assertions about class membership to have broad and useful implications. This is the hardest part of ontology development. If you can find an existing ontology that has already undergone extensive use and refinement, it makes sense to adopt it." [17]

WordNet [5] is both an ontology that represents natural language as well as an instance database of English terms, which is based on research into human lexical memory. It provides a lattice work of the English language, where terms (words or compound words) have relations to descendant terms (hyponyms) and ancestor terms (hypernyms). The structure of the relations of the lattice imply an "is a" (subclass) relationship. For instance fig. 1.i shows the lattice for the

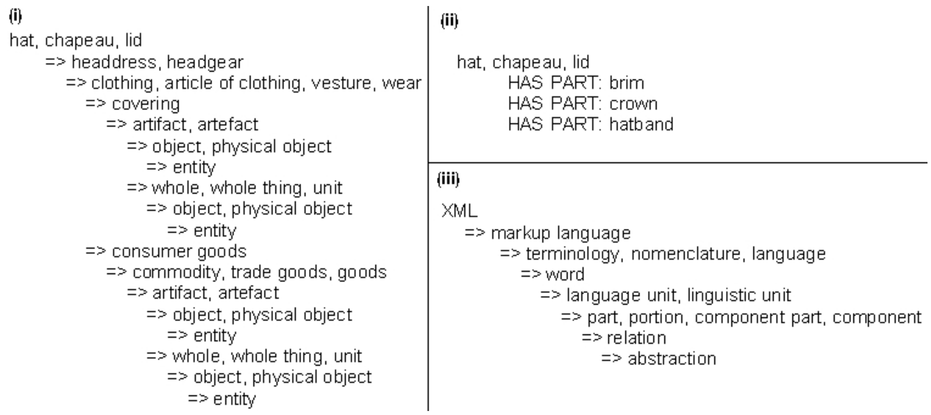

Fig. 1. Examples from the WordNet lattice: (i) shows the lattice for the first sense of "hat". (ii) shows the meronyms or parts of a hat. (iii) shows how we can add a term "XML" to extend the WordNet lattice with new terms.

first sense of term "hat". Multiple senses of a term therefore implies a lattice for each sense, where based on sense of a term, there are different "is a" relationships. Further, it provides synonymous terms, antonyms (negative senses of the term), and any derived terms for each individual sense, if they exist. WordNet's "part of" relationships are shown in 1.ii. Here we see the meronym relation, <hat> <has part> <brim>. Finally, the last relationship we will mention is, an instance. An instance, is a real-world example of a term; <"Sir Walter Scott"> is an <instance> of <author>.

3.2 Putting CommonFolks to Work

We will now show how we can leverage WordNet and collaborative tagging techniques to be applied in metadata creation. The CommonFolks lattice contains the WordNet lattice plus any terms that have been added. Tagging in CommonFolks, as in RDF, is simply adding a predicate and an object to a given subject (a resource, typically the learning object). Rather than designing an ontology and then providing instances that fit the ontology for given resources, we are able to skip the design process and begin using instances to describe our resource, since instances will be added to or exist in the WordNet lattice. We can later infer a domain ontology for a set of tags (instances used to describe resources) at any point, based on our own or the communities annotations. We will refer to the CommonFolks Tools (CFT) as a web application that facilitate CommonFolks metadata creation and collaborative tagging techniques. The general process of using CFT is now shown in a example of annotating a learning object.

Consider a learning object about XML. The URI we have associated for our learning object is http://example.org/lessons/XML-Intro. Our first step is to declare what a learning object is. Considering that we have not first declared what a "learning object" is using CFT, and it is not a term available in WordNet, we must first add it to the lattice. We will now either search or navigate through the WordNet lattice to find what we consider to be the best place for “learning object”. CFT be able to give recommendations (provided other CFT users have already added the term “learning object”) - imagine that CFT recommends a few possibilities and the most widely used parent for “learning object” is “school text”. The most popular choices for the position of “learning object” represents a form of community consensus. We decide that this is a good idea and add it to have “school text” as a parent. We have now implied the “is a” relationship and that “learning object” is a “school text” and that “learning object” is also all the other ancestors in the lattice for “school text”. Now we use the same process to add the URI, which represents the XML tutorial, which is an instance. We choose the parent of http://example.org/lessons/XML-Intro as our new term “learning object”. We have now added the minimal amount of information required in CommonFolks. We have already implied our first “tag” or predicate-object relationship to our subject; the URI is a “learning object”. However, let us now explore how to declare more information about our “learning object”. It should be noted, that from now on, CFT will keep track of our new terms, and they will no longer need to be added; further, any new additions we make will be checked to disallow any contradictory or inconsistent statements. Considering that people have already added “learning object” as a term, they have probably also made some tags on “learning object” objects. This being said, CFT will point out the most commonly used properties for “learning object”, the URI, or any of the ancestors in the lattice.

Next, consider adding an annotation to describe the XML tutorial and receiving two community suggestions that are very similar. For this example we will want to declare the subject of the tutorial (which we will say is XML). In our selection list of properties assume we can see two properties that could accurately denote what we want, “has subject” or “is about”. This will come down to personal preference, which could be based on reading some of the most common definitions associated with the properties, on how they are ranked or some other preference. If nothing is appropriate then we will just add our property using the same method as for adding the term, “learning object”. For our example we will choose, “has subject”. Making a wrong choice is not an issue, since we can consolidate our RDF representations by changing the property value at any time in the future. Mappings between different choices are also reflected in our new RDF representation, so as to accommodate for any previous knowledge conflicts. Having the property for our annotation selected (“has subject”) we need to define the object part of our triple statement. Assume that XML is not available as a term in WordNet or as an added term in CommonFolks. With no community recommendation, we are on our own, but we can search for a closely coupled term. “Markup language” is an existing term in WordNet, and seems like a good candidate. We receive confirmation by seeing that HTML already exists in WordNet and has “markup language” as a parent. By adding XML, the lattice is extended as in 1.iii. We can also, add a definitions for XML, which is just a RDF literal (text), this will used by others to decide if they like our definition and placement in the lattice for XML. The new statements describing the our learning object which is a XML tutorial are: <"http://example.org/lessonsXML-Intro"> <is a> <learning object> and <"http://example.org/lessonsXML-Intro"> <has subject> <XML>.

We have now seen an example of how CommonFolks facilitates the metadata creation process. We have also shown community support during authoring and have a method to move towards community consensus using our collaborative tagging techniques. Since we have our terms associated on a natural language framework, contrary to Shirky's earlier quote, we have "injected linguistic meaning" [15] and thus semantics, using a keyword-like tagging approach.

3.3 Prototype Assessment

We have created a prototype of the CommonFolks Tools approach to annotating metadata. It allows for the annotation of resources as described in section 3, without the collaborative suggestion functionality. The prototype was created to better assess if the approach would actually facilitate the creation of metadata annotations. We compared this method for creating metadata with that found in Aloha2,[5] which provides a form for annotating a learning object with IMS Learning Resource Metadata (IMS LRM - the precursor to the IEEE LOM). Our observations suggest that our method would be more intuitive for new users because of the availability of definitions and reference to the natural language lattice. Further, CommonFolks will provide the ability to extend metadata not represented in Aloha2 as it only represents the information provided in the IMS LRM. However, our prototype does not provide a reference to the requirements for standards such as the IMS LRM, but such standards can be predefined in CFT with the use of domains, as described in the last sub-section.

We are currently running a more detailed study on the kinds of keywords that instructors and learners associate with learning objects. In this study, we have collected a set of keywords from subject matter experts on content written in the domain of artificial intelligence. At the same time, a group of 100+ entry level university students are providing similar content-related keywords. We plan to qualitatively analyze the differences between these two groups to assess the usefulness of collective novice behavior versus limited expert opinion. In a second round of study we intend to introduce another set of students and instructors to the CommonFolks approach, and compare the quality of annotations created, the time it takes to create these annotations, and breadth of coverage of the annotations.

4 Future Work

4.1 Dealing with Breadth and Disagreements

We foresee two possible methods for dealing with the breadth of tags or annotations made by users. This breadth needs to be dealt with in a manner to allow

[5] http://aloha2.netera.ca/

users to receive the best possible support in using the system and contributing to the community of annotations. The first is to use collaborative filtering to make predictions for recommendations to users. While the second is to use meta-tagging to create domains and sub-domains (effectively sets and sub-sets used in recommendations).

It should be noted that within community disagreements are surely to arise. The goal is not necessarily to resolve disagreements between all parties but rather to acknowledge them and allow them to occur, as is the practice in collaborative tagging. CommonFolks then is not a way to create an authoritative ontology, rather it is a method to encourage consensus through use by annotators. Due to CommonFolks being a thesauri (as is WordNet), automatically consolidating differences by referencing the relations between terms (if they exist) is possible, and where no relation exists the level of community support can be referenced.

4.2 CommonFolks in RDF

The fundamental building block of web ontology languages is RDF. CommonFolks metadata will be written purely in RDF, using the Simple Knowledge Organization Scheme[6] (SKOS) or by extending an existing WordNet implementation using the Web Ontology Language[7]. Thus data interchange using CommonFolks and a way to fully express the instances, terms and mappings created in the CommonFolks tools will have strong compatibility with current adaptive e-learning systems.

5 Conclusion

In this paper we have introduced a framework, to facilitate human metadata creation for consumption by humans and machines. Having partially implemented CommonFolks Tools in a prototype (where annotations can simply be made using terms existing in WordNet) we are encouraged by the outlook for our continuing research in this area. The issue of creating high quality consistent metadata for learning resources in a timely fashion has plagued the e-learning area since the inception of the LOM as a standard. We have suggested a collaborative approach which leverages work done in human lexical memory to overcome the problems associated with formal ontology creation.

6 Acknowledgments

This project was funded in part by the Government of Canada through the LORNET NSERC research network.

[6] http://www.w3.org/2004/02/skos
[7] http://www.w3.org/2001/sw/BestPractices/WNET/wordnet-sw-20040713.html

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American **20** (2001) 34–43
2. Nejdl, W., Wolf, B., Staab, S., Tane, J.: Edutella: Searching and annotating resources within an rdf-based p2p network. In Frank, M., Noy, N.F., Staab, S., eds.: Proceedings of the WWW2002 International Workshop on the Semantic Web, Hawaii, May 7, 2002. Volume 55 of CEUR Workshop Proceedings., CEUR-WS.org (2002)
3. Dolog, P., Henze, N., Nejdl, W., Sintek, M.: The personal reader: Personalizing and enriching learning resources using semantic web technologies.. In Bra, P.D., Nejdl, W., eds.: Adaptive Hypermedia and Adaptive Web-Based Systems, Third International Conference, AH 2004, Eindhoven, The Netherlands, August 23-26, 2004, Proceedings. Volume 3137 of Lecture Notes in Computer Science., Springer (2004) 85–94
4. Lytras, M.D., Pouloudi, A., Korfiatis, N.: An ontological oriented approach on e-learning: Integrating semantics for adaptive e-learning systems. In: Proceedings of the 11th European Conference on Information Systems (ECIS 2003), Naples, Italy 16-21 June 2003. (2003) http://is2.lse.ac.uk/asp/aspecis/20030090.pdf.
5. Fellbaum, C., ed.: An Electronic Lexical Database. MIT Press (1998)
6. Mathes, A.: Folksonomies - cooperative classification and communication through shared metadata. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (2004)
7. Golder, S., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science **32** (2006) 198–208
8. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition **5** (1993) 199–220
9. Borst, W.N.: Construction of Engineering Ontologies. PhD thesis, University of Twente, Enschede, NL Centre for Telematica and Information Technology (1997)
10. IEEE Learning Technologies Standards Committee: IEEE LOM working draft 6.1. http://ltsc.ieee.org/wg12/index.html (2001)
11. Brooks, C., McCalla, G., Winter, M.: Flexible learning object metadata. In: 3rd International Workshop on Applications of Semantic Web Technologies for E-Learning, Amsterdam, The Netherlands (2005) Held in conjunction with the 12th International Conference on Artificial Intelligence in Education.
12. IMS Global Learning Consortium Inc.: IMS meta-data best practice guide for IEEE 1484.12.1-2002 standard for learning object metadata, version 1.3. (2004)
13. Greenberg, J., Robertson, W.D.: Semantic web construction: An inquiry of authors' views on collaborative metadata generation. In: Proceedings of the International Conference on Dublin Core Metadata for e-Communities, Florence, Italy, Firenze University Press (2002) 45–52
14. Guy, M., Tonkin, E.: Tidying up tags? D-Lib Magazine **12** (2006)
15. Shirky, C.: Ontology is overrated: Categories, links, and tags. http://www.shirky.com/writings/ontology_overrated.html (2005)
16. Gasevic, D., Hatala, M.: Searching web resources using ontology mappings. In: Integrating Ontologies. (2005)
17. Smith, M.K., Welty, C., McGuinness, D.L.: OWL Web Ontology Language Guide, W3C recommendation. http://www.w3.org/TR/owl-guide (2004)