Personal metadata for cataloguing microcontent – linking to large ontologies?

Daniel Rodríguez, Karsten Oster Lundqvist
S. Williams, K. Baker
Dept of Compt Science, The University of Reading (United Kingdom)

Abstract: Microlearning objects can be related by means of freely created annotations or tagging. Although such solution can help with filtering and searching, it is not enough for formalizing microlearning towards the semantic Web approach. In this paper, we propose the use of upper ontologies and more concretely, OpenCyc, to provide the required formal semantics needed by the semantic Web.

1. Microlearning

The tools to create microcontent have made easier to generate and publish information based on one's personal, subjective view of the world. Microlearning approaches may benefit from reusing microcontents in informal learning settings (Chisholm, 2005). This in turn calls for a reconsideration of how to find postings in common microcontent systems (Langreiter and Bolka, 2005). However, the volume of microcontent items being generated daily needs mechanisms to categorize them so that they can be more easily selected through information filtering technology. Standard metadata specifications as IEEE LOM could be used for that purpose, and public ontologies or taxonomies could be used for categorization. Furthermore, the subjective character of blogs and other microcontent systems suggests that kinds of "personal annotations" are required to retain the essence of microcontent which contrasts with approaches to metadata in which typically an information science professional or an specialized expert categorizes the content using common standards.

An approach to create a more personal and loosely organized cataloguing is „tagging" (as known from del.icio.us and Flickr). An arbitrary number of tags (lightweight keywords) can be assigned to every bookmark posted. What makes this service interesting from an information filtering perspective is the multitude of possibilities to filter bookmarks – by

user, by tag or even by a combination of tags (Langreiter and Bolka, 2005). Tagging can be organized through the metaphor of *folksonomies,* which have the required characteristics to integrate seamlessly with microcontent creation systems as blogs. Folksonomies are defined[1] as *"collaboratively generated, open-ended labeling system that enables Internet users to categorize content such as Web pages, online photographs, and Web links. The freely chosen labels – called also tags – help to improve search engine's effectiveness because content is categorized using a familiar, accessible, and shared vocabulary".*

In spite of its flexibility, tagging and folksonomies are not enough for becoming the raw material for a Semantic Web approach (Berners-Lee et al., 2001) for filtering and personalizing microcontent. The missing element is the formal approach provided by ontologies. Ontologies provide an explicit, shared representation of a domain providing unambiguous definitions for the main concepts and relations describing the phenomena under consideration (Gruber, 1995). Description logics (Baader et al., 2003) are well-known and thoroughly studied knowledge representations that can be used to develop formal ontologies and systems that use them, enabling reasoning services of a diverse kind. Formal ontologies thus provide a foundation for representing the diverse aspects of virtual communities and the artifacts they deal with in terms of ontologies, which would eventually result in advanced tools that are aware of the structural and cultural issues that are part of a community. Ontologies can be expressed in Web markup languages for ease of processing and interoperability. Notably, the OWL language has reached the status of W3C recommendation, and several tools are yet available to edit or process OWL representations. Thus, the problem with bringing the Semantic Web to microlearning can be approached by bridging with the actual practice of tagging to the formal annotations provided by logical ontological languages. This can be seen as a form of providing normative usage semantics to microcontent (Sánchez-Alonso and Sicilia, 2005). The key issue is doing that in a form that still retains the informal and easy way to tagging. This paper describes a possible approach for that issue that uses public large ontologies as the way to connect informal to formal semantics.

The rest of this paper is structured as follows. Section 2 deals with the form of micro-annotations and their interpretation in common interfaces. Then, Section 3 describes how such tags can be linked to large ontologies for reuse of common, shared semantics. Finally, conclusions and outlook are provided in Section 4.

1 http://en.wikipedia.org/wiki/Folksonomy (accessed Oct, 2006)

2. Typing in micro-annotations

The concept of annotation in the Semantic Web refers to the creation of semantic metadata, which essentially entails the creation of some statements that use the concepts and relations of one or several domain ontologies.

Euzenat (2002) formalized semantic annotation in the context of the Semantic Web as follows: from two sets of objects, documents and formal representations, two functions can be created: a *function from document to formal representations*, called annotation and a function from formal representations to documents called index. Thus, formal metadata is a requirement. According to the Wikipedia, "a folksonomy is a collaboratively generated, open-ended labeling system that enables Internet users to categorize content such as Web pages, online photographs, and Web links. The freely chosen labels – called tags – help to improve search engine's effectiveness because content is categorized using a familiar, accessible, and shared vocabulary". The flexibility of tagging is clearly incompatible with the approach to annotation in consistent and formal ways in the Semantic Web. However, a strategy for proving stricter semantics to tagging could be based on the following aspects:

- Provide easy ways for more detailed tagging.
- Provide mappings from folksonomies (systems of tags) to formal ontology.

This section deals with the first aspect. To illustrate the strategy, an example from del.icio.us, shown in figure 1, will be used.

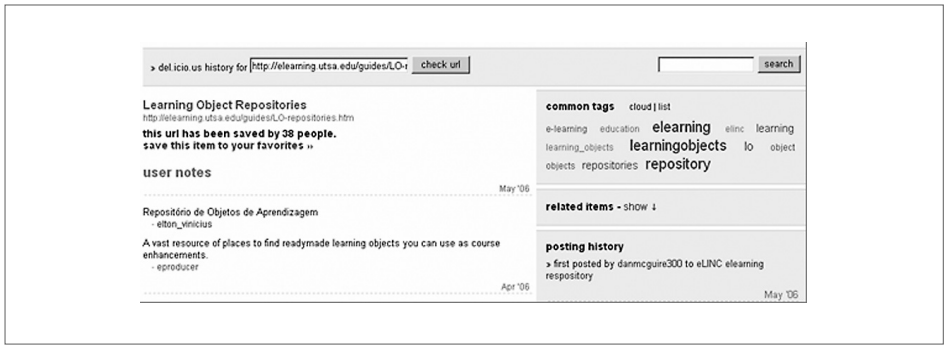

Figure 1: An example of tagging in del.icio.us

Tagging consists of placing a content (or reference to it) under a label. This entails that a single predicate is used, that in description logics form could be understood as a property of a class named generically **InternetResource** in the following form:

∀*classifiedIn.Tag.* This only provides a basic definition, but it introduces the restriction that the tagging process has an implicit meaning. An alternative may be that of using different "predicates". This is not a new concept, since it has been applied in research at least since the seminal work of Trigg (1986). Further, annotation tools like the *Annotea* technology (Kahan et al., 2002) can be easily integrated to carry RDF annotations as those to shared repositories in an easy way, integrated in popular browsers.
This basic enhancement has important implications for learning design, since the semantics of the different predicates can be used to aid in different pedagogies. For example, a position biased to reflective learning will seek for criticizes. Tag predicates. For example, the popular critical article of Wiley[2] on learning objects would be described with that predicate and thus make available to tools that are looking for controversial contents on a given topic. The interface in figure 1 will need to reflect the kind of predicate somewhat, e.g. several "common tag" clouds.

Further, predicates could be extended to link InternetResource instances, and not only as a mean to connect resources to tags. This way, for example, the same criticizes predicate could be applied to two microcontents, one replying to the other in a critical way – which is a very common pattern in weblogs. In figure 1 these links could be showed as decorations of the way links are displayed.

3. From tags to ontologies: linking to upper ontologies

The simple mechanism of differentiated predicates described above can then be complemented by other more complex approaches that add full formal semantics to tagging. A non-intrusive approach for this could be that of linking tags to concepts (or other ontology elements as instances) in shared ontologies. For a maximum reuse, large, commonsense ontologies and ontological semantic lexicons such as OntoWordNet described by Gangemi, Guarino, Masolo, Oltramari (2003) and W3C (2004) are obvious

2 http://opencontent.org/blog/archives/230 (accessed Oct, 2006)

candidates, even though a combination with more specific domain ontologies also could be used.

OpenCyc[3] is the open source version of the Cyc Knowledge Base (Lenat, 1994), which contains over one hundred thousands atomic terms. It is provided with an associated efficient inference engine, and it attempts to provide a comprehensive upper ontology of "commonsense" knowledge. In what follows, OpenCyc is used as an example of the potential of the approach proposed.

The non-intrusive approach can be summarized in the following:

- The tagging capabilities of the systems are retained, with the predicate typing described above.
- A separate tool is provided to connect tags in the available folksonomies to elements in the upper ontology.
- The tag-to-ontology connections are used for building "clouds" and/or finding related items.

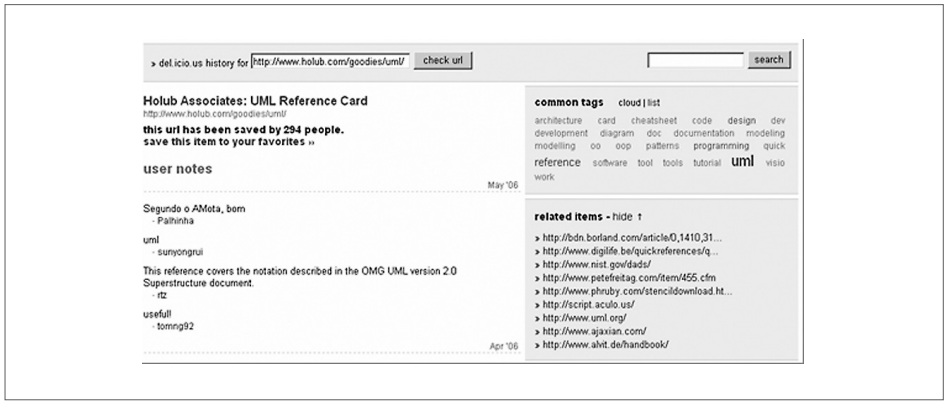

Figure 2: Another example of tagging in del.icio.us

The approach has the advantage that common microcontent creators can still use the easy tagging mechanism, and the formal mapping to the ontology can be done by other users, or even by specialists. Figure 2 shows an example of the approach. First, the *uml* tag can be directly mapped to the OpenCyc individual **oc_UnifiedModelingLanguage**,

3 http://www.opencyc.org/ (accessed Oct, 2006)

since they represent the language. This simple mapping (in this case, an equivalence), entails that the subsumers of the OpenCyc term, as **oc_ComputerLanguage**, could be used for finding related items.

In another direction, provided that a predicate **oc_umlElement** exists, there could be also navigation to links that explain different aspects of the language, as **oc_umlClass** or **oc_umlInstance**. Once this connection is made, the problem on how to traverse the ontology and/or find related links is a matter of user interface tactics in ontology browsing (García-Barriocanal and Sicilia, 2003). These kinds of mappings are a way to provide indirect formal semantics to informal folksonomies or collections of tags.

4. Conclusions and Future Work

The classification and organization of microcontent is one of the main challenges in turning microlearning into an effective pedagogy. Folksonomies and tagging can be used as subjective, informal classifications created openly by link collectors or micro-content creators. However, they do not provide the required formal semantics that Semantic Web applications need to build richer and more advanced information seeking technologies. A practical and non-intrusive approach to bringing a degree of formal semantics to tagging has been sketched in this paper. The approach is based on the use of predicate types, and the separate linking of tags to formal ontology elements.

Future work will continue in the direction of fully specifying the kind of annotations and mappings sketched here, and also in the direction of studying the user interface techniques that better fit the informal way of creating or collecting microcontents in current systems as weblogs.

5. References

Baader, F. et al (eds.). (2003). *The Description Logic Handbook. Theory, Implementation and Applications.* Cambridge.

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), pp. 34–43.

Chisholm, L. (2005). Micro-Learning in the Lifelong Learning Context. In *Proceedings of the Microlearning 2005 conference*, pp.5–6.

Euzenat J (2002) Eight Questions about Semantic Web Annotations. *IEEE Intelligent Systems* 17(2), pp. 55-62

García-Barriocanal, E. and Sicilia, M.A. (2003). User Interface Tactics in Ontology-Based Information Seeking. *Psychnology e-journal* 1(3),243–256.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. (2003) Restructuring WordNet's Top-Level, AI Magazine, Fall.

Gruber, T.R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human and Computer Studies*, 43(5/6), 907.

Kahan, J., Koivunen, M. R., Prud'Hommeaux and Swick, R. R. (2002) ANNOTEA: an open RDF infrastructure for shared Web annotations. *Computer Networks* 39, pp. 589–608.

Langreiter, C. and Bolka, A. (2005). Snips and Spaces: Managing microlearning. In *Proceedings of the Microlearning 2005 conference*, pp.79–97.

Lenat, D. B. (1995). Cyc: A Large-Scale Investment in Knowledge Infrastructure. Communications of the ACM 38(11), pp. 33-38.

Sánchez-Alonso, S. and Sicilia, M. A. (2005). Normative Specifications of Learning Objects and Learning Processes: Towards Higher Levels of Automation in Standardized eLearning. International Journal of Instructional Technology and Distance Learning, 2(3), 3–12.

Trigg, R. H. and Weiser, M. (1986) TEXTNET: A Network-Based Approach to Text Handling. ACM Transactions on Office Information Systems, 4(1), pp.1–23.

Sicilia, M. A. and Lytras, M. (2005). On the representation of change according to different ontologies of learning. International Journal of Learning and Change, 1(1), pp. 66–79.

W3C. (2004) Wordnet in RDFS and OWL, [online], Available: http://www.w3.org/2001/sw/BestPractices/WNET/wordnet-sw-20040713.html (accessed Oct, 2006)