

Folksonomies behind the scenes

Leyla Jael García-Castro¹, Alexander García²

¹ Universität der Bundeswehr München, Werner-Heisenberg-Weg 39,
85779 Neubiberg, Germany
w31blega@unibw.de

² University of Arkansas for Medical Sciences, Biomedical Informatics.
agarcia@uams.edu /alexgarcia@gmail.com

Abstract. In this position paper we analyze the similarities amongst folksonomies, semantic wikis, and ontology building; we also propose an alignment and orchestration of ontologies representing these scenarios. We argue that such alignment enables a more direct application of folksonomy-based approaches over these set-ups. The rationale behind folksonomies is shared across environments such as collaborative ontology building and semantic wikis. The three of them aim to facilitate knowledge sharing across communities; a social environment in which individual and community objectives are achieved supports them all.

Keywords: Ontology engineering, social web, semantic web, folksonomy, wiki

1 Introduction

Social tagging systems (STS) have become increasingly popular within the Web 2.0 era; they allow users to freely associate terms, *i.e.* tags, to resources. Tags support a variety of tasks such as information retrieval, personal organization strategies, and share-ability. Conceptual structures emerging from STS are known as folksonomies [1, 2]; they have been used mainly to improve the retrieval on tagged resources [3-5] as well as to discover shared conceptualizations and make explicit the semantic behind tags [6, 7]. Simplicity and immediate benefits for end users, *e.g.* bookmarks available online, are part of the rationale behind the fast adoption of STS [8].

Ontologies are shared conceptualizations that aim to represent an abstraction of a particular domain [9, 10]. Ontologies play a central role in Semantic Web because they are intended to enable data and information exchange in a machine-accessible format; establishing in this way common vocabularies and semantic interpretations of concepts [11]. Whilst agreements in folksonomies are implicit and mainly reached by common use and popularity; agreements in ontologies are explicit as well as documented and supported on evidences.

Wikis enable users to collaboratively create, share, and edit information via a browser interface, thus the final content is the result of everybody's effort [12]. Semantic Wikis, introduced in 2004 [13], aim to facilitate ontology content integration in Wiki as well as to support the evolution of knowledge: moving from term lists to logical constraints while granting users the freedom over the creative

process. Most of the existing semantic Wikis rely on RDF and mainly support subject-predicate-object structures [13]. Semantic Wikis exhibit a similar structure as folksonomies supporting both semantic annotations and selection within documents: a semantic annotation takes as subject a document or a portion of it and relates it to an object by means of a semantic qualifier, *e.g.* *skos:broader*. This structure could also be extracted from links in traditional Wikis; in this case the annotations would establish a *relatedTo* relation.

In this position paper we analyze the similarities amongst folksonomies, semantic wikis, and ontology building; we also propose an alignment and orchestration of ontologies representing these scenarios. We argue that such alignment enables a more direct application of folksonomy-based approaches over these set-ups.

2 Folksonomies behind the scenes

Annotation Ontology. The Annotation Ontology (AO) [14] represents the annotation process within social environments. AO is built upon the Annotea Project (<http://www.w3.org/2001/Annotea/>); it is also compatible with Newman's tagging ontology (www.holygoat.co.uk/projects/tags/), the Meaning of a Tag ontology (<http://moat-project.org/>), and the Simple Knowledge Organization System (SKOS) (<http://www.w3.org/2004/02/skos/>). AO supports both *free* as well as *semantic annotations*, namely qualified annotations in AO. It enables users to freely attach terms to resources –*free annotations*, as well as terms related to ontological entities – *semantic annotation*. Annotations can be attached to the entire resource as well as to portions of it, *e.g.* text, images, or tables. As annotations on specific parts of a document do not necessarily apply to the whole document, implementations should take care of it by enabling users to define whether or not such annotations should be also global. AO also supports the curation process over the annotations and offers different types of annotations such as notes, comments, erratum, etc. It offers provenance support by reusing the Provenance Authoring and Versioning ontology (PAV, <http://swan.mindinformatics.org/spec/1.2/pav.html>).

Collaborative Ontology Building. The Changes and Annotations Ontology (ChAO) [15] provides a model to track the modifications on ontology classes, properties and instances. It contains two main classes: *Change* represents changes –add, edit, and delete, in the ontology, and *Annotation* that stores related information such as comments, examples, explanations and votes. ChAO is currently in use by the collaborative Protégé project (http://protegewiki.stanford.edu/wiki/Collaborative_Protege). The ontology building process entails negotiation practices, *i.e.* ontologies are social agreements to accomplish shared objectives. When building ontologies, people are pursuing: (i) retrieving related information, (ii) sharing information, and (iii) improving and broadening both knowledge and performance. Interestingly, these are also the motivations for tagging resources [16]. We consider the ontology building process as a structured folksonomy in which the main document being tagged is that one representing the ontology; also, the participants are aware of the purpose of their contributions, *i.e.* annotations.

Mapping ChAO and AO makes it possible to use the flexibility of folksonomies

into the ontology building process; Fig. 1 shows the proposed mapping. Users in ChAO are identified by user names or accounts whilst AO uses *foaf:Agent* for that purpose; this brings benefits such as a unique URI to identify a contributor participating in different ontology developments, regardless of the methodology or editor. AO can also facilitate reusing information from other folksonomies as it is compatible with Newman's ontology, MOAT, and SKOS.

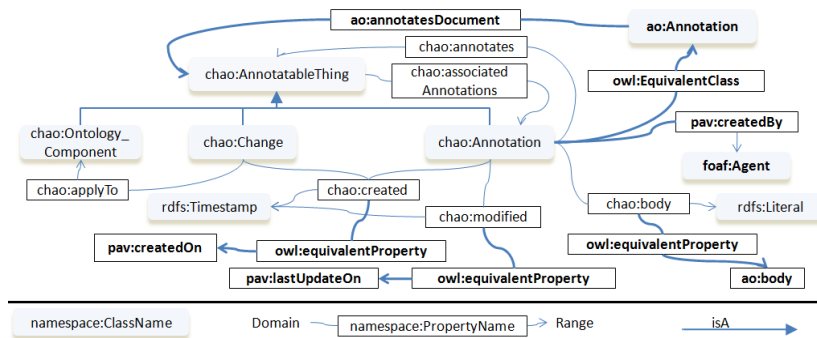


Fig. 1 AO and ChAO mapping

Fig. 2 shows an annotation from the ChAO and AO perspectives: a user named Daniel works on the Pizza ontology document; he creates a property *#hasTopping*, *chao:Property_Created*, on a *chao:Ontology_Component*, and adds an annotation, *chao:Annotation*, explaining why the property was created. From the AO perspective Daniel is represented as a *foaf:Agent*; he creates an annotation, *ao:Annotation*, corresponding to a creation, *ao:hasTopic*, on a portion of the ontology named *#hasTopping*, *ann:context*. In AO, the property *#hasTopping* is represented by means of XPointer (www.w3.org/TR/xptr/) selector, *i.e.* an element in an XML document.

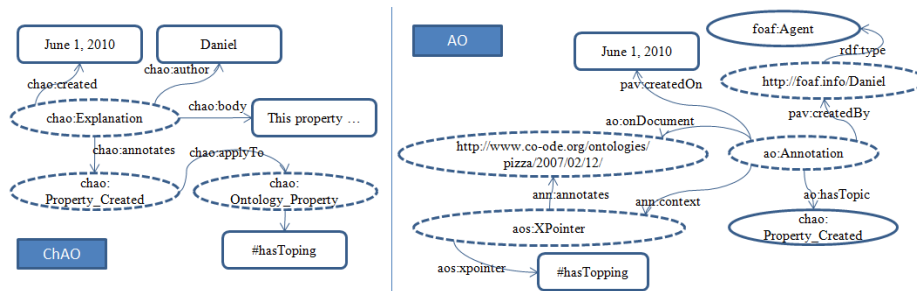


Fig. 2 Changes on an ontology component, ChAO and AO perspectives

During the ontology building, contributors perform activities such as adding, editing, and deleting ontological entities; those activities produce classes, properties or axioms that can be seen as portions of the ontology, easily identified by a URI. Annotations such as comments, notes, and votes are attached to particular entities. Consequently, it is possible to identify a contributor attaching annotations to pieces of

a document, *i.e.* the ontology; this process is fully supported by AO. Conjugation of ChAO, AO, PAV and FOAF makes it possible to use SPARQL in order to answer questions such as who has worked on this class, on which ontologies have contribute Andy and Tony, or what ontologies have been created for a particular domain.

Wikis. SweetWiki [17] proposes an ontology to represent the wiki structure; concepts include *document*, *page*, *tag*, *link*, *backward link*, *contributor*, *version*, *attached file*, etc. These concepts are found in both semantic and non semantic wikis; most of them are also covered by well known vocabularies such as Dublin Core, SKOS, SIOC, and FOAF. Semantic annotations within semantic wikis follow the structure subject-predicate-object [13]; by using AO, it is possible to model those annotations as qualified ones: the contributor corresponds to the annotator *-pav:createdBy* and *foaf:Agent*, the subject to the wiki page or a portion of it *-ao:onDocument* and *ann:context*, the topic to the type of the annotation *-skos:broader*, *dcterms:isVersionOf*, *sioc:attachment*, etc., and the object to the annotation *-ao:body*. AO currently supports semantic annotations corresponding to *skos:exactMatch*, *skos:closeMatch*, *skos:broader*, and *skos:narrower*; however, it is possible to extend AO in such a way that other qualifiers are also allowed. This extension is based on Hypertag [18] and consists of a new type of annotation *Relationship* that relates either two resources, or one resource as *subject* and the other one as *object*; it also enables both reusing relations, *e.g.* *dcterms:isVersionOf*, or creating new ones.

3 Conclusions and Future Work

We have presented an alignment between folksonomies and collaborative ontology building that facilitates collaboration across decentralized settings, pacing with dynamics on evolving domains, and monitoring the quality and consistency of the model by using the wisdom of crowds. The proposed alignments will likely facilitate the use of folksonomy-based approaches in wiki environments and vice versa as well as the knowledge emergence from both of them. Semantic annotations as well as provenance do not solve all semantic issues that folksonomies lack of; however, it reduces the gap between the social and the semantic web. The proposed alignment makes it easy to integrate knowledge gathered from social platforms into knowledge elicitation phases in ontology development methodologies. The alignment fits into methodologies with a collaborative component reusing non-structured or semi-structured existing knowledge such as Mature Project [19], NeOn Methodology [20], and Melting Point [21]. Integrated to the Mature Project, our approach makes it easier to consolidate and axiomatize the ontology built by the community as it includes semantic links; facilitating in this way the extraction of hierarchies as well as *ad hoc* relationships and mappings. Similarly, our approach facilitates reusing and reengineering non-ontological resources; one of the phases proposed by the NeOn methodology, as well as the conceptualization activity proposed in the Melting Point. The aforementioned methodologies reuse knowledge while our alignment facilitates to add and extract semantics from social environments. This combination makes possible the evolution of the extracted ontology as this one becomes part of the ontologies used to qualify annotation, thus new mappings and relations can emerge.

In this a way, users contribute to the ontology building process without being aware of the process in which they are taking part.

We have identified three scenarios that could benefit from such an alignment, improving the way that content from folksonomies is currently exploited. The first scenario belongs to the bioinformatics domain and is related to the collaborative annotation of proteins. In such a scenario, documents representing protein sequences are enhanced by semantic annotations that can be applied to the whole sequence or a portion of it as well as to other annotations on the protein. In this way, users provide content in the form of annotations, facilitating the publication of experimental data related to proteins. It also enables the immediate discovery of such information as annotations are modeled with AO and linked to protein specialized vocabularies; thus, it will be available as part of the Linked Open Data cloud. It will also facilitate ontology evolution by using an AO extension that enables the representation of relationships.

The second scenario belongs to the biological domain, it is related to annotations in laboratory notebooks. Tags4Labs [22] is a prototype supporting the annotation of experimental data for some of the processes routinely run at the Center for International Tropical Agriculture (CIAT) biotechnology laboratory. With the proposed alignment it will be also possible to use annotations as an enrichment mechanism for those ontologies being used to annotate experimental procedures. The third scenario belongs to the medical domain and is related with the annotation of medical images. Ceballos et al. [23] (<http://72.167.51.20:8888/webprotege/>) propose and environment in which medical images can be annotated with ontological terms or just by “tagging”. With the proposed alignment it will be also possible to enrich existing ontologies by capturing the evidence behind a “tag” so that ontology engineers can decide on the inclusion of the term in the ontology. Also, other users will be able to access such information, making it easier for them to evaluate the relevance of the term and its corresponding use.

Documents should be able to “know about” their own content for automated processes in order to “know what to do” with them. With the proposed alignment we aim to make it possible, *i.e.* both knowledge discovery and knowledge emergence.

References

1. Jaschke, R., Hotho, A., Schmitz, C., Ganter, B., Stumme, G.: Discovering shared conceptualizations in folksonomies. *Web Semantics: Science, Services and Agents on the World Wide Web* 6 (2008) 38-53
2. Helic, D., Strohmaier, M., Trattner, C., Muhr, M., Lerman, K.: Pragmatic evaluation of folksonomies. *International World Wide Web Conference*. ACM, Hyderabad, India (2011)
3. Begelman, G., Keller, P., Smadja, F.: Automated Tag Clustering: Improving search and exploration in the tag space. *World Wide Web Conference - Collaborative Web Tagging Workshop*, Scotland (2006)
4. Heymann, P., Garcia-Molina, H.: Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems Stanford University (2006)
5. Yeung, C.A., Gibbins, N., Shadbolt, N.: Understanding the Semantics of Ambiguous Tags in Folksonomies. *International Workshop on Emergent Semantics and Ontology Evolution*, Korea (2007)

6. Angeletou, S., Sabou, M., Specia, L., Motta, E.: Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. European Semantic Web Conference - Bridging the Gap between Semantic Web and Web 2.0 Workshop, Austria (2007)
7. Van Damme, C., Hepp, M., Siorpaes, K.: FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. European Semantic Web Conference - Workshop "Bridging the Gap between Semantic Web and Web 2.0", Austria (2007)
8. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking. *The Semantic Web: Research and Applications* (2006) 411-426
9. Gruber, T.: What is an Ontology? , Vol. 2009 (1992) Retrieved May. 23, 2009, from <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
10. Guarino, N., Poli, R., Gruber, T.: Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Workshop on Formal Ontology*. Kluwer Academic Publishers (1993)
11. Gendarmi, D., Lanubile, F.: Community-Driven Ontology Evolution Based on Folksonomies. On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Vol. 4277. Springer (2006)
12. Schaffert, S., Gruber, A., Westenthaler, R.: A Semantic Wiki for Collaborative Knowledge Formation. *Semantics 2005*, Vienna, Austria (2005)
13. Kuhn, T.: AceWiki: A Natural and Expressive Semantic Wiki. *Semantic Web User Interaction*, Florence, Italy (2008)
14. Ciccarese, P., Ocana, M., Garcia Castro, L.J., Das, S., Clark, T.: An Open Annotation Ontology for Science on Web 3.0. *BmC Bioinformatics* (accepted) (2010)
15. Noy, N., Chugh, A., Liu, W., Musen, M.: A Framework for Ontology Evolution in Collaborative Environments (2006)
16. Braun, S., Schmidt, A., Walter, A., Nagypal, G., Zacharias, V.: Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering. *International World Wide Web Conference - Workshop on Social and Collaborative Construction of Structured Knowledge (CKC)*, Canada (2007)
17. Buffa, M., Gandon, F.: SweetWiki: semantic web enabled technologies in Wiki. *International Symposium on Wikis*. ACM, Odense, Denmark (2006)
18. García-Castro, L.J., Hepp, M., García, A.: Tags4Tags: Using Tagging to Consolidate Tags. *International Conference on Database and Expert Systems Applications*, Linz, Austria (2009)
19. Braun, S., Schmidt, A., Walter, A., Nagypal, G., Zacharias, V.: Ontology Maturing: a Collaborative Web 2.0 Approach to Ontology Engineering. *International World Wide Web Conference - Workshop on Social and Collaborative Construction of Structured Knowledge (CKC)*, Canada (2007)
20. Suárez-Figueroa, M., Dellschaft, K., Montiel-Ponsoda, E., Villazón-Terrazas, B., Yufei, Z., Aguado-de-Cea, G., García, A., Fernández-López, M., Gómez-Pérez, A., Espinoza, M., Sabou, M.: NeOn Methodology for Building Contextualized Ontology Networks (NeOn Deliverable D5.4.1.). NeOn Project (2008)
21. Garcia, A., O'Neill, K., Garcia, L.J., Lord, P., Stevens, R., Corcho, O., Gibson, F.: Developing Ontologies within Decentralised Settings. In: Chen, H., Wang, Y., Cheung, K.-H. (eds.): *Semantic e-Science*, Vol. 11. Springer US (2010) 99-139
22. Garcia Castro, A., Giraldo, O., Garcia Castro, L.J.: Annotating experimental records using ontologies. *International Conference on Biomedical Ontology*, Buffalo, NY, USA (2011)
23. Ceballos, O., Garcia Castro, A., Garcia Castro, L.J., Millan, M.: Anotación semántica de imágenes médicas. *Acta Biológica Colombiana* 15 (2010)