INTEGRATING SOCIAL WEB WITH SEMANTIC WEB: ONTOLOGY LEARNING AND ONTOLOGY EVOLUTION FROM FOLKSONOMIES

Ademir Roberto Freddo, Cesar Augusto Tacla
Universidade Tecnológica Federal do Paraná, Av. Sete de Setembro 3165, Curitiba, Paraná, Brasil
freddo@utfpr.edu.br, tacla@utfpr.edu.br

Keywords: Social Web, Semantic Web, Ontology Learning, Ontology Evolution, Folksonomy, Ontology Alignment.

Abstract: In this paper, we present an approach for integrating Social Web with Semantic Web by combining the easiness of annotation of resources in the Social Web and the expressiveness of ontologies to describe the resources in the Semantic Web. Our approach combines ontology learning and ontology evolution techniques to provide an integrated Web. Besides, we show how ontology alignment can be used to enrich ontologies in this context.

1 INTRODUCTION

The Semantic Web requires resources to be annotated with machine understandable metadata (Berners-Lee et al, 2001) being ontologies the knowledge representation technique most used nowadays to describe such metadata.

Even knowledge engineers and experts have some difficulty of maintaining consistency between resources and the ontologies. It is necessary to catch the changes of web resources and keep update the ontology. Besides, the lack of an imagined class or wrong classification when users annotate a Web resource with applications built on ontologies is a recurrent problem.

The knowledge acquisition bottleneck has limited ontology use. Ontology development starts with an initial ontology which is later revised, refined and filled with details (Heflin et al, 1999; Noy and McGuinness, 2001). Besides, new information which was previously unknown or unavailable needs to be added to the initial ontology. Nowadays changes to ontologies have to be captured and introduced by knowledge engineers (Zablith et al, 2008).

On the other hand, in the Social Web, social tagging systems such as Flickr (*http://www.flickr.com*) for photo sharing, and delicious (*http://delicious.com*) for social bookmarking are becoming more popular in the Web. The reason for their immediate success is the fact that no specific skills are needed for annotating. Users annotate, assign tags (any keyword, label), and categorize web resources easily and freely without using or even knowing taxonomies or ontologies. In the social tagging systems, user's resources and associated tags constitute the personomy. The collection of personomies constitutes a folksonomy (Jaschke et al, 2008). The folksonomy is dynamic as long as users learn new things and review their personomies, including and excluding their tags. Creating personomies is easy and does not require expert users. However, it let users to introduce ambiguities. Thus, in folksonomies content retrieval activities such as searching are limited, because results can present low recall and precision. So, in the Social Web the meaning of the tagging data has limited useful reasoning with the data.

1.1 Folksonomy and Ontology

Folksonomy is created by open and uncontrolled systems (social tagging systems) where users can annotate resources with different tags depending on their social or cultural backgrounds, expertise and perception of the world (Belelman et al, 2006; Golder and Huberman, 2005; Peterson, 2006; Wu H. et al, 2006).

Ontology is "an explicit and formal specification of a conceptualization" (Gruber, 1993). Ontologies specify common conceptualizations, independent of data model, so people can align their systems semantically by adopting ontologies.

In contrast to ontology, folksonomy does not explicitly state shared conceptualizations. It is a free-form annotation of web resources, done by users, and without the constraints of a predefined taxonomy or ontology (Wu H et al, 2006). Thus, in the Social Web, the meaning of the tagging data is completely unspecified. We need technology for reasoning with the folksonomies in such a way that computations can discover and conclude new things (Gruber, 2008).

The integrated Web combines the facilities in annotating web resources with tags that characterizes Social Web applications with ontologies to better describe the resources in the Semantic Web. According to Tim Berners-Lee (Berners-Lee et al, 2001) "the Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”.

1.2 Contribution and organization of the paper

Our approach aims at extracting structured data (ontologies) from unstructured data (folksonomies). As result of our approach, we obtain an ontology whose elements are linked to their source tags enabling to trace folksonomy changes back to the ontology.

The folksonomies in Social Web can provide new impulse to Semantic Web in by reducing the burden on users and engineers in tasks related to knowledge engineering (e.g. knowledge acquisition, Web resource annotation, and ontology construction). On the other hand, the Semantic Web can improve inferences and provide better query results in Social Web.

The remainder of the paper is organized as follows. Section 2 describes the basic definitions and current developments in ontology learning and ontology evolution from folksonomies. In section 3, we describe the approach to ontology learning and evolution. Section 4 presents a case study. Finally, Section 5 gives a conclusion and presents envisaged works.

2 STATE OF THE ART

In this section we describe the state of the art in ontology learning and evolution.

2.1 Ontology learning

According to Maedche and Staab (2001), “ontology learning greatly facility the construction of ontologies by the ontology engineer. It is a task of (semi)-automatically construct an ontology by using machine learning or data mining algorithms that are applied on data”.

In folksonomies, the use of tags by people with common interests tends to converge to a shared vocabulary. Following this intuition, a variety of approaches have been proposed to discover shared conceptualizations that are hidden in a folksonomy. Some of them (Schmitz, 2006; Wu et al., 2006; Belelman et al., 2006; Mika, 2005) analysis the co-occurrence of tags. Schmitz (2006) finds candidate subsumption relations, Wu et al (2006) and Belelman et al. (2006) create clusters of tags, and Mika (2005) builds graphs relating tags. However these cited approaches focus on finding groups of related tags rather than identifying the semantics of those relations.

Another set of recent approaches (Angeletou et al, 2007; Basso and da Silva, 2008; Specia and Mota, 2007) build ontologies from folksonomies thus going beyond the mentioned approaches that identify implicitly inter-related tags. Angeleton et al (2006) propose a method to enrich the tag space of folksonomies by exploring ontologies. Basso and da Silva (2008) presents a proposal for the ontology construction/evolution from the folksonomies based on WordNet. Specia and Mota (2007) propose the integration of folksonomies and ontologies to enrich tag semantics – identify semantic relation between tags using ontologies available on the semantic web.

These approaches has some limitations: difficulty in finding online ontologies; syntactic mapping to link tags to ontology concepts; different ontologies reflect different views which often lead to contradictory; limited reasoning because of plain structure of folksonomies; and they not focus on the ontology evolution.

In social tagging systems (e.g. *http://www.flickr.com, http://delicious.com*) there is no formal semantic and no formal agreement on the representation of the tagging process. This means that every system uses a different format to publish its tagging data.

Thus, other approaches (Knerr, 2006; Gruber, 2007, Kim et al, 2007) propose a solution for tag data representation. Gruber’s conceptual model (Gruber, 2007) describes tagging as a relation between an object (resource – bookmark, picture), a tag, a tagger (a person or agent that created the link between the tag and the object) and a source (the space where the tagging action has been performed – flickr, delicious). Knerr (2006) develop an ontology using the tripartite tagging (user, resource, tag) model Newman et al, (2005). The Social Semantic Cloud of Tags (SCOT) (Kim et al, 2007) ontology aims to describe the structure and the semantics of tagging data and to offer social interoperability of the data among heterogeneous sources.

2.2 Ontology evolution

Ontology evolution is defined by (Haase and Stojanovic, 2005) as the "timely adaptation of an ontology to the arisen changes and the consistent management of these changes". Ontology evolution is a process that supports the enrichment of the ontology by adding new entities (concepts, properties, and instances) or by modifying existing entities when new knowledge is acquired.

Ontology changes may come from explicit and implicit requirements (Cimiano and Volker, 2005). Explicit requirements are generated by ontology engineers. Implicit requirements are reflected in the behavior of the system and can be induced by applying change discovery methods from existing data (Cimiano and Volker, 2005; Stojanovick, 2004). Stojanovick (2004) defines three types of change discovery: structure-driven deduces changes from the ontology structure itself, usage-driven identifies changes from the usage patterns creates over a period of time, and data-driven generates changes by modifications to the underlying data (text documents, database) that represents the knowledge modeled by an ontology.

The current state of the art in ontology evolution, as well as a list of existing tools that aid the process can be found in (Haase and Sure, 2004). Some of these approaches are simple ontology editors, like Protégé (Noy et al, 2000) and OilEd (Bechhofer et al, 2001). One set of approaches (Alani et al, 2006; Bloehdorn et al, 2006; Novacek et al, 2007; Novacek et al, 2008, Ottens and Glize, 2007; Cimiano and Volker, 2005; Zablith, 2007, Zablith et al, 2008) in ontology evolution identify potential novel information that should be added to the ontology by exploiting the changes occurring in the various data sources. Such approaches do not consider folksonomies as an information source where ontology changes can be discovered. In this work, this is exactly the case, as we focus on change capturing from data-driven implicit requirements. Our collaborative approach for ontology evolution starts with an existing ontology (base or domain ontology) and supports the acquisition of new knowledge from folksonomies when users change their tags as long as their vocabularies change. We evolve ontologies exploiting folksonomy versioning and linking learned ontology entities to the source tags in the folksonomy.

3 SOCIAL WEB WITH SEMANTIC WEB

This section presents the approach to ontology learning and evolution from folksonomies.

3.1 Ontology learning from folksonomies

The ontology learning process from folksonomies is articulated in the following phases (Figure 1): populating the tag ontology, identifying relations between tags, and interacting with user.

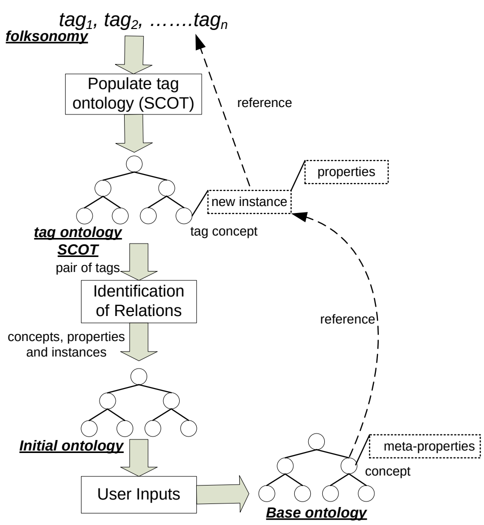

Figure 1. Ontology learning from folksonomies.

The populating tag ontology phase consists of building a representation for folksonomy tags based on entities defined in SCOT, an ontology to describe the structure and the semantics of folksonomies.

The SCOT ontology uses concepts and proprerties of Newman's model (Newman et al, 2005).

After *populating SCOT ontology*, we identify for each pair of tags, $((tag_1,tag_2), (tag_1,tag_3), (tag_1,tag_n), (tag_2,tag_3), (tag_2,tag_n), (tag_3,tag_n))$, the tag type (whether concept or instance), and the relation between them based on properties defined in SCOT such as textual description, synonym, and spelling variant. In order to identify tag types, we use Text2Onto (Cimiano and Volker, 2005). For tag relations, we use WordNet (Miller, 1990) for identifying meronyms, hyponyms, synonyms, and hyperonyms, and the Hwang's work (2007) for performing analysis of classes' hierarchy.

A tag can be described as an instance or a concept in the base ontology. Thus, it is possible to obtain the following types of pairs: (*instance*, *instance*), (*instance*, *concept*), and (*concept*, *concept*). In pairs of type (*instance*, *concept*), we verify if the tag instance can be considered an instance of the concept tag. For example, for the pair of tags (*Porto*, *University*), *university* is the concept and *Porto* is an instance (*University of Porto*). So, we add to the base ontology the concept *University* and the instance *Porto*. For pairs (*concept*, *concept*), we identify meronym, hyponym, and hyperonym relations. Finally, for pairs (*instance*, *instance*), we identify the concepts related to the instances. In this case we create a new concept or use an already identified concept. The approach to ontology learning from folksonomies is semi-automatic. It suggests to the user concepts, instances, properties, and relations between concepts. In the *interacting with user* phase, the user takes all decisions concerning the creation of concepts, instantiation of concepts and relations. After the user decision taking, the base ontology is created. Each entity in the base ontology is linked by means of the reference metaproperty to the source tag in order to maintain the traceability with the folksonomy.

3.2 Ontology evolution from folksonomies

Figure 2 shows our approach to evolve ontologies from folksonomies. A folksonomy is produced by user tagging activity in any social tagging system. Tags represent the domain according to the users' perspective and they are formally described using SCOT ontology. The task ontology contains the rules to extract data from folksonomies and to populate the SCOT. The base ontology is created by ontology engineers or some ontology learning process. Our purpose is to respond to changes in the folksonomies to update the base ontology. If the folksonomy is changed, the base ontology may also be modified. Our approach starts with reading a folksonomy in order to extract new and relevant information to be added to or removed from the base ontology. As the SCOT is populated with information from the folksonomy, the addition or removal of instances means possible changes to be done in the base ontology.

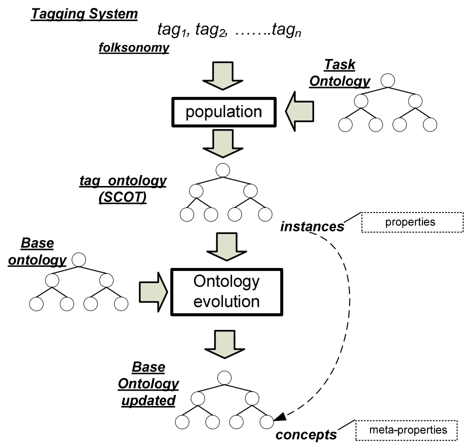

Figure 2. Approach to ontology evolution.

3.2.1 Add concepts in base ontology

When new tags are added to the folksonomy, the base ontology may have to be updated with new entities. After identifying the changes, we identify similarity relations between the extracted tags and the entities in the base ontology (Figure 3) using an ontology alignment method. With this method, we know which parts of the base ontology are affected by the changes in the folksonomy.

Besides, we identify the proper position where the new entity should be added. In this work we use the partial ontology alignment method named *Partial Ontology Alignment Method* - POAM (Freddo et al., 2007). For example, a new instance and its associated properties in the SCOT ontology are compared to the other SCOT's instances obtaining a set of similar instances. We search in the base ontology the concepts linked to the instances in the set by the reference metaproperty. At this moment, the user has to decide which relation the new instance has with the linked concepts of the base ontology: none, subsumption, equivalence, sibling or instance of.

3.2.2 Remove concepts in base ontology

Each concept in the base ontology is referenced by one or more instances of SCOT's concepts. When a tag is removed from the folksonomy, we verify which concepts are linked to this tag. If the tag is linked only to one concept in the base ontology, we suggest to the user to remove the concept from the base ontology. If the tag is linked to two or more concepts in the base ontology, we do not remove the concept. We only update the reference meta-property. As in the ontology learning, this approach to ontology evolution creates suggestions for changes which may be help knowledge engineers in taking final decisions.

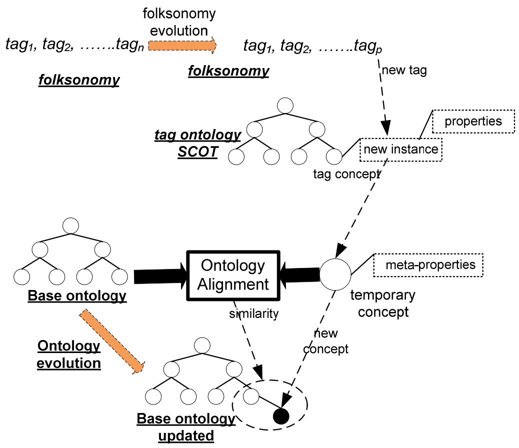

Figure 3. Ontology evolution and ontology alignment.

4 CASE STUDY

A case study is presented to demonstrate the complete approach comprising ontology learning and evolution.

4.1 Case study to ontology learning from folksonomies

We demonstrate the ontology learning with the tags *hotel*, *accommodation*, *room*, *luxury*, *Paris* were retrieved from some folksonomy in the tourism domain. We use these tags to populate the SCOT ontology. Figure 4 shows the SCOT ontology where each instance of the tag concept has some properties such as *reference_tag*, *equivalentTag*, *spelling_variant*, *used_by*, and *associatedTag_in*. The *reference_tag* property maintains the traceability between a tag in the folksonomy and the instance of the tag in SCOT.

After populating the SCOT ontology, we identify the relation between pair of tags. In this example, we have the following pairs: (*hotel*, *accommodation*), (*hotel*, *room*), (*hotel*, *luxury*), (*hotel*, *Paris*), (*accommodation*, *room*), (*accommodation*, *luxury*), (*accommodation*, *Paris*), (*room*, *luxury*), (*room*, *Paris*), and (*luxury*, *Paris*).

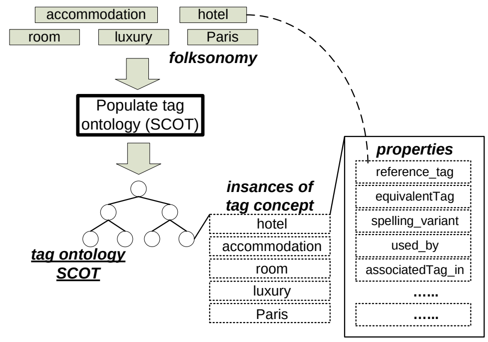

Figure 4. Instances of the SCOT's tag concept.

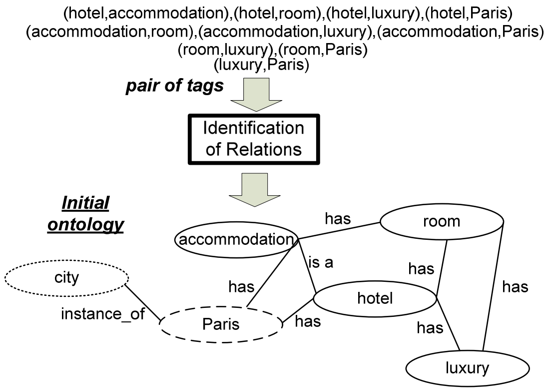

Figure 5. Initial ontology.

Figure 5 shows the initial ontology with some relations such as *instance_of*, *has*, and *is a*. Each tag can be an instance or concept in the initial ontology.

With the suggestions represented in the initial ontology (Figure 5), the user builds the base ontology. In the base ontology, each concept has properties and meta-properties (properties in SCOT are meta-properties for concepts in the base ontology and temporary concept). For instance, the concept *hotel* has properties *hasCity*, *hasRoom*, and *hasClassification*. The same concept has meta-properties *reference_tag*, *equivalentTag*, *spelling_variant*, etc.

4.2 Case study to ontology evolution from folksonomies

We demonstrate the ontology evolution approach with the inclusion of new tags (*hostel*, *address*) in folksonomy (Figure 6). The new tags are instances of SCOT's tag concept. Based on the properties defined for the tag concept in the SCOT ontology, each tag can be seen as a temporary concept with meta-properties. Thus, we align hostel and address with concepts in the base ontology based on the POAM's similarity value computed with several metrics including property similarity ones. The POAM detects the following alignments: hostel and hotel, and address and city (Figure 7).

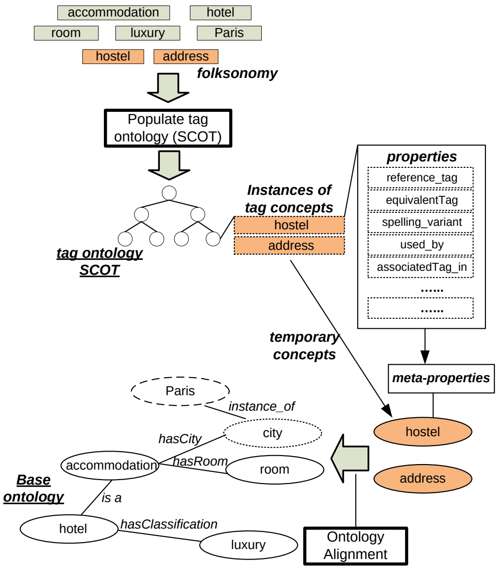

Figure 6. Evolving an ontology from new tags in the folksonomy.

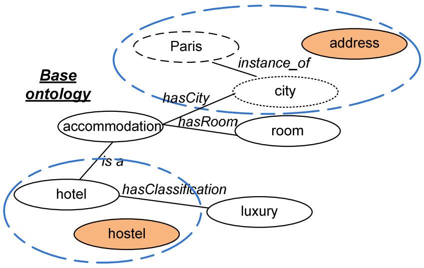

Figure 7. Similarity between temporary concepts and concepts in the base ontology.

Hostel and address are temporary concepts that were aligned with concepts city, accommodation, room, hotel and luxury in base ontology. Based on the found alignments, the user knows the proper position where the concepts *hostel* and *address* are added. Then we identify the relations between following pair of tags: *(hostel, hotel), (hostel, accommodation), (address, city).*

5 CONCLUSIONS

In Social Web, we have users building their personomies online. However, the meaning of tags is completely unspecified. Ontologies can describe semantically such data.

By combining the facilities in annotating Web resources in the Social Web and the expressiveness power of Ontologies to describe resources in the Semantic Web, we could provide an integrated Web.

In this work we describe an approach for combine the Social Web and the Semantic Web. According to Gruber (2008), "the challenge for the next generation of the Social and Semantic Web is to find the right match between what is put online and methods for doing useful reasoning with the data".

Motivated by the challenges of ontology engineering and inspired by the success of social web applications, we presented an approach to ontology learning and evolution from folksonomies. We use ontology alignment to support ontology enrichment (add or remove entities) when changes are detected in the folksonomy.

The implementation of this approach is currently in progress. In a near future, we intend to evaluate it analyzing the inferences that become possible with the integration, the precision and recall in queries, and the degree of user overloading with tasks related to knowledge engineering related to the ontology development and evolution.

ACKNOWLEDGENT

The authors thankfully acknowledge the financial support granted for this study by Fundação Araucária (AMO 9902 – conv. 234/2007).

REFERENCES

Alani, H., Harris, S., O'Neil B. 2006. Winnowing ontologies based on application use. *In Proceedings of 3rd European Semantic Web Conference (ESWC)*, Montenegro.

Angeletou, S., Sabou, M., Specia L., Mota, E. 2007. Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. *In: Workshop Bridging the Gap between Semantic Web and Web 2.0 - 4th European Semantic Web Conference, Insbruck. Workshop Bridging the Gap between Semantic Web and Web 2.0, p. 30-43.*

Basso, C. A. M., da SILVA, S. R. P. 2008. Uma Proposta para a Evolução de Ontologias a partir de Folksonomias. *In: Webmedia - Workshop de Teses e Dissertações, 2008, Vila Velha - ES. Webmedia. Sociedade barsileira de Computação, v. 1. p. 197-200.*

Belelman, G., Keller, P., Smadja F., 2006. Automated Tag Clustering: Improving search and exploration in the tag space. In *Proc. of the Collaborative Web Tagging Workshop*.

Berners-Lee, T., James H., Ora L., 2001. The Semantic Web. Scientific American Magazine. http://www.sciam.com/article.cfm?id=the-semantic-web&print=true. Retrieved on 2008-03-26.

Bloehdorn, S., Haase, P., Sure, Y., Voelker J. 2006. Ontology Evolution. In: *John Davies, Rudi Studer and Paul Warren, Semantic Web Technologies - Trends and Research in Ontologybased Systems*. pp. 51-70. John Wiley & Sons.

Cimiano, P., Volker, J., 2005. Text2Onto - a Framework for Ontology Learning and Data-driven Change Discovery. *In:Proc. NLDB,* LNCS 3513, pp. 227-238, Springer, Alicante.

Freddo, A. R., Brito, R. C., Gimenez-Lugo G., Tacla, C. A., 2007. Partial and Dynamic Ontology Mapping Model in Dialogs of Agents. *EPIA Workshops*: 347-356. Services. AI Magazine 26(1): 71-82.

Golder S., Huberman B.A., 2005. The Structure of Collaborative Tagging Systems. HPL Technical Report.

Gruber, T. R., 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing., *In Formal Ontology in Conceptual Analysis and Knowledge Representation*, Nicola Guarino et Roberto Poli (eds.), Kluwer Academic Publishers, International Workshop on Formal Ontology, Padova, Italy.

Gruber T. R., 2007. Ontology of folksonomy: a mash-up of apples and oranges. *Intl Journal on Semantic Web & Informatin Systems*.

Gruber, T. R., 2008. Collective knowledge systems: Where the Social Web meets the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*. Volume 6, Issue 1, Semantic Web and Web 2.0, p. 4-13.

Haase, P., Sure, Y., 2004. D3.1.1.b State of the Art on Ontology Evolution Available at: http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/SEKT-D3.1.1.b.pdf

Haase, P., Stojanovic, L. 2005. Consistent Evolution of OWL Ontologies. In: *Proceedings of the 2nd European Semantic Web Conference*. Heraklion, Greece.

Heflin, J., Hendler, J., Luke, S., 1999. Coping with Changing Ontologies in a Distributed Environment. In *Proceedings of the Workshop on Ontology Management of the 16 National Conference on Artificial Intelligence*, pp. 74—79.

Hwang Suk-Hyung. 2007. A Triadic Approach of Hierarchical Classes Analysis on Folksonomy Mining. Intl. Journal of Computer Science and Network Security (IJCSNS).

Jaschke R., Hotho, A., Schmitz C., Ganter, B., Stumme G., 2008. Discovering shared conceptualizations in folksonomies. In *Web Semantics: Science, Services and Agents on the World Wide Web*. vol. 6, Issue 1, Semantic Web and Web 2.0, p. 38-53.

Kim H. L., Yang S. K., Breslin, J. G., Kim H. G., 2007. Simple algorithms for representing tag frequencies in the scot exporter. *In IAT, pages 536-539. IEEE Computer Society*.

Knerr, T., 2006. Tagging Ontology- Towards a Common Ontology for Folksonomies. Available at: http://tagont.googlecode.com/files/TagOntPaper.pdf.

Maedche, A., Staab, S. 2001. Ontology learning for the semantic web. IEEE Intell. Syst. 16 (2) 72–79.

Mika, P., 2005. Ontologies are us: A unified model of social networks and semantics. *Proceedings of the 4th International Semantic Web Conference (ISWC*), pp.522--536.

Miller, G.A. 1990. WordNet: An on-line lexical database. In *International Journal of Lexicography* 3(4), 235–312.

Newman R., Ayers, D., Russell. S. 2005. Tag ontology, Available at: http://www.holygoat.co.uk/owl/redwood/0.1/tags/.

Novacek, V., Laera, L., Handschuh, S. 2007. Semi-automatic integration of learned ontologies into a collaborative framework. *In International Workshop on Ontology Dynamics (IWOD)*.

Novacek, V., Laera, L., Handschuh, S., Zemanek, J., Volkel, M., Bendaoud, R., Hacene, M. R., Toussaint, Y., Delecroix, B., Napoli. A., 2008. D2. 3.8 v2 report and prototype of dynamics in the ontology lifecycle.

Noy, N. F., McGuinness, D., 2001. Ontology development 101: a guide to creating your first ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Ottens K., Glize. P., 2007. A multi-agent system for building dynamic ontologies. *In Proceedings of AAMAS*.

Peterson E., 2006. Beneath the Metadata: Some Philosophical Problems with Folksonomy. D-Lib Magazine, 12(11).

Schmitz, P., 2006. Inducing ontology from Flickr tags. *In Collaborative Web Tagging Workshop, 15th WWW Conference*, Edinburgh.

Specia, L., Motta, E., 2007. Integrating Folksonomies with the Semantic Web. *In: European Semantic Web Conference (ESWC 2007)*, The Semantic Web: Research and Applications. LNCS 4519, pp. 503—517, Springer. Heidelberg.

Stojanovic, L., 2004. Methods and Tools for Ontology Evolution. Phd thesis, University of Karlsruhe.

Wu, X., Zhang, L., Yu, Y., 2006. Exploring social annotations for the semantic web. *In 15th WWW Conference*, Edinburgh.

Wu H., Zubair, M., Maly. K., 2006. Harvesting Social Knowledge from Folksonomies. *In Proc. of HYPERTEXT*.

Zablith F., 2007. Dynamic ontology evolution. In: International Semantic Web Conference (ISWC) Doctoral Consortium, Karlsruhe, Germany.

Zablith, F., Sabou, M., d'Aquin, M., Motta, E., 2008. Using Background Knowledge for Ontology Evolution. Workshop: International Workshop on Ontology Dynamics (IWOD). The International Semantic Web Conference (ISWC), Karlsruhe, Germany.