

# TOWARDS A MULTILINGUAL SEMANTIC FOLKSONOMY

Murad Magableh, Antonio Cau, Hussein Zedan, Martin Ward  
Software Technology Research Laboratory (STRL)  
Faculty of Technology  
De Montfort University  
The Gateway, Leicester LE1 9BH  
United Kingdom  
{mmurad, cau, hzedan, mward}@dmu.ac.uk

## ABSTRACT

The content of collaborative tagging systems (so-called folksonomies) is generated, consumed, and annotated by the end-users. Users annotate and categorise their data using free-keywords, so-called tags. Consequently, several linguistic problems come to the surface in folksonomies such as; synonyms, polysemy, multilinguality, and others which produce ambiguous and inconsistent classification of data. Therefore, relevant results are not retrieved in the user's query. In this paper, we suggest a novel approach to enhance the “*social vocabulary*” presented in folksonomies with the “*controlled vocabulary*” presented in Semantic Web ontologies. Therefore, our proposed approach uses the online WordNet lexical ontology in addition to the EuroWordNet multilingual lexical resource. Our approach tries to employ the ontological relations presented in WordNet in the folksonomy, it focuses only on the problems of synonyms, tag relations, and multilinguality.

## KEYWORDS

Social Web, Semantic Web, Collaborative Tagging System, Folksonomy, WordNet Ontology, EuroWordNet.

## 1. INTRODUCTION

By introducing Web 2.0 (Social Web), end-users became at the heart of Web content generation and classification processes. In collaborative tagging systems (folksonomies), users generate contents and they use free-text keywords, so-called tags, to classify their contents. Therefore, users create metadata as well as data. This new approach of data categorisation and metadata creation is simple, easy, fast, low cost, and flexible compared to traditional metadata creation process by professionals and authors. Nevertheless, lack of semantics among data in such communities represents a real challenge regarding the information retrieval.

The ethos of Semantic Web vision is to represent the data in such a way that computers can understand. Thus, Semantic Web ontologies offer an efficient resource of structured data that can be exploited by the Social Web. Together, Social Web and Semantic Web can produce a harmonised duet.

This paper suggests the usage of the Semantic Web ontologies, WordNet and EuroWordNet, for better information retrieval in folksonomies. Section 2 is devoted for the challenges of folksonomies. We demonstrate our approach in Section 3, followed by a discussion in Section 4. In Section 5, we review some related work, and conclude in Section 6.

## 2. CHALLENGES OF FOLKSONOMIES

By analysing the current collaborative tagging systems, we can notice that the main problems are ambiguity, inconsistency, and redundancy problems [Li and Lu, 2008, Marchetti et al., 2007, Mathes, 2004, Angelidou et al., 2008]. This is normal since the collaborative tagging systems (by their nature) are shared by

many users. These users came from different backgrounds, cultures, countries, domains, and tongues. The diversity of the users' behaviours would inevitably create inconsistent tags that would give ambiguous identification of the tagged objects.

The ambiguity and inconsistency of the tags in folksonomies emerge mainly because of linguistic reasons such as; word synonyms [Dix et al., 2006, Golder and Huberman, 2006, Lee and Yong, 2008, Li and Lu, 2008, Marchetti et al., 2007, Mathes, 2004], polysemy (homonym) [Dix et al., 2006, Golder and Huberman, 2006, Lee and Yong, 2008, Li and Lu, 2008, Marchetti et al., 2007], different lexical forms [Dix et al., 2006, Golder and Huberman, 2006, Lee and Yong, 2008, Marchetti et al., 2007], alternative spellings [Marchetti et al., 2007], misspelling errors [Li and Lu, 2008, Marchetti et al., 2007], and use of different languages [Al-Khalifa and Davis, 2006, Angeletou et al., 2008, Zamora and Nistal, 2009]. When searching the folksonomy, these problems cause irrelevant result to be retrieved, and relevant results not to be retrieved. Our concern in this paper is the latter case.

### 3. OUR APPROACH

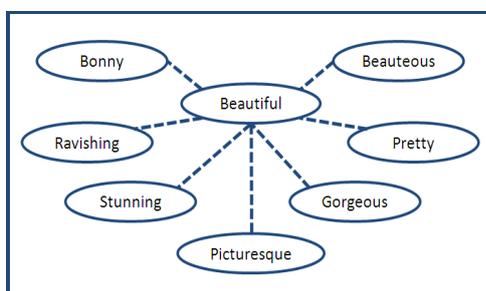
As aforementioned, we focus in our approach on synonyms, multilinguality, and initiating relations among tags in folksonomy based on the semantic relations existing in the ontology. Since all these challenges are lexical ones, we suggest the use of the lexical ontology WordNet. WordNet is a lexical ontology which has set of synonym words, called *synset*, that defines a particular concept. It includes a lot of lexical and semantic relations between words and synsets. It is restricted to no specific domain and covers all common parts of speech; nouns, adjectives, verbs and adverbs [Morato et al., 2004].

#### 3.1 Synonyms

Usually, when a user is tagging, (s)he is not aware of all synonyms for the tags (s)he uses. If the tagger is British, (s)he will use the word “*lift*” whilst the American one will use the word “*elevator*” to describe the lifting device used to move people from one floor to another in a building. Also, when we want to express the beauty of something, we will use words (synonyms) like “*beautiful*”, “*pretty*”, and maybe “*gorgeous*”. Always we miss some of the synonyms. In the first example, if the tag that was used is “*lift*”, the future search will retrieve nothing if we use the word “*elevator*” as a search keyword.

Our idea is to add “*system tags*” every time the user adds tags. The system tags will be added automatically by the collaborative tagging system by consulting the WordNet ontology, these tags are all the existing synonyms in WordNet for the “*user tags*”. Figure 1 shows subset of the synonyms set that can be added by WordNet ontology for the tag “*beautiful*”.

Figure 1. Some synonyms for the word “*Beautiful*” obtained from WordNet ontology



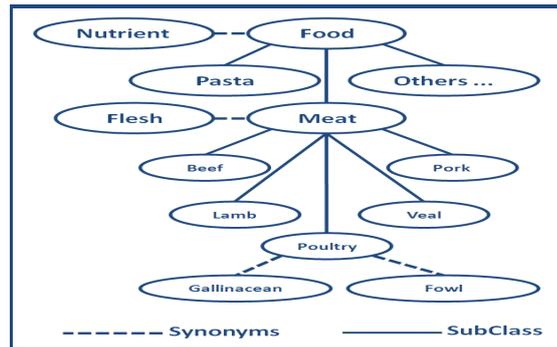
When the user adds the tag “*beautiful*”, the system will add all related synonyms from the WordNet. Future search using any of the synonyms added by the system (system tags) will be able to retrieve the tagged object. Thus, it ensures the retrieval of relevant results.

### 3.2 Tag relations

Imagine if a user tagged a resource as “poultry”. Poultry is indeed kind of meat and it is expected to be retrieved when searching using the keyword “meat” because it is relevant to the search keyword. Unfortunately, it will not be in the result set since this word is not in the tags set for that resource. The same problem is faced again; relevant results are not being retrieved due to lack of semantics in the folksonomy.

The WordNet ontology has such semantic relations among words. Figure 2 shows a part of the WordNet ontology.

Figure 2. Part of WordNet for the word “Poultry”



The system will add the synonyms of the word “poultry” (*gallinacean, fowl*). Also it will add the parent of that word (*meat*) and its synonym (*flesh*) as system tags. To keep a certain level of relevance, only the first parent will be considered. Therefore, anyone who searches using the keyword “meat” will retrieve the resource originally tagged with “poultry”.

### 3.3 Multilinguality

So far, the tagged resource in the abovementioned examples is accessible and visible only if the search keywords are English words. If a non-English speaker is searching using non-English keywords, nothing will be retrieved. If an Italian is searching using the word “bello” (it means: beautiful), the tagged resource in the previous example will seem as irrelevant and thus will not be retrieved. As humans, we can see clearly that it is relevant, but the machines do not.

As a solution for the multilinguality problem, we will use the EuroWordNet. EuroWordNet relates and unites WordNets in different European languages (Dutch, Spanish, Italian, German, French, Czeck, and Estonian) in a single multilingual lexical resource, and it links them to the English WordNet [Vossen, 2002].

We propose that EuroWordNet will find the equivalent words for the tag “beautiful” in the abovementioned languages using so-called Inter-Lingual-Index (ILI). These equivalent words (In addition to their synonyms and parent words as aforementioned) will be added as system tags. This guarantees that future searches by non-English speakers using their own languages will retrieve the relevant resource even if these resources were tagged originally by only English tags, and vice versa.

## 4. DISCUSSION

The proposed approach requires replicating the WordNet and EuroWordNet words and storing them in the folksonomy as system tags. This redundancy of data is justified in the following paragraphs.

Alternatively, we can avoid adding system tags at *tagging time* by consulting and deducing the relations from the lexical resources at *search time*. In the case of synonyms in the previous example, when the user uses the keyword *pretty* in the search, the system will send it to the WordNet. The WordNet will send all the

found synonyms to the folksonomy, and thus all objects that are tagged by any of these synonyms will be retrieved (See Figure 1).

This communication between the folksonomy and the ontology and the searching process inside the ontology itself are time consuming while the user is waiting for a response. We have the choice either to save time or to save space. Time is the critical factor in such a case.

## 5. RELATED WORK

Many researchers have tried to address the abovementioned challenges of folksonomies using different approaches. One of these approaches was to use the power of the Semantic Web in decreasing the ambiguity and inconsistency of tags. Having a glance at these attempts, we can see that there are still many gaps to fill.

In [Al-Khalifa and Davis, 2006], tags are filtered and normalised, then these tags will be adhered to different domain ontologies' concepts, and only the terms that appear in the ontologies will be selected. In this method they remove some users' tags which reflect part of the users' understanding of the tagged object. Moreover, the changes in the users' vocabulary will not be reflected in the semantic ontologies.

In [Ghali and Cristea, 2008], they correct the misspelled tags and group similar tags together, and then the tags are mapped to online ontologies. This method then replaces some tags with corresponding concepts in the online ontologies. We argue that the interference in users' tags will conflict with the ethos of folksonomies (free-keywords).

In [Lee and Yong, 2008], they developed their own folksonomy system using a domain-specific ontology and the WordNet ontology. They detect the domain of the most popular tags, and then they manually build an ontology for that domain. The problem in this method is the necessity of building the domains ontologies, even worse; the domain ontology needs to be built manually.

In [Laniado et al., 2007], they used the WordNet concepts' relations to show the user an additional panel on the browser's interface. This extra visualisation displays related tags organised according to a semantic criterion to facilitate navigation and searching in the folksonomy. It is only visualisation nothing more and some tags are not recognised in the lexicon.

In [Bindelli et al., 2008], they map the unstructured tags to more structured domain ontologies. These ontologies are used for refining the queries to combine results of different tag-based systems. This method uses an ontology-based navigation interface allowing the user to retrieve more related results through graphical navigation of the ontology concepts. This method cannot deal with unmatched tags; which are the tags that do not exist in the domain ontologies.

In [Marchetti et al., 2007], they use WordNet and Wikipedia to substitute semantic assertions for the current tags. These assertions are not simple strings to describe a particular resource; each semantic assertion describes a specific property of a resource. Therefore, the possibility of tagging using free words is absent which contradicts the ethos of folksonomy.

In [Sluijs and Houben, 2008], they apply both syntactic and semantic techniques for connecting tag to ontologies in order to get more semantics about the tag and provide tag suggestions for the users. This method, in addition to offering suggestions to the users, asks the users to give feedback about these suggestions. Hence, we argue that it puts more effort on the users' side to improve the quality of the tags by changing the conventional way by which the users used to interact with the folksonomy.

## 6. CONCLUSION

Folksonomies lack semantics among users' tags which causes relevant results not to be retrieved. Semantic Web ontologies are considered a rich source for semantic relations that, if exploited properly, will improve the searching process in folksonomies. Our approach focused on addressing the problems of synonyms, semantic relations among tags, and multilinguality. It is based on the idea of adding *system tags* as complements to the *user tags* for a wider coverage of potential future search keywords, therefore, more relevant results will be retrieved.

## 7. FUTURE WORK

In the future, this proposal will be implemented therefore more empirical results will follow. EuroWordNet is limited to only some European languages. Our approach is extendable to other languages by using intermediate online dictionaries. These dictionaries might be used to translate from one WordNet to another for languages that are not included in EuroWordNet (e.g. from English WordNet to Arabic WordNet).

A unifying architecture for collaborative tagging systems is under construction. This architecture includes clustering techniques to address the problem of shorthands usage in tagging. Such tags are written using special words that do not belong to any language. Therefore, we can consult the social networks to predict their meanings.

Our proposal needs a software agent that is responsible of reflecting any prospective future changes in the online lexical resources on the folksonomy to keep the system tags in the folksonomy chimed with the linguistic changes that might emerge in the ontology.

The scope of our approach, as mentioned before, is limited to address the problems of synonyms, tag relations, and multilinguality. Therefore, the disambiguation process of some words that might appear in more than one synset (polysemy) is not addressed in this work (i.e. turkey).

## REFERENCES

- Al-Khalifa H. and Davis H., 2006. FolksAnnotation: A semantic meta-data tool for annotating learning resources using folksonomies and domain ontologies. *Innovations in Information Technology*.
- Angeletou S. et al, 2008. Semantically enriching folksonomies with FLOR. *European Semantic Web Conference Workshop: CISWeb*.
- Bindelli S. et al, 2008. Improving search and navigation by combining ontologies and social tags. *1st International Workshop on Ambient Data Integration*.
- Dix A. et al, 2006. Semantic Halo for collaboration tagging systems. *Workshop on the Social Navigation and Community based Adaptation Technologies*.
- Ghali M. et al, 2008. Folksonomies and ontologies in authoring of adaptive hypermedia. *A3H 6th International Workshop on Authoring of Adaptive and Adaptable Hypermedia Workshop*.
- Golder S. and Huberman B., 2006. Usage patterns of collaborative tagging systems. *In Journal of Information Science*, Vol. 32, No. 2, pp 198-208.
- Laniado D. et al, 2007. Using WordNet to turn a folksonomy into a hierarchy of concepts. *Semantic Web Application and Perspectives - Fourth Italian Semantic Web Workshop*.
- Lee S. and Yong H., 2008. Ontosonomy: Ontology-based extension of folksonomy. *Proceedings of the 2008 IEEE International Workshop on Semantic Computing and Applications*, pp 27-32.
- Li Q. and Lu S., 2008. Collaborative tagging applications and approaches. *In Multimedia*, Vol. 15, No. 3, pp 14-21.
- Marchetti A. et al, 2007. Semkey: A semantic collaborative tagging system. *Proceedings WWW 2007 Workshop on Tagging and Metadata for Social Information Organisation*.
- Mathes A, 2004. Folksonomies-cooperative classification and communication through shared metadata. *In Computer Mediated Communication - LIS590CMC*.
- Morato J. et al, 2004. Wordnet applications. *Proceeding of the Second Global Wordnet Conference*.
- Sluijs K. and Houben G., 2008. Relating user tags to ontological information. *Proceedings of 5th International Workshop on Ubiquitous User Modeling*.
- Vossen P, 2002. WordNet, EuroWordNet and global WordNet. *In Revue Franaise de Linguistique Appliquee /RFLA*, Vol. 7, No. 1.
- Zamora F. and Nistal M., 2009. Visualising tags as a network of relatedness. *39th ASEE/IEEE Frontiers in Education Conference*.