

Replacing the Monolithic LOM: A Folksonomic Approach

Hend S. Al-Khalifa & Hugh C. Davis

Learning Societies Lab- ECS, University of Southampton, UK

{hsak04r/hcd}@ecs.soton.ac.uk

Abstract

IEEE-LOM is a well-known metadata standard for describing learning resources. However, many problems are associated with this kind of representation, which include the number of fields to be filled and the amount of time needed to fill them.

To overcome this hurdle, we propose the use of cheap unstructured metadata to create structured semantic metadata, this metadata is called folksonomy. In this paper we show an approach that uses folksonomy tags to create structured metadata using semantic web technologies. The generated folksonomic metadata are then evaluated against a human expert annotation.

1. Introduction

Metadata standards are formal specifications used to semantically annotate educational materials of any kind. They have been developed to support both machine interoperability (information exchange) and resource discovery by human users.

The importance of metadata has also evolved to include the domain of the Semantic Web. At the heart of the Semantic Web is the idea of adding formal metadata that describes the content, context and/or structure of a web resource [7].

There are two widely accepted metadata standards in education [8], namely:

1. DC (Dublin Core) educational version, and
2. IEEE-LOM (Institute of Electrical and Electronic Engineers/Learning Object Metadata).

Most e-learning developers do not adhere strictly to these standards, but prefer to use “application profiles” which more accurately reflect their application’s metadata needs.

Duval et al. [5] have defined application profiles as “... mixing and matching metadata elements, in order

to meet specific requirements for a particular context”. Examples of application profiles include CanCore¹, UK LOM² and ARIADNE³.

To utilize application profiles, their elements need to be populated with appropriate descriptors. This brings us back to the main dilemmas of creating standard metadata, which are: the number of fields to be filled and the amount of time required to fill them.

A possible solution is “*Electronic Forms Must Die*” [4], Duval’s famous slogan to evangelize the automation of metadata creation. Erik Duval, a well-known member in IEEE-LOM standardization board, has realized the need for more automated process to create metadata so that the burden of creation can be alleviated by machines.

Despite Duval’s vision of metadata automation, it is not possible within the existing standards to represent sufficiently fine grained semantic information about learning resources, which would allow the selection of appropriate learning materials from a number of resources within some domain. This drives us to the use of semantic metadata techniques that employ ontologies to generate specific domain semantics.

Therefore, to remove the burden of metadata generation and to generate semantic metadata that handles particular domain semantics, we propose the use of folksonomies.

Folksonomies, as one of Web 2.0 signatures, are considered a free source of unstructured metadata. They can unfold a lot about a web resource subject, its type and possible applications. Social bookmarking services such as del.icio.us⁴ are considered good sources of folksonomies.

The problem of metadata granularity and the need for automating the process of metadata generation are two important issues that led to the idea of using folksonomies in the process of creating semantic

¹ <http://www.cancore.org>

² <http://www.cetis.ac.uk/profiles/uklomcore>

³ <http://www.ariadne-eu.org/>

⁴ <http://del.icio.us>

metadata. This realization can be exploited using the power of semantic metadata representations.

In this paper we show how we successfully used folksonomies to generate semantic metadata for a case study in the domain of teaching Web Design with Cascading Style Sheets (CSS).

2. The domain ontologies and semantic metadata

The three ontologies in our system were chosen based on observed patterns in peoples' tags in the del.icio.us bookmarking service, which are:

- A) **Web Design Domain Ontology**: The domain ontology represents an abstract level of the domain of 'Web Design' and the relation of the concepts in that domain.
- B) **CSS Subject Ontology**: The CSS ontology gives a fine grained listing of the concepts used in the subject of CSS. The concepts were derived from assorted websites that classify the subject of CSS.
- C) **Resource Type Ontology**: The resource type ontology models the different possible pedagogical and technical resource type a given web resource might have e.g. editor, example, etc.

Moreover, the semantic metadata elements used to describe CSS web resources were constructed by mixing elements from the IEEE LOM standard and elements specific to the domain of CSS. Thus, the resultant metadata record consists of 15 elements, which are: *Title, Description, Keywords, Resource Type, Recommendation, Property, Selector, Unit, Attribute, Technique, Application, Subject, Layout, Difficulty level* and *Instructional level*.

Table 1 shows the candidate elements from the IEEE LOM standard that were used in our semantic metadata along with the appropriate RDF namespace bindings.

Table 2 shows the extra elements that were not described by IEEE-LOM but we have created them to add extra useful semantics to a CSS resource.

Table 3 shows two examples of CSS domain specific elements that are used to describe the fine grained semantics of the CSS domain.

Table 1: IEEE-LOM properties used in the Semantic Metadata

LOM Property	RDF Binding	Description
--------------	-------------	-------------

1.2- Title	dc[#]:title with a literal value	Name given to this learning resource.
1.4- Description	dc:description with a literal value	Description of the content of the learning resource.
1.5- Keyword	dc:subject with a literal value	A keyword describing a topic in the learning resource.
5.2- Learning Resource Type	RTO[*]:hasType	Specify the pedagogical and technical type of learning resource. Example: tutorial, code, etc.
5.8- Difficulty	lom⁺:hasDifficulty	How hard it is to work with the learning resource. Example: easy, medium, difficult, very difficult.

Table 2: Extra elements used with the semantic metadata

Property	RDF Binding	Description
Recommendation	SDO[^]:hasRecommendation with a numeric value	Used to describe how popular a resource is for the people who voted for it.
Instructional Level	SDO:hasInstructionalLevel	Describes the instructional level of a web resource. Example: basic, intermediate, advanced

Table 3: Excerpt of the domain specific elements used with the semantic metadata

Property	RDF Binding	Description
Application	SDO:hasApplication	Describes the range of applications a CSS web resource is promoting.
Layout	SDO:hasLayout	Describes the range of layouts a CSS web resource is promoting.

3. FolksAnnotation tool architecture

The FolksAnnotation tool (Figure 1) is a stand-alone application that takes as an input a URL of a bookmarked web resource in del.icio.us, and generates

[#] Dublin Core Namespace

^{*} Resource Type Ontology Namespace

[^] Subject Domain Ontology Namespace

⁺ LOM Educational category namespace

<http://kmr.nada.kth.se/el/ims/schemas/lom-educational>

in the background the appropriate semantic metadata as an RDF format. The tool was built using Java SWT library and uses Jena API for ontology manipulation and inference.

The implemented system consists of two processes: 1) tags extraction and normalization pipeline and 2) semantic metadata creation pipeline.

The process starts by fetching a bookmarked web resource from the del.icio.us bookmarking service, then the tag extraction process begins by extracting viable information from the web page of the bookmarked web resource, this includes: Title, URL, Number of people who bookmarked the resource and the list of all tags assigned to the bookmarked web resource.

The extracted tags are then passed to the normalization process which performs a series of filters for cleaning the tags. The filters are performed sequentially in the following order:

- **Lower-case filter:** Tags are converted to lower case so that string manipulation (e.g. comparison) can be applied to them easily,
- **Non-English filter:** Non-English characters are dropped; this step is to insure that only English tags are present when doing the semantic annotation process,
- **Stemming filter:** Tags are stemmed (e.g. convert plural to singular) using a modified version of the Porter Stemmer⁵,
- **Tags sense Disambiguation filter:** stemmed tags are passed to this module to remove ambiguous tags, i.e. those tags that have multiple meanings in different context (i.e. polysemy)
- **Grouping filter:** similar tags are grouped (e.g. inclusion of substrings),
- Finally, **the removal filter**, where the general concept tags (e.g. programming, web, etc) and ambiguous ontological terms in our domain of interest are eliminated.

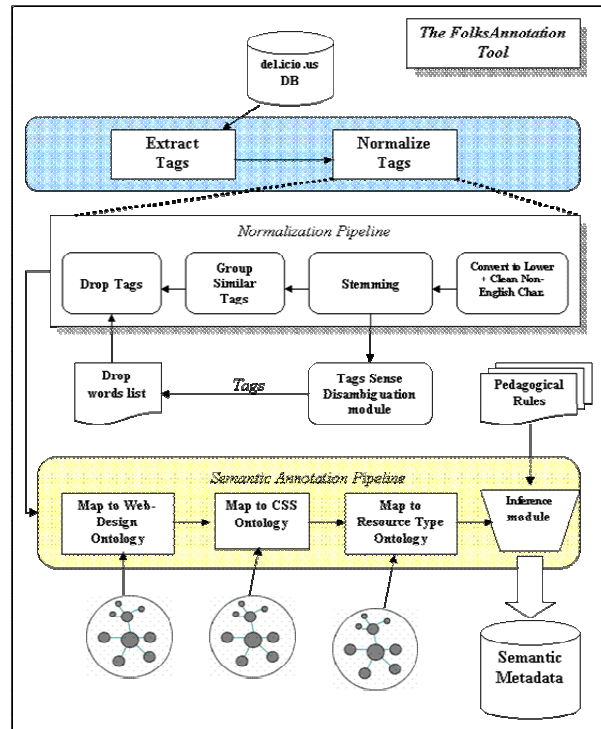


Figure 1. An overview of the Tool illustrating the interplay of the different components.

At the end of the normalization pipeline we get a normalized list of folksonomy tags. The normalized list is then ready to be used in the semantic annotation process, where each normalized folksonomy tag is associated with a corresponding instance in one of our three ontologies. This process will attach ontology instances as descriptions for the web resource.

The inference module is responsible for associating pedagogical semantics (i.e. difficulty level and instructional level) to the annotated web resource. These two values are generated from a set of inference rules feed to the inference engine by a separate file. The pedagogical rules will only function if there is enough information available in the basic semantic descriptors. When finishing the annotation process, each item of the generated semantic metadata is saved in a triple store.

4. Evaluation results

In previous work [1, 2] we have evaluated the representative-ness of our metadata elements and the performance of our generated semantic metadata based on ontology search. In this paper, however, we report our final evaluation procedure where we evaluate the

⁵ <http://www.tartarus.org/~martin/PorterStemmer/>

search-ability and quality of folksonomy metadata assignment against the assignment of a human subject-matter expert.

The subject-expert was asked to annotate a set of ten randomly chosen web resources from our data set using our three ontologies. We also annotated all web resources beforehand to be considered as a *benchmark* standard to which the two parties were compared against. Then we evaluated the expert annotation against the folksonomy annotation using IR measures of recall (1), precision (2) and f-measure (3).

$$Recall = \frac{Retrieved}{Relevant} \quad (1)$$

Relevant: means all resources marked relevant by us as a benchmark standard.

$$Precision = \frac{Relevant}{Retrieved} \quad (2)$$

$$F - Measure = 2 * \frac{Precision * Recall}{Precision + Recall} * 100\% \quad (3)$$

We also compared manually the quality of folksonomy assignments against the human expert assignments.

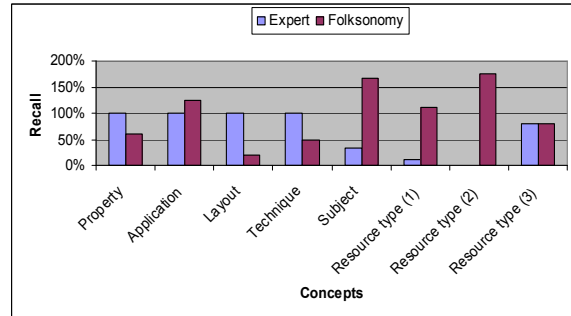
Figure 2 shows the results obtained after evaluating the search performance of the human expert assignment and the folksonomy assignment against the benchmark standard.

The recall results, Figure 2(A), show that the folksonomy results were better than the expert results in almost half of the queries. However, the precision results, Figure 2(B), show that the human expert outperforms the folksonomy results in two cases and equals them in the rest, except for one case (i.e. resource type (2)) where the human expert did not assign a value to the web resource; this affected its precision and recall scores accordingly.

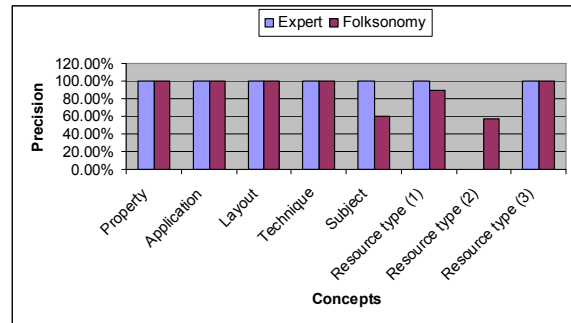
As for the f-measure results, Figure 2(C), we can assume that in our case study, folksonomy search has performed better in most of the cases compared to the expert search results. This outcome can be attributed to the high values of the recall which boosted the f-measure results.

Finally, as for the qualitative manual evaluation of both assignments, we have found that although human expert is more precise than folksonomy users when annotating a web resource, folksonomy tags have added more potential contextual dimensions to most of our sample web resources. Moreover, when tagging the web resources with elements from the resource type ontology, the human expert tends to annotate resources with pedagogical instances and forgotten about the technical aspect of the resource. This observation

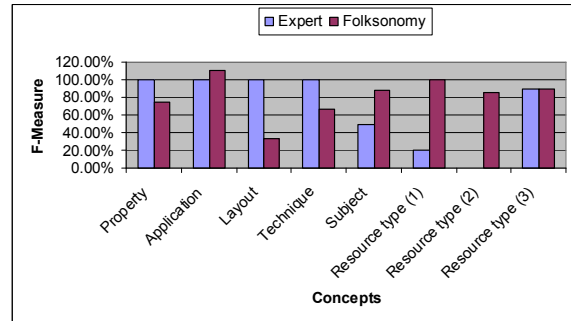
demonstrates the power of aggregating people's intelligence against a human subject-matter expert annotation.



(A)



(B)



(C)

Figure 2: The Recall (A), Precision (B) and F-Measure (C) of folksonomy search results and the human expert search results against the benchmark. The performance of both techniques is shown for eight different queries.

5. Related work

Automatic metadata generation techniques in our field of interest can be categorized into two types: Techniques to generate Semantic Metadata, and Techniques to generate metadata from folksonomies.

Our paper will talk about the two genres and how do they relate to our current work.

5.1 Semantic metadata techniques

The most recent example is the TANGRAM system [6]. TANGRAM is a learning web application for the domain of Intelligent Information Systems (IIS) where users (students and teachers) can upload, describe, search or compose a new learning object using components in the system repository. The system provides a solution for automatic metadata generation of learning objects (LO) components. Thus, each generated semantic metadata attached to a LO allows the TANGRAM system to assemble these objects into new LOs personalized to the users' goals, preferences and learning style.

Despite the similarity between the purpose and outcome of the TANGRAM system and our system, our system does not rely on any algorithms to generate or extract metadata from web resources (that are equivalent to LO). Moreover, our system uses a freely accessible web service (i.e. del.icio.us) for generating semantic metadata, while the TANGRAM system operates from within a learning management system.

5.2 Folksonomic metadata techniques

To the best of our knowledge, utilizing folksonomies in the process of creating semantic metadata for eLearning applications is not that well-researched. However, there is a nascent prototypical tool called 'CommonFolks' [3] that is being developed in the laboratories of Advanced Research in Intelligent educational Systems (ARIES), Canada, to create ontological metadata (i.e. semantic metadata) from people's tags to annotate learning resources to be used in adaptive eLearning systems.

The system goal is to employ collaborative tagging in order to make metadata creation fast, easy and machine consumable using the English language ontology (WordNet). The approach works by appending WordNet with tags in a 'is-a' relationship. The tool is still in its early stages and no evaluation results have been reported yet.

6. Conclusion

This paper demonstrated that folksonomies can be considered a useful source for creating semantic metadata for educational purposes. The generated folksonomic metadata was evaluated against a human expert annotation and the results showed that

folksonomic metadata generated using folksonomy tags were better in terms of search and contextual coverage than the metadata created by the human expert. Despite that folksonomy tags were neither perfect nor complete, yet, they added potential contextual dimension to the generated metadata.

6. References

- [1] Al-Khalifa, H. S. and H. C. Davis (2006). *FolksAnnotation: A Semantic Metadata Tool for Annotating Learning Resources Using Folksonomies and Domain Ontologies*. Proceedings of the Second International IEEE Conference on Innovations in Information Technology, Dubai, UAE, IEEE Computer Society.
- [2] Al-Khalifa, H.S., H.C. Davis, and L. Gilbert. *Creating Structure From Disorder: Using Folksonomies To Create Semantic Metadata*. In Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST). 2007. Barcelona, Spain: Springer.
- [3] Bateman, S., C. Brooks and G. McCalla (2006). *Collaborative Tagging Approaches for Ontological Metadata in Adaptive ELearning Systems*. the Proceedings of the Fourth International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL 2006) in conjunction AH2006, Dublin, Ireland.
- [4] Duval, E. (2004). *We're on the road to*. Proceedings of the ED-MEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications, Lugano, Switzerland, AACE.
- [5] Duval, E., N. Smith and M. V. Coillie (2006). *Application Profiles for Learning*. Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies, ICALT 2006, Kerkrade, The Netherlands, IEEE Computer Society.
- [6] Jovanovic, J., D. Gašević and V. Devedžić (2006). "Ontology-Based Automatic Annotation of Learning Content." *International Journal on Semantic Web and Information Systems (IJSWIS)* 2(2): 91-119.
- [7] Marshall, C. C. and F. M. Shipman (2003). *Which semantic web?* Proceedings of the fourteenth ACM conference on Hypertext and hypermedia, Nottingham, UK, ACM Press.
- [8] Stratakis, M., V. Christophides, K. Keenoy and A. Magkanaraki (2003). *E-Learning Standards*: p. 45.