

Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data

Alexandre Passant
Electricité de France R&D
1, Avenue du Général de Gaulle
92141 Clamart, France
alexandre.passant@edf.fr
&
LaLIC, Université Paris-Sorbonne
28, rue Serpente
75006 Paris, France
alexandre.passant@paris4.sorbonne.fr

Philippe Laublet
LaLIC, Université Paris-Sorbonne
28, rue Serpente
75006 Paris, France
philippe.laublet@paris4.sorbonne.fr

ABSTRACT

This paper introduces MOAT, a lightweight Semantic Web framework that provides a collaborative way to let Web 2.0 content producers give meanings to their tags in a machine-readable way. To achieve this goal, this approach relies on Linked Data principles, using URIs from existing resources to define these meanings. That way, users can create inter-linked RDF data and let their content enter the Semantic Web, while solving some limits of free-tagging at the same time.

Keywords

Semantic Web, Web 2.0, Tagging, MOAT, Linked Data, Architecture of Participation, SIOC

1. INTRODUCTION

Among the various tools and principles that Web 2.0 introduced such as blogging, collaborative knowledge management with wikis or on-line social networking, tagging is one of the most interesting phenomena. While one of the main ideas of Web 2.0 is to let users play an important role in the process of creating content, tagging goes a step further by letting them control the way they organise it. By adding simple keywords, or tags, to their data or to data they browse on-line, they can decide themselves which meta-data must be related to any content. These tags can refer to various and really different levels of annotation, as Golder and Huberman [4] identified, from content meta-data (topic(s) of a blog post) to quality meta-data (opinion about a webpage) or even self reference. Yet, whatever the way people use it, tagging raises various issues regarding information retrieval. The Semantic Web and especially semantic annotation [7] offers better perspectives regarding information retrieval, by

using URIs that uniquely identify resources used to annotate data. Nevertheless, this process is generally a task harder to overcome than free-tagging.

This paper introduces MOAT - <http://moat-project.org> -, a framework based on Semantic Web principles designed to bridge this gap between free-tagging and semantic annotation. Its goal is to provide a simple and collaborative way to annotate content thanks to existing URIs with as little effort as possible and by keeping free-tagging habits. We will first briefly introduce the well-known limits of free-tagging and why, while a human can solve them, computers are not able to do it easily. We will then introduce how we represent the meaning of a tag and the way we model it in a machine-understandable way. We will then present the MOAT project, and start with related work about tagging vocabularies on the Semantic Web. We will describe the MOAT ontology, used to model these relationships between tags and their meanings, using URIs of existing Semantic Web resources. Next, we will see how our framework helps users to take advantage of these ideas thanks to simple tools and collaborative principles that ease the task of semantic annotation. Finally, we will overview how this approach is related to the Linked Data movement.

2. TAGS AND THEIR MEANINGS

2.1 Limits of Tagging for Software Agents

While tagging became popular thanks to Web 2.0 services such as del.icio.us and Flickr, or blogging and dedicated agregation websites like Technorati, it raises various issues from an information retrieval point of view. These limits mainly consists in the ambiguity and heterogeneity of tags, as well as the flat organisation of folksonomies. Ambiguity and heterogeneity may produce too much noise or silence while the lack of relationship between tags makes difficult to find related content from a given entry point. Thus, Mathes argues that *"a folksonomy represents simultaneously some of the best and worst in the organization of information"* [11].

These limits cannot be easily overcome since tags, from a machine point of view, do not carry any semantics about what they represent, while a human can interpret such se-

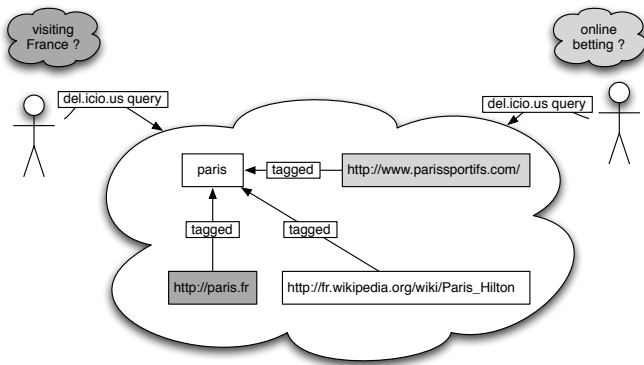


Figure 1: User perception and tag search

semantic when tagging or reading some content. For example, when tagging a blog post "paris", a user has in mind an existing concept from the real-world, which can be a city, a person, or anything else. Yet, from a computer point of view, there is no way to make a difference since the only thing it will consider is a text string. Thus, when retrieving data, the user himself will have to manually deal with tags ambiguity to find, from a resulting dataset, in which post this tag has been used in a similar way that he had in mind (see Fig.1).

2.2 Defining and Representing the Meaning of a Tag Thanks to the Semantic Web

In order to represent the meaning of a tag, we first consider the meaning it can be assigned to in a particular tagging context (e.g. in that post context, "paris" means a french city). Thus, we extend the tripartite model of tagging and folksonomies [12], by adding a *local meaning* for each tagging action:

$$\text{Tagging}(\text{User}, \text{Resource}, \text{Tag}, \text{Meaning}) \quad (1)$$

From this definition, we define the *global meanings* of a tag, i.e the list of all different meanings a tag can be assigned to. To keep a social aspect within this definition, each meaning is related to the set of users that used it:

$$\text{Meanings}(\text{Tag}) = \{(\text{Meaning}, \{\text{User}\})\} \quad (2)$$

In order to represent these meanings in a machine-readable way, which can help to solve some of the issues raised before, we think that Semantic Web, and especially URIs of existing resources can play an important role. Since they provide unique identifiers for resources of the real-world, we believe that they are one of the most efficient way to define it, either globally or in a tagging context. Thus, in the two previous definitions, meanings are defined thanks to URIs of existing resources, which can be part of any knowledge base, as GeoNames¹ or DBpedia [1], but also internal corporate datasets. For example, the meaning of the tag "paris" can be, in a given blog post context, the URI `<http://dbpedia.org/resource/Paris>` while it can

have a completely different meaning on another one, e.g. `<http://dbpedia.org/resource/Paris_Hilton>`.

3. THE MOAT ONTOLOGY

3.1 Tagging and the Semantic Web

Various work has been done regarding tagging and the Semantic Web, in which we can distinguish work related to modeling tags thanks to Semantic Web technologies and work related to mining folksonomies from ontologies [6] [15] or linking ontologies and folksonomies [14]. Here, we will focus on the first aspect. The Tag Ontology [13] provides a model that introduces **Tag** and **Tagging** classes in order to represent tags and tagging actions. Its **Tag** class inherits from `skos:Concept` and the model relies on FOAF[3] for modeling the user aspect which makes the ontology compliant with existing standards. Gruber [5] defined a similar model but extends this concept by notifying the source (i.e. the webspace) where the action takes place whereas it does not consider its temporal aspect contrary to the previous approach. Yet, his ideas have not been implemented. The SCOT ontology [10] focuses on a way to share tags by modeling tagclouds and also provides various properties to link tags together (e.g.: synonymy, case-variation ...). Finally, while they do not define any semantic modeling, Flickr machine tags² allow people to embed RDF-like assertions within their tags.

3.2 Classes and Properties of the Ontology

The MOAT ontology features a **Tag** class that extends the one defined in the Tag Ontology. Indeed, in our case, a **Tag** instance must have a single label (thus the ontology uses an OWL restriction) and its URI must respect a certain pattern that is defined by a MOAT server, as we will see later. That way, it offers common URIs for tags that can be shared across communities and social media sites.

In order to represent tag meanings, the ontology can be divided in two parts. The first one, dedicated to *global meanings* (2), introduces a `moat:hasMeaning` relationship and a `moat:Meaning` class, that are used to link a **Tag** instance to all its meanings. Each `moat:Meaning` instance features a unique property called `moat:meaningURI` in order to link to the URI of an existing Semantic Web resource that represents the given meaning. Moreover it features at least one `foaf:maker` link (using once again a cardinality constraint in the ontology) in order to keep a trace of the user(s) that defined that URI as a meaning for the tag. The following snippet of code represents the global meanings assigned to the tag "paris". Here, three different meanings have been represented, while one of them is shared by two users. While the following example shows how MOAT can be used to represent tag ambiguity, the ontology can also be used to deal with heterogeneity since two different tags can share common meanings (using `meaningURI`), which can be helpful in multi-lingual systems where different tags refer to the same resource (e.g. *paris* and *parigi*).

¹<http://geonames.org>

²<http://tech.groups.yahoo.com/group/yws-flickr/message/2736>

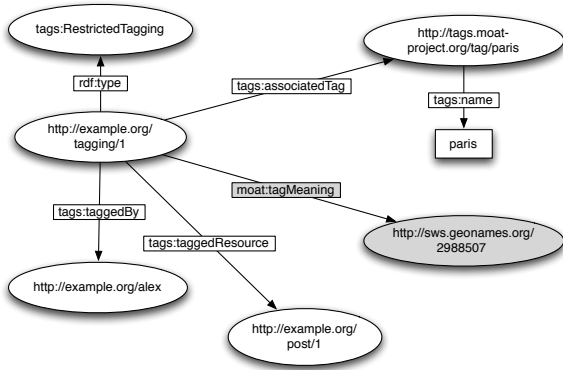


Figure 2: Tagging and the local meaning of a tag

```
@prefix moat: <http://moat-project.org/ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
```

```
<http://tags.moat-project.org/tag/paris> a moat:Tag ;
moat:name "paris" ;
moat:hasMeaning [
  a moat:Meaning ;
  moat:meaningURI <http://sws.geonames.org/2988507/> ;
  foaf:maker <http://example.org/alex/>
] ;
moat:hasMeaning [
  a moat:Meaning ;
  moat:meaningURI <http://sws.geonames.org/4402452/> ;
  foaf:maker <http://example.org/alex/> ;
  foaf:maker <http://myblog.net/user.rdf#me>
] ;
moat:hasMeaning [
  a moat:Meaning ;
  moat:meaningURI dbpedia:Paris_Hilton ;
  foaf:maker <http://myblog.net/user.rdf#me>
] .
```

The second part of the ontology defines a model for the *local meaning* (1) of a tag. Here, we rely on the `RestrictedTagging` class from the Tag Ontology, which identifies a tagging relationship between a post, a user, and a single and only tag. Thus, we introduced a `moat:tagMeaning` property to link a `RestrictedTagging` instance to the meaningful URI in this context, as show on Fig.2.

4. FRAMEWORK ARCHITECTURE

4.1 Global Architecture

While the MOAT ontology provides a model to define relationships between tags and their meanings, this is not enough to let users easily define these meanings, either globally or locally. In order to achieve this goal, we designed a client-architecture that (1) lets users define which URIs they want to assign as meanings to their tags and (2) let them choose a URI from an existing set in a given tagging context. Furthermore, this process relies on an architecture of participation, since meanings are shared across a community and can evolve among time thanks to users themselves, that can add new meaning URIs.

Thus, the framework consists of the association of (1) a MOAT server that can deliver the list of all global meanings for a given tag in a community and that can be updated by users themselves and (2) clients that provide interfaces to define and choose the meaningful URI for a given tag in a tagging action and then produce RDF data describing it. Both interact with each other using HTTP, and exchange data thanks to the previously defined ontology. To benefit from this architecture, users simply install a client on their favourite blogging tool while subscribing to a tag server as they could have done with Annotea [9]. Then, users create their content and tag it as usual. As soon as the content is saved, the client automatically queries the server to get the list of all of the URIs associated to the given tag(s). The user then choose which one he wants to assign to each tag in this context - or define new ones if nothing relevant is found - and then saves its content, which is instantaneously exported as interlinked RDF data (see Fig.3).

4.2 Interaction Between Parties

MOAT clients can interact with a server both in reading and writing data, respectively to retrieve meanings for a tag or to add new ones. Both actions are performed over HTTP in a RESTful way with normalized API calls³.

In order to retrieve the set of meanings for a tag, clients can use various URLs to query the server. First, since the server uses content-negotiation principles, clients can simply request the tag URI to get its related RDF description, as soon as they send the correct `accept` header to the server. Moreover, it means that the tag itself carry some semantics, since its deferencable URI gives information about all the meanings it can be related to. As explained before, the tag URI must respect a given pattern so that the server understand it, which is: `tag_uri = SERVER_BASE+urlencode(tag_label)`. Yet, in order to ease the task of writing MOAT clients to developers, clients can directly request the RDF file URL, or even a JSON description of the content, available at `tag_uri/json`, so that they can write clients without to deal with RDF. The description sent to the client is created upon request thanks to a SPARQL query sent to the triple-store which is used as a back-end of the server. That way, a MOAT server acts as a layer between a triple-store and various clients.

To update the server data, i.e. add new meanings - since a MOAT server is initially empty when a community install it - , clients simply send the URL of a file containing `Tag` instances and related `Meanings` (that is automatically created by the client itself thanks to user actions). Depending on the choice of the community, this action may require an additional API key to allow this operation to be performed. When getting the file URL, the MOAT server imports it in the back-end store, thus merging the new meanings with already existing ones for the given tag and adding new `foaf:maker` statements, so that decentralized `Meaning` instances related to a particular tag are then combined. Hence, we immediately benefit from the collaborative aspect of this process, since as soon as one user defines a new meaning for a tag, the whole community can reuse it.

³<http://moat-project.org/server>

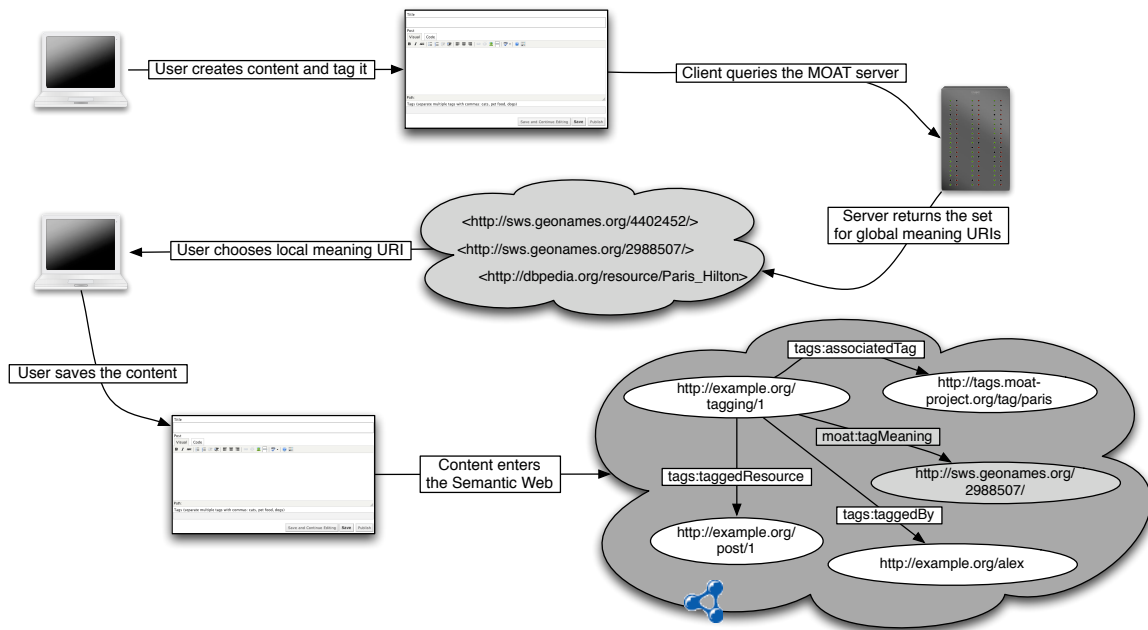


Figure 3: Global architecture

4.3 Implementations

The MOAT server is currently available as a PHP5 application⁴ that must be plugged on the top of a triple-store. It uses a **Connector** class to provide an interface between the server and a RDF storage system and currently features connectors for ARC2⁵ and for the 3store API⁶, using SPARUL LOAD queries to add new data.

A first client implementation was developed as a Drupal module⁷. It provides an interface with any MOAT server, export of the tagging object as a RDF file, and uses SIOC [2] to describe other meta-data about the tagged item. In order to help users find new URIs for their tags, the module uses the Sindice [16] Widget⁸, as shown on Fig. 4. Moreover, already used URIs are displayed as links so that user can browse them. A client implementation was recently added in Openlink Virtuoso⁹ and another one may be available soon for Wordpress through the SparqlPress¹⁰ add-on.

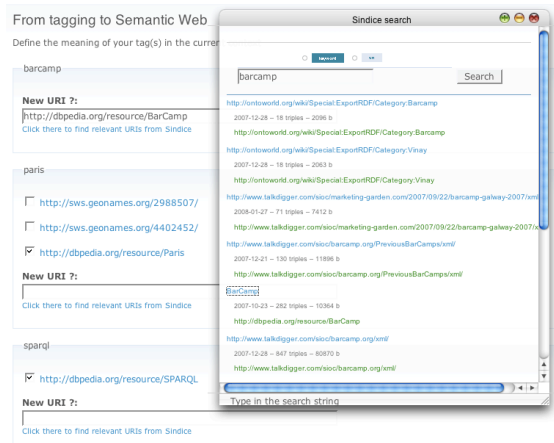


Figure 4: Drupal module interface

5. MOAT AND THE LINKED DATA WEB

While MOAT can be mainly seen as a way to solve some issues of free tagging by giving meaning to tags in RDF, it must also be considered from a Linked Data point of view. By providing a way to link any Web 2.0 content to existing URIs, it may help to discover content related to these URIs thanks to lookup services such as Sindice. Furthermore, it can also be helpful to provide new tag-based search engines using Semantic Web principles that could answer advanced

queries like *"Find all blog posts tagged with french cities"*. Finally, it can also be used to suggest related content by looking at all resources linked to a meaning URI and find posts linked to one of this resource (Fig. 5). This example shows how, from a simple free-tagging scheme, posts could be finally related thanks to the way they are linked to URIs and existing links between these URIs. Moreover, in order to provide a direct link between the tagged content and a meaning URI, one can directly rely on the `sioct:topic` property, thus letting SIOC further enter the Linked Data Web.

6. CONCLUSIONS

In this paper, we introduced MOAT, an ontology and a collaborative framework which goal is to let users bridge the gap between free-tagging and semantically-annotated content in a simple way. This framework relies on an architec-

⁴<http://moat-project.org/server>

⁵<http://arc.semsol.org>

⁶<http://threestore.sf.net>

⁷<http://moat-project.org/clients>

⁸<http://sindice.com/dev/widget>

⁹<http://www.openlinksw.com/virtuoso/>

¹⁰<http://wiki.foaf-project.org/SparqlPress>

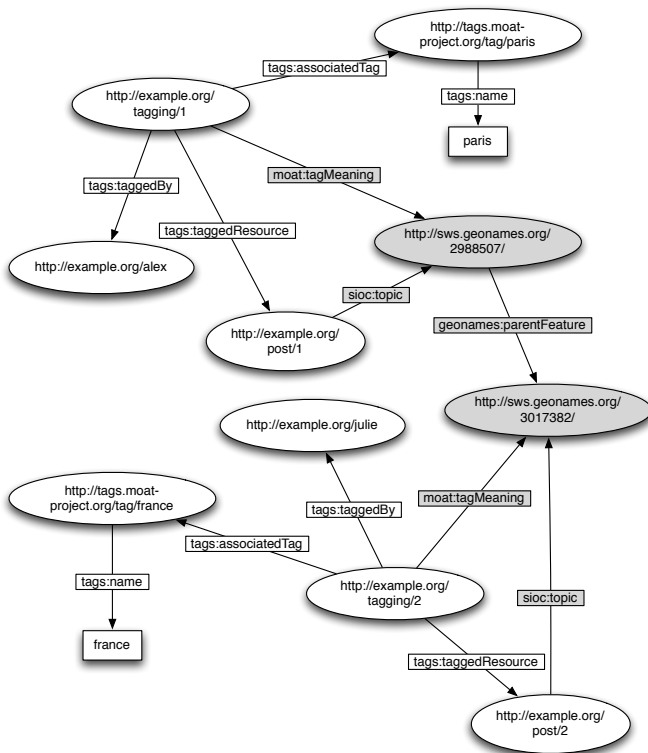


Figure 5: Tagging and the Linked Data Web

ture of participation and allows people to interlink content with any URI from existing resources, thus letting tagging and related services as blogs enter the Linked Data Web. Future works regarding MOAT will mainly focus on implementations for other platforms, as well as integrating social networking aspects when retrieving a list of URIs from a server for a given tag, so that a user could be suggested in priority to use meanings that have been assigned by one of his friend. Moreover, while MOAT does not currently focuses on semi-automatic meaning definition for tags [17] or clustering [8], we think it could provide a nice machine-readable representation model for such work, and that it could also be combined with such approaches in the future.

7. REFERENCES

- [1] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. *6th International Semantic Web Conference*, November 2007.
- [2] J. Breslin, A. Harth, U. Bojars, and S. Decker. Towards Semantically-Interlinked Online Communities. *2nd European Semantic Web Conference*, May 2005.
- [3] D. Brickley and L. Miller. FOAF Vocabulary Specification 0.91, November 2007. <http://xmlns.com/foaf/0.1/>.
- [4] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, April 2006.
- [5] T. Gruber. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on*

Semantic Web and Information Systems, 3(2), 2007.

- [6] H. Halpin, V. Robu, and H. Shepard. The dynamics and semantics of collaborative tagging. In *Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW06)*, November 2006.
- [7] S. Handschuh and S. Staab. *Annotation for the Semantic Web*. Number 96 in *Frontiers in Artificial Intelligence and Applications*. IOS Press, Amsterdam, 2003.
- [8] C. Hayes and P. Avesani. Using tags and clustering to identify topic-relevant blogs. In *Proceedings of the 1st International Conference on Weblogs and Social Media (ICWSM 07)*, Boulder, Colorado, March 2007.
- [9] J. Kahan and M.-R. Koivunen. Annotea: an open RDF infrastructure for shared Web annotations. In *Proceedings of the tenth international conference on World Wide Web*, pages 623–632, New York, 2001. ACM Press.
- [10] H. Kim, J. Breslin, S. Yang, and H. Kim. int.ere.st: Building a Tag Sharing Service with the SCOT Ontology. In *Proceedings of the AAAI 2008 Spring Symposium on Social Information Processing*, Stanford University, California, 2008.
- [11] A. Mathes. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>, December 2004.
- [12] P. Mika. Ontologies are us: A unified model of social networks and semantics. In *Proceedings of the 4th International Semantic Web Conference, ISWC 2005*, volume 3729 of *Lecture Notes in Computer Science*, pages 522–536, Galway, Ireland, November 2005.
- [13] R. Newman. Tag ontology design. <http://www.holygoat.co.uk/projects/tags/>, March 2005.
- [14] A. Passant. Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *Proceedings of International Conference on Weblogs and Social Media*, Boulder, Colorado, March 2007.
- [15] L. Specia and E. Motta. Integrating folksonomies with the semantic web. In *Proceedings of the European Semantic Web Conference (ESWC2007)*, LNCS, pages 624–639, July 2007.
- [16] G. Tummarello, R. Delbru, and E. Oren. Sindice.com: Weaving the open linked data. In *6th International Semantic Web Conference*, pages 552–565, 2007.
- [17] C. M. A. Yeung, N. Gibbins, and N. Shadbolt. Understanding the semantics of ambiguous tags in folksonomies. In *Proceedings of the International Workshop on Emergent Semantics and Ontology Evolution (ESOE2007) at ISWC/ASWC2007*, November 2007.