Enabling Semantics-Aware Collaborative Tagging and Social Search in an Open Interoperable Tagosphere

Javier Soriano
School of Computing
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Madrid (Spain)
(+34) 913367396
jsoriano@fi.upm.es

Javier López
School of Computing
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Madrid (Spain)
(+34) 913367394
jlopez@fi.upm.es

Miguel Jiménez
School of Computing
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Madrid (Spain)
(+34) 913367394
mjimenez@fi.upm.es

Fernando Alonso
School of Computing
Universidad Politécnica de Madrid
Campus de Montegancedo s/n
28660 Madrid (Spain)
(+34) 913367430
falonso@fi.upm.es

ABSTRACT

To make the most of a global network effect and to search and filter the Long Tail, a collaborative tagging approach to social search should be based on the global activity of tagging, rating and filtering. We take a further step towards this objective by proposing a shared conceptualization of both the activity of tagging and the organization of the *tagosphere* in which tagging takes place. We also put forward the necessary data standards to interoperate at both data format and semantic levels. We highlight how this conceptualization makes provision for attaching identity and meaning to tags and tag categorization through a Wikipedia-based collaborative framework. Used together, these concepts are a useful and agile means of unambiguously defining terms used during tagging, and of clarifying any vague search terms. This improves search results in terms of recall and precision, and represents an innovative means of semantics-aware collaborative filtering and content ranking.

Categories and Subject Descriptors

H.3.5. **[Online Information Services]**: Web-based services

H.3.3. **[Information Search and Retrival]**: Retrieval models, Search process, Information filtering, Relevance feedback

H.5.3. **[Group and Organization Interfaces]**: Web-based interaction

General Terms

Design, Standardization, Languages, Experimentation

Keywords

Web 2.0, Social Tagging, Social Bookmarking, Social Search, Content Ranking, Collaborative Filtering.

1. INTRODUCTION

Despite all the attention it is garnering recently, social search is not really new. It has been around in one form or another from the early days of the Web, even before the first search engines emerged in the early 90s and whenever human judgments about Web contents quality, relevance and interest have been taken into account to improve the results of searching. The really innovative thing about recent approaches to social search is their alignment with the Web 2.0 vision that exploits the collective intelligence of Web community collaboration when working social media Web sites.

In a collaborative tagging- and rating-based mash-up approach to social search, tagging provides Web content with user-contributed metadata that helps to distinguish high-quality contents from all the noise and to counter spam-induced noise in current search engines. Additionally, it gives text-based search engines a fighting chance in media sharing. Meanwhile, rating helps to improve search results by voting the contents tagged by others and obtained by searching.

To make the most of a global network effect and to search and filter the long tail, a collaborative tagging- and rating-based mash-up approach to social search should be based on the global activity of the entire tagging community across the whole range of existing and future social media applications and aggregators, such as Del.icio.us, Yahoo MyWeb 2.0, Flickr or Technorati.

However, the existing *tagosphere* is made up of an ever-increasing number of separate, disconnected systems and aggregators (i.e. each Web site acts as a separate tagosphere). Therefore, they are missing out on an opportunity by not making the most of the millions of active participators that could provide valuable knowledge work for developing social search engines that tap the power of such collective intelligence.

There have been only a few noteworthy attempts at interconnecting these social media systems and at aggregating and building on their data to enhance the user search experience (e.g. Whonu or TagBulb). These initiatives have neither an explicit nor a shared conceptualization that would allow seamless interoperability. Instead, they are all based exclusively on the use of part-fledged REST (REpresentational State Transfer) [2] APIs and/or applicable data standards, such as the well-known xFolk MicroFormat [1], or even on rough scraping from different sources, i.e. they are hardly representative of a social search-enabling infrastructure for interoperating at data format and/or API levels, let alone at the semantic level. For both different social media systems to interoperate and a social search engine to be logically consistent when combining and building on data from different sources, they all need to make ontological commitments to the semantics of tagging, aside from any agreements on formats, APIs and protocols.

iiWAS2008, November 24–28, 2008, Linz, Austria.

Additionally, they all suffer from the inherent ambiguity and imprecision of language during the tagging process. Both problems have a significant negative impact on both the precision and the recall of the search results.

In this paper, we tackle these shortcomings jointly and take a further step towards enabling a global activity of tagging, rating and filtering by proposing a shared conceptualization of both the activity of tagging/rating/filtering and the organization of the "tagosphere" in which these activities take place (section 2). We then highlight how this conceptualization makes provision for attaching identity and meaning to tags through a Wikipedia-based collaborative framework (section 3), thus making search results more precise. In section 4 we explain how our conceptualization accounts for the concept of "metatagging" or tag categorization and its benefits. We then put forward the necessary data standards to interoperate at both data formats and semantic levels (section 5), and place these data standards in the context of an architectural stack for interoperability (section 6). Next we take into account related work in section 7. Finally, we conclude the paper and present work in progress in section 8.

2. TOWARDS A SHARED CONCEPTUALIZATION OF THE TAGOSPHERE

Considering what happened eight years ago, when Brad L. Graham successfully coined the term *blogosphere*, which in turn resembles *logosphere* and *noosphere*, it is tempting to again turn to the geologist Eduard Suess, who first coined the term *biosphere* in 1875, and following recent trends, suggest the term *Tagosphere* to define the place on the Web's surface where collaborative tagging systems dwell. We envision a tagosphere as being a collective term covering all tagging systems and repositories, as well as the whole tagging community, as a kind of densely interconnected social network.

As opposed to an ever-increasing number of separate, disconnected tagging systems and aggregators (as is the case today), we suggest that the tagosphere should be thought of as an open, interoperable ecosystem of densely interconnected social media systems that enables any system, and particularly any social search engine, to interoperate with other heterogeneous tagging sources and tools in a way that helps to combine and add value to the knowledge work done by users in the ecosystem as a whole.

As pointed out recently by knowledge researcher Tom Gruber in [5], there will be many possible ways for social media systems to collect, interpret, or use tag data, but if we want them to interoperate, there must be at least an ontological commitment to the semantics of tagging, i.e. a common conceptualization of what tagging means and at least some way for them to correlate or connect tag data from one application to another. We add to this the need for a shared conceptualization of how the tagosphere is organized around its core concepts —taggers, networks and sources—, which affect the semantics of the tagging process.

We focus on identifying a shared conceptualization of the tagging activity and the organization in which it is performed, highlighting its immediate benefits for both the tagging and the searching activities. We have left out its formalization (i.e. its specification as an ontology [8]) for the sake of legibility. The process of developing such a conceptualization is open. This conceptualization is open to enhancements that we encourage, and there are other possible approaches.

2.1 Identifying the Organization and Dynamics of the Tagosphere

As shown in Figure 1, the tagosphere is the entire environment of collaborative tagging that emerges from the holistic integration of different folksonomies being dynamically created by folks through the tagging, i.e. their association of tags to Web resources in the tagging system of their choice.

Each tag is anchored to a concept definition from a semantics-enabled tag space (e.g. a Wikipedia entry), or directly to an ontology entry (an ontology is then a valid tag space), in order to identify its intended meaning when applied to a given Web resource. This way each application-specific tag space can be mapped to a shared tag space. This is a valid way of reasoning about the relationship among tag data from different applications without any one application owning a global tag space. Thanks to this, for example, tags from different tag spaces would be said to be equivalent when anchored to the same concept definition.

In addition, we encourage support for the complementary ideas of polarity and voting. These are similar ideas, but convey different semantics On the one hand, the negative tag feature, asserting that a tag should not apply to a given Web resource, helps to handle the collaborative filtering of user-induced spam and incorrect tagging. This feature is even more important when considering an extended tag, i.e. a meaningful, possibly disambiguated tag, because it indicates which concrete meaning does not apply to the resource. On the other hand, the rating feature expresses, for any one search, the user's opinion or vote about how relevant the result is to the keyword used.

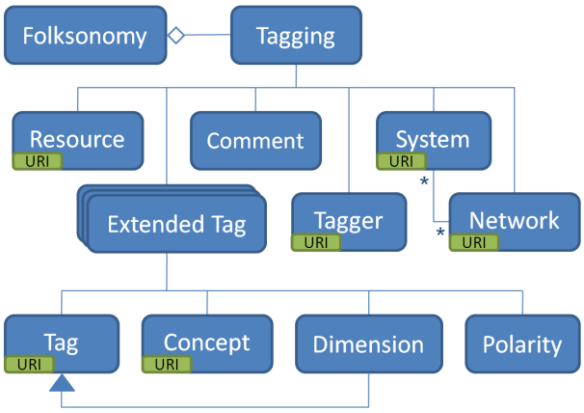

Figure 1. Tagosphere conceptualization.

Finally, we also consider a *dimension* feature for organizing tags when they are intended to express ideas other than what topics a Web resource is about [6]. This feature can be explained as a kind of "metatagging", as explained below.

Each extended tag then can be represented by the following tuple expression:

```
ExtendedTag(Tag, Dimension,
            Concept, Polarity)
```

This expression captures what is to be communicated in different data formats, such as XHTML MicroFormats [1], as we will explain later. An extended tag can exist on its own, as a useful means for anchoring author-designated annotations to Web resources (as many bloggers do today). However, it commonly is part of a tagging assertion about a given Web resource. In this last case, the tagging not only includes the tagged resource and its assigned set of extended tags, but it also accounts for tagging authorship in terms of who did that tagging and the context of which community or social network and which system it was done in.

Each tagging assertion can then be represented by:

```
Tagging(Resource, {ExtendedTag}+,
        Folk, [Network], System, Comment)
```

According to Reed's later refinement of Metcalfe's law on network effects, growth in "value" can even be exponential (instead of quadratic) to the number of users of those networks that can form groups [9], as is the case of most social media Web sites that connect people with similar interests and/or expertise in tightly-knit tagging communities. It is thus important to preserve the nature and the identity of these communities in the tagging expression. To be able to deal with systems that do not explicitly support the notion of community, we consider a tagging system as a kind of community.

We need formal definitions of identity for most core tagging concepts. Following the ideas behind the REST architectural style, we regard each of these concepts as being a resource having a unique identifier (URI) that can be accessed through a uniform interface to obtain its representation:

- Web resources are identified by their URI.
- Extended tags are identified by both a URI from the system-dependant tag space (a term scoped by the system URI, as is common today) and by a global permalink representing a concept definition from a semantics-enabled tag space. This permalink can be considered as the canonical name, irrespective of specific tag syntax used by each application.
- Taggers are identified by a URI. OpenID is also considered as a global, cross-system means to identify taggers, irrespective of how each system identifies them internally.
- Communities are identified by a URI.
- Systems are themselves identified by their URL, which usually serves as an URI prefix for tag, tagger, and community internal identities, and supplies the system-dependant tag space with a valid namespace.

Another interesting tagging-intrinsic notion is that extended tags play a role in the meaning of a tagging that is different from the one played by the resource or the tagger. In principle, they are not interchangeable in a tagging assertion without loss of meaning [5]. For example, if you want to use some sort of "metatagging" without altering the meaning of the different tagging assertions this will give rise to (e.g. "this tag represents a broader concept subsuming other tags" as in http://del.icio.us/adobe), then it would need a different sort of family of relations for metatagging.

We now thoroughly analyze two features of this conceptualization that are closely related to social search: attachment of meaning to tags, and metatagging. Used together, they represent a useful and agile means of unambiguously defining terms used during tagging, and clarifying vague search terms. This improves search results in terms of recall and precision.

3. USING WIKIPEDIA TO ATTACH IDENTITY AND MEANING TO TAGS

In both current tagging systems and search engines, terms chosen as tags or keywords are intended to represent real-world concepts assigned to a given Web resource. However, they cannot explicitly identify these concepts. Imagine you are bookmarking the latest version of the first (introductory) part of the W3C Recommendation for the Web Ontology Language (http://www.w3.org/TR/owl-features), then you will be tempted, and even be advised, to use the term 'owl' as a suitable tag. Nevertheless, the term 'owl' can also refer to "any one of about 220 species of mainly nocturnal birds of prey", to "Owl, the Winnie the Pooh character", etc. As an acronym, OWL also stands for "Object Windows Library". On Jule 15, 2008, Wikipedia offered 16 different meanings for "owl", and 14 more for the acronym.

As far as we know, there is no tagging system or search engine today that can attach the intended meaning to the term used as tag or keyword. This has a negative impact on the search results in terms of recall and precision. To tackle this important problem, any useful conceptualization of the tagosphere needs to consider some notion of tag identification that can attach meaning.

One use of ontologies is for people to state what they mean by formal terms used in any data that they generate or consume (commonly referred to as semantic annotation). In this sense, ontologies will be a useful means for unambiguously defining terms used during tagging, and for clarifying vague search terms.

You could try to build an ontology of all the world's knowledge, and some people still do, but it is hard work and requires solid knowledge/ontology engineering skills and expertise, even if considered as a collaborative development (as in the OntoWorld initiative). Ontologies are, unquestionably, useful for associating terms with concepts, but it is hard to believe that they will ever become widely used by the Web community as the engineered artifact of choice for tagging and for unambiguously and understandably searching resources.

Instead of developing ontologies to tag resources and clean up the emerging folksonomy, we suggest using Wikipedia as a commonly agreed and shared lightweight conceptualization for this purpose. We view Wikipedia entries (each one conveying a different meaning and having an assigned permalink to identify it) as a good way of anchoring a reference to a tag in a manner that two people (or two systems) could agree that they are talking about the same thing. This way, you can continue choosing a tag from any "tag space" and then use a Wikipedia reference to univocally identify it and indicate the semantic concept that you are conveying when you associate that tag with a Web resource.

Wikipedia's built-in disambiguation services (e.g. http://en.wikipedia.org/wiki/Owl_%28disambiguation%29) will also serve as part of the necessary conceptualization infrastructure

to build a recommendation system that will help people to attach meaning to tags and keywords when they are either tagging or searching.

Figure 2. Tag Identification.

Figure 3 illustrates the tagging process as it would be performed in a system that supports tag identification. Continuing with the previous example, your Web agent (either a browser or a plug-in) would ask for the intended meaning of each tag you want to apply to the Web Ontology Language page you are trying to bookmark. To do this, it would list all Wikipedia entries related to the tag. In our example, it would return up to 23 possible meanings for the tag "owl". Then you would have to choose the meaning "Web Ontology Language". After you have chosen the meanings of all selected tags (i.e. identified them) a new tagging assertion would be added to the tagging repository. As we will explain later, the notion of metatagging will help to relax the need for user interaction in this identification process. Figure 2 shows a possible implementation of the above described tagging process through a screenshot taken from a real Web system that is being developed by the authors as a proof of concept for these ideas. The system can currently be accessed and tried at http://jupiter.ls.fi.upm.es/tagosphere, as we mention later in the future work section.

Figure 3 also illustrates how tag identification can also help to find more accurate results in a given search. Imagine you are now looking for resources relating to "Web Ontology Language" using the "owl" keyword. All the different meanings would be listed and you would select your preferred option, as explained above for tagging. Consequently, a Wikipedia identification-aware search engine would retrieve every resource associated with the preferred meaning, whatever terms had been used to convey this meaning in the tagging repository. Therefore, this approach solves —or at least, minimizes— some other inherent tagging-related problems [4] that can stymie people tagging or searching, unless handled as synonyms (i.e. results tagged with both "owl" or "Web Ontology Language" would now be retrieved), lexical anomalies that can emerge in uncontrolled vocabularies, plurals and parts of speech and spelling, and, also, some system constraints like unsupported characters (i.e. "Web.Ontology.Language" and "Web Ontology Language" should have the same meaning).

User feedback based on negative tagging and voting about search results can considerably improve future searches. User votes express opinions such as "this resource is not relevant to that search keyword", therefore influencing the behavior of the search engine as regards the relevance attributed to that result when ranking future related searches. Negative tagging asserts that a tag should not apply to a given search result, and, consequently, corrects a wrong resource tag.

When applied to traditional tagging systems that do not anchor meaning to tags, a negative tagging asserted to a result from a given search for a meaning-aware keyword across multiple systems expresses that the result in question does not apply to the meaning anchored to the keyword. It should not therefore be returned in successive searches for that meaning. Nevertheless, that result could continue being returned in other searches for the same keyword anchored to different meanings.

The proposed mapping between each application's and the Wikipedia-defined tag space is a valid way of reasoning about the relationship (e.g. equivalence) among data tagged in different applications. It makes it possible to exchange, compare, and reason about the tagged data without any one application owning a global 'tag space' or folksonomy. This way, http://del.icio.us/tag/owl, http://del.icio.us/tag/webontology-language, and http://ma.gnolia.com/ tags/owl could be said to be equivalent tags if they were all anchored to the same Wikipedia entry http://en.wikipedia.org/wiki/Web_Ontology_-Language, but not if they had other meanings.

We also want to highlight how this Wiki-based approach leads to a collaborative process of convergence in terms of (a) the set of concept definitions (i.e. Wikipedia entries) used to identify the semantics conveyed by a tag when applied to a given Web content, and, more importantly, (b) the set of tags identified by the same concept (i.e. Wikipedia entry). This assures that the tag set used by the user over time is coherent (i.e. the user will always be advised to use the same tag to convey the same concept), as well as consensus among users on which is the preferred tag set for conveying a given concept. This would otherwise be difficult, if not impossible, to achieve [7] and, again, will result in better search results in terms of both recall and precision in the end. All of this takes the form of a recommendation, and does not restrict the eligible set of terms.

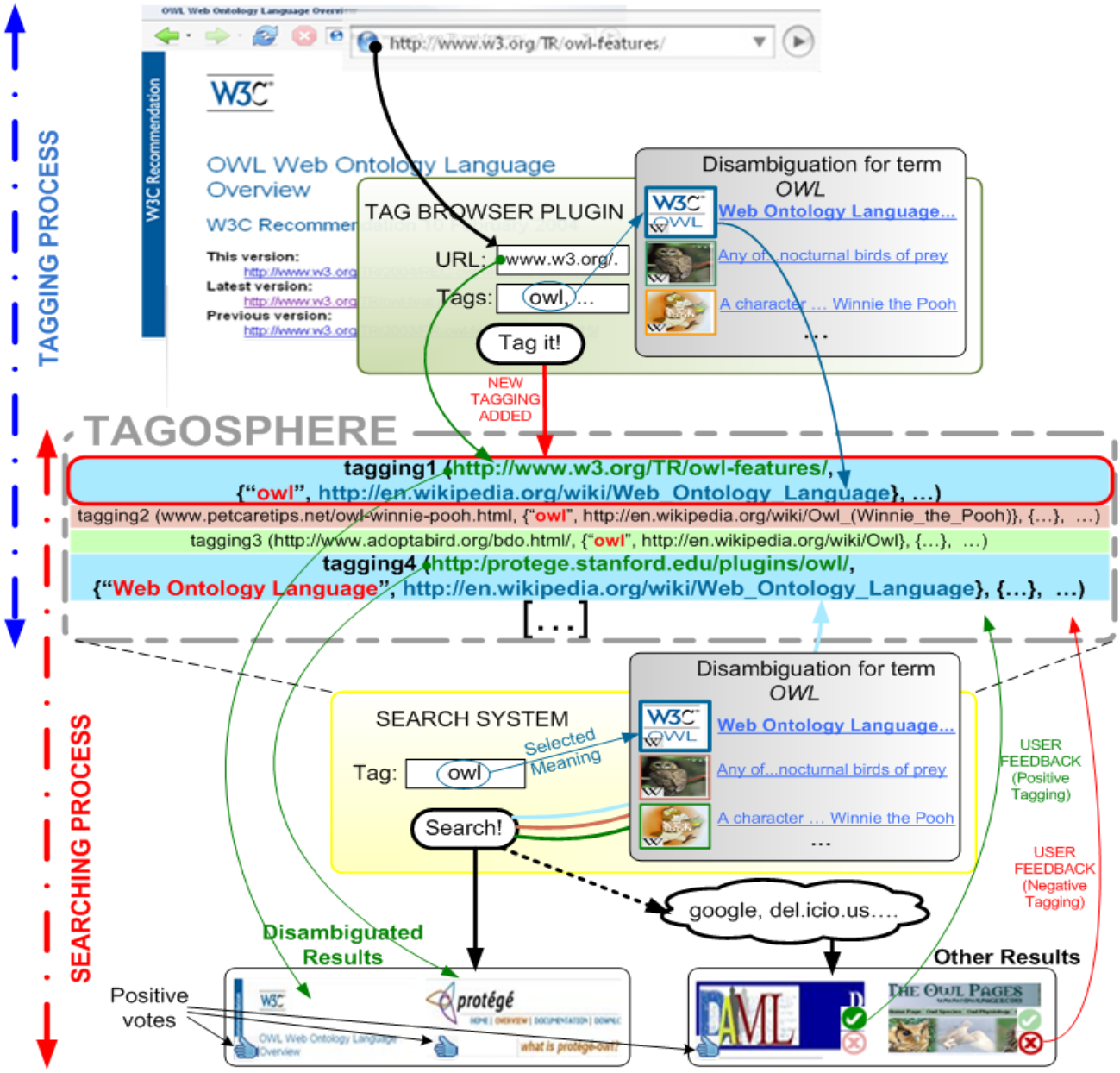

Figure 3. Wikipedia-based framework for handling meaning for both tagging and searching

4. DEALING WITH TAG CATEGORIZATION

Our conceptualization accounts for the concept of "metatagging" or tag categorization. One use of metatagging is as a good means for organizing identified tags within a given context, represented by another tag expressing a broader concept (e.g. a subsuming concept). For example, during a search session a user may want to find resources about "Semantic Web" (e.g. his/her user profile or navigation history would help in automatically identifying this preference). It would therefore be reasonable to assume that he/she wants to use tags (e.g. our example for 'owl') with meanings already related to Semantic Web, which then do not need to be identified (and possibly disambiguated) individually.

This use of metatagging improves both the tagging and search process by relaxing the need for user interaction while identifying concepts for every tag used in a given tagging/search or even throughout a tagging/search session. Also, networks can benefit enormously from the metatagging performed by their members.

In addition to the above benefits, metatagging also helps to deal with other habits observed by analyzing current tagging community activity, mainly the use of tags to express ideas belonging to cognitive dimensions or metadata other than what express the topics that the Web resource is about. These include what kind of thing a resource is (e.g. 'blog'), who owns it, how you organize tasks on it ('to read'), or even why it is important for you ('research'). The information provided by a number of these dimensions is related to and/or only relevant for the tagger, and is

of little or no use for other people's searches. Others, like the information acting as metadata, can be used in searches as filters based on media type, kind of resource, etc.

Metatagging can be carried out in a wiki-based collaborative form, where users can freely organize their topics of interest around broader concepts of their choice.

5. MICROFORMATS AS AN OPEN EXTENDIBLE DATA STANDARD FOR INTEROPERATION

A shared conceptualization is not all we need to achieve interoperability in the tagosphere. Lack of an open data standard is also a major issue. Used together with an ontological commitment to the semantics of tagging, a data standard would make it possible to easily collect and remix tag data, enabling the development of social search engines that work across tagging services and bookmark repositories. It would also make it possible to write cross-application AJAX scripts and other innovative services, enabling considerable improvements in user search experience.

The well-known Rel-Tag and xFolk MicroFormats constitute such an open data standard. In their present form, they are useful for identifying a set of tagged resources in a XHTML document, along with all the tags associated with each resource. Generally, they add lightweight semantics to web content. Nevertheless, they currently do not convey the necessary elements to enable systems to interoperate at a semantic level according to the proposed conceptualization.

Thanks to their design, Rel-Tag and xFolk are both easy to extend. Instead of developing a new "standard" from scratch, we opted to extend these MicroFormats, and add additional ones, to include the necessary elements of our conceptualization of the tagosphere.

Note that several alternative mechanisms, apart from MicroFormats, would also be suitable for this purpose. These include (a) creating a separate RDF/XML description, and (b) creating a separate XML description (and using the <link> element to link it from HTML/XHTML in both cases, if necessary).

5.1 Extending Rel-Tag to Anchor Wikipedia URIs to Author-Designated Tags

By adding rel="tag" to a hyperlink, a resource indicates that the destination of that hyperlink is an author-designated "tag" for this resource or for a major portion of it. Following both XHTML and MicroFormat principles for defining tags as a part of a more specialized format, we have built on and extended Rel-Tag to anchor the identifying Wikipedia URI to the author-designated tag in order to provide a specific meaning. For example, by placing this link on a page,

```
<div class="meta">  <!--extended Rel-Tag-->
  <a href="http://del.icio.us/tag/owl"
     rel="tag">owl</a>
  <a href=http://en.wikipedia.org/wiki/Web_
     Ontology_Language
     rel="tagmeaning" >Web Ontology Language,
     a markup language for […]</a>
</div>
```

the author is indicating that the page (or some portion of the page) has the tag "OWL", meaning Web ontology Language. It is now the sum of the linked page and the link expressed in the tagmeaning class that defines the tag.

Each extended tag is now comprised of the proposed Rel-Tag extension, along with its associated reminder information —scope and polarity— represented by class attributes. Therefore, an extended tag will be as follows:

```
<div class="meta"><!--extended tag-->
   […]<!--extended Rel-Tag-->
 <div class="scope">about</div>
 <div class="polarity">positive</div>
</div>
```

Since the last path segment is the only part of a tag space URL of which any structure is required, a tag space URL can be hosted at any domain. Therefore, page authors may even be tempted to choose to link to a tag directly at Wikipedia in order to provide a specific meaning. This is considered system-specific behavior, and is not the rule. Therefore, the proposed extension to Rel-Tag is still needed.

5.2 Conveying Semantics-aware Tagging Information

The xFolk MicroFormat is a general decentralized syntax for tagging arbitrary URLs or external content. xFolk currently describes the data published by bookmarking services using a simple schema. This schema consists of a set of tagged links, each of which is characterized by a title for the entry, tags for that link, and an extended description or summary of that link.

The primary goal behind the design of the xFolk recommendation is to ease adaptation to current practices. Therefore, few assumptions are made as to the exact kinds of elements used for an xFolk entry. Rather, the work of defining semantics is left entirely to the class and rel (in the case of Rel-Tag) attribute values. Semantic elements within xFolk entries may also be nested at arbitrary depths.

We have added elements for conveying a set of extended tags anchored to the resource identified by the taggedlink element to the remaining data-entries considered in the original recommendation.

```
<div class="xfolkentry"> <!— An xFolk entry
considered as aggregated data -->
  <div><a class="taggedlink" href="a link">Web
Ontology Language Primer</a></div>
  <div class="description"></div>
  <div class="meta">
    <div class="meta">
    <a href="http://del.icio.us/tag/owl"
rel="tag">owl</a>
    <a rel="tagmeaning"
href="http://en.wikipedia.org/wiki/Web_Ontology_La
nguage">Web Ontology Language, a markup language
for [...] World Wide Web.</a></div>
    […]<!--more extended tags-->
  </div>
</div>
```

Note that xFolk is still valid for representing aggregated data from a tagging repository. Nevertheless, we also aim to use xFolk as a means to represent a tagging. Therefore we further extended it to convey:

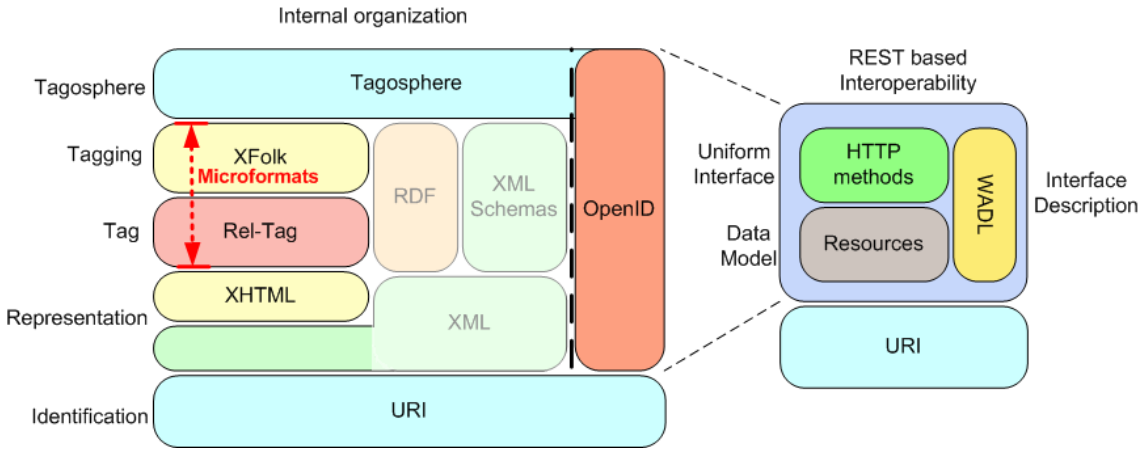

Figure 4. Language and protocol stacks supporting interoperability

```
<div class="xfolkentry"> <!—An xFolk entry
considered as a tagging -->
  <div><a class="taggedlink" href="a link">Web
Ontology Language Primer</a></div>
  <div class="description"></div>
  <div class="meta">
    <div class="meta">
      <a href="http://del.icio.us/tag/owl"
rel="tag">owl</a>
      <a rel="tagmeaning"
href="http://en.wikipedia.org/wiki/Web_Ontology_La
nguage">Web Ontology Language, a markup language
for [...] World Wide Web.</a></div>
    […]<!--more extended tags-->
 <a rel="folk" href="Folk's valid openID
URI">Folk's valid openID URI</a>
 <a rel="system"
href="http://del.icio.us">Delicious</a>
</div>
```

In the context of an architecture for interoperability, these data formats are part of a stack of existing and widely adopted Internet languages and protocols. The tagosphere is located on top of that stack, and represents the global tagging aggregate on which search engines operate. The following section describes the proposed architectural stack for interoperability.

6. ARCHITECTURAL STACK FOR INTEROPERABILITY

The tagosphere has an architecture for interoperability based on a hierarchy of existing and widely adopted Internet languages and protocols. Each language and protocol exploits the features and/or extends the capabilities of the layers below. The relationships between the languages and protocols actually leads to a stack, like the one illustrated in Figure 4, where data format-oriented languages, e.g. xFolk, rel-tag, etc., and communication-oriented languages/ protocols are split up in two separate towers.

All XML data formats rely on a URI scheme for identification and on OpenID to convey user information. Microformats are based on XHTML at a representation level and cover both tag and tagging layers (i.e. the proposed extensions to rel-tag and xFolk, respectively). As previously mentioned, RDF/XML and XMLSchema are also considered as valid data formats for the same purpose. The tagosphere layer, which is located on top of the stack, represents the tagging aggregate on which search engines operate.

Finally, the protocol stack for communicating tagging data between Web applications is based on the REST [2] architectural style. This REST-based interoperability model is founded on (1) a resource oriented data model, (2) the well-known HTTP verbs (or methods) as a uniform operational interface to these resources and (3) the Web Application Description Language (WADL) as a standard means of describing in XML both the data model and the operational interface to the resources in the data model.

7. RELATED WORK

In [5] the author lays out some of the issues and challenges for designing a specification of tag concepts that might enable services for analyzing and reasoning over tag data across applications. Nevertheless, it remains work in progress, and focuses exclusively on interoperability. It does not tackle specific considerations for improving search. Wikipedia is garnering growing attention in a number of related research areas as a means to represent and reason about meaning. Results from these areas would fuel some of the ideas presented here. As an example, [3] proposes a method to represent the meaning of texts in a high-dimensional space of concepts derived from Wikipedia. Methods alike would help improving our approach by assessing the relatedness of keywords in complex search queries.

8. CONCLUSIONS AND FUTURE WORK

Even though it is considered the next big breakthrough in search and social search, tagging-based search is still in its infancy and currently faces a number of unresolved, both sociological and technical, problems. We have evidenced some significant examples of the technical snags. Current tagging systems are still tapping into their own community of users to designate contents as share-worthy and to search what other users find relevant. Additionally, they all suffer from the inherent ambiguity and imprecision of language. Both problems have a significant negative impact on both the precision and the recall of the search results.

We have pointed out that the availability of a common conceptualization of what tagging means and what the tagosphere is, suitably furnished with the capability of attaching meaning to and categorizing tags gives tag-based search engines a fighting chance. Search results are then more precise because they refer exclusively to the user's intended keyword meaning, and user-induced spam and incorrect tagging is filtered out. Likewise, results recall is significatively greater, because, first, the search space goes beyond the system boundaries and occupies the global tagosphere and, second, synonyms, acronyms, lexical variations, etc., can now be considered. From the tagging point of view, it also leads to a collaborative process of convergence that boosts the coherence of the tag set used by the user over time and favors consensus among users on which are the preferred tags for a given meaning. Like disambiguation, this helps to clean up the emerging folksonomy, without restricting the tag set.

As a proof of concept for the ideas presented in this paper, we are currently experimenting with a full-fledged Web application prototype developed by the authors. This Web application represents a semantics-aware social bookmarking system that implements the rationale and fundamentals expressed above. It can currently be accessed and tried at http://jupiter.ls.fi.upm.es/tagosphere. Following one of the defining characteristics of Internet era software, which considers it as a service delivered and maintained on a daily basis instead of as an artifact delivered as a product, our Web system is offered in a "perpetual beta" and new features are continuously being added on a regular basis as part of the normal user experience. We are therefore engaging users as real-time testers and we have instrumented the service so that we know how people use the new features to make the most of them.

9. ACKNOWLEDGMENTS

This work is supported in part by the European Social Fund and Comunidad Autónoma de Madrid under their Researcher Training programs.

10. REFERENCES

[1] T. Çelik and D. Powazek. Rel-Tag: Draft specification. Microformats, 2005. Bud Gibson, The Community Engine. xFolk (RC1): Draft specification, Microfornats, 2006, Available at http://microformats.org/wiki.

[2] R. T. Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. Ph. D. Dissertation. University of California, Irvine 2000

[3] E. Gabrilovich and S. Markovitch. Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, January 2007.

[4] S. Golder and B. Huberman. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2):198-208, 2006

[5] T. Gruber. Ontology of Folksonomy. Invited keynote to the *First on-Line conference on Metadata and Semantics Research (MTSR'05)*, 2005.

[6] M. Kipp and D. Campbell. Patterns and Inconsistencies in Collaborative Tagging Systems. *Annual General Meeting of the American Society for Information Science and Technology*, 2006.

[7] K.F. Lawrence, M.C. Schraefel. Freedom and Restraint Tags, Vocabularies and Ontologies. *Information and Communication Technologies (ICTTA'06)*, pages 1745-1750, IEEE Computer Society, 2006.

[8] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W. R. Swartout. Enabling technology for knowledge sharing. *AI Magazine*, 12(3):16-36, 1991. Conger., S., and Loch, K.D. (eds.). Ethics and computer use. Commun. ACM 38, 12 (entire issue).

[9] D. P. Reed. The law of the pack. *Harvard Business Review*, Harvard Business School Publishing Corp., February 2000.