Crowdsourced Semantics with Semantic Tagging: "Don't just tag it, LexiTag it!"

Csaba Veres

Institute for Information and Media Science,
University in Bergen, Norway
Csaba.Veres@infomedia.uib.no

Abstract. Free form tagging was one of the most useful contributions of "Web2.0" toward the problem of content management and discovery on the web. Semantic tagging is a more recent but much less successful innovation borne of frustration at the limitations of free form tagging. In this paper we present LexiTags, a new platform designed to help realize the potential of semantic tagging for content management, and as a tool for crowdsourcing semantic metadata. We describe the operation of the LexiTags semantic bookmarking service, and present results from tools that exploit the semantic tags. These tools show that crowdsourcing can be used to model the taxonomy of an information space, and to semantically annotate resources within the space.

1 Introduction

The emergence of "Web2.0"[1] brought a number of innovations which changed the way people interact with information on the World Wide Web. The new paradigms made it easy for anyone to contribute content rather just consume. One of the early success stories was *social tagging*, which gave rise to *folksonomies*[2] as a way to organise and find information on the Web through emergent shared vocabularies developed by the users themselves. Social tagging for content management and discovery became very popular in commercial services like the photo sharing site flickr.com and the bookmarking site delicious.com. These successes prompted some commentators to declare victory of user driven content tagging over the "overly complex" technologies of the semantic web. Perhaps most famously, in a web post entitled "Ontology is Overrated: Categories, Links, and Tags" Clay Shirky argued that any technology based on hierarchical classification (including ontologies) was doomed to fail when applied to the world of electronic resources [1]. Instead, simple naive tagging opened the door to crowdsourced content management, where dynamic user contributed metadata in a flat tag space offered a breakthrough in findability.

However, researchers and information architects soon began to point out the limitations of unconstrained tagging for enhancing information findability. [2] identified a number of problems with tagging, which can limit its effective usefulness. Among the problems were tag ambiguity (e.g. apple - fruit vs. apple - company), idiosyncratic

---

treatment of multi word tags (e.g. vertigovideostillsbbc, design/css), synonyms (e.g. Mac, Macintosh, Apple), the use of acronyms and other terms as synonymous terms (e.g. NY, NYC, Big Apple), and of course mis spelled and idiosyncratic made up tags. These factors limit the use of tags in large scale information management. For example [3] discuss limitations of searching with tags, which is necessarily based on syntactic matching since there are no semantic links between individual tags. Thus, searching with "NYC" will not guarantee that results tagged with "NY" will be retrieved.

Semantic tags or *rich tags* as they are known in the context of social tagging, emerged as a way to impose consistent and refined meanings to user tags [4]. The two best known semantic tagging sites were Faviki[3] and Zigtag[4] (the latter now appears to be defunct). Each site expected users to use tags from a large collection of provided terms. Faviki used WikiPedia identifiers, while Zigtag used a "semantic dictionary". Both sites also allowed tags which were not in their initial knowledge base. In the case of Faviki, users can link undefined tags to a web page which best represents the tag. The web page is located by a simple web search. Zigtag allowed the use of undefined tags, with the expectation that users would later return and provide definitions of the tags. However a large proportion of tags remained without definition, leading to a mishmash of defined and undefined tags.

LexiTags [5] was initially developed for similar reasons, as a tool for content management with rich tags. But as a semantic application it also had higher aspirations, to provide a platform and a set of front end tools designed to crowdsource the semantic web. By providing intuitive access points (APIs) to user generated metadata with semantic tags, the platform was expected to outgrow its initial purpose and provide novel new benefits for its users while at the same time generating semantic metadata for general consumption. The intuition is that users should have systems which behave as Web2.0 at the point of insertion, yet as Semantic Web at the point of retrieval. In this paper we present the core LexiTags system, and describe some tools we have developed to capitalise on the crowdsourced semantic metadata.

2. LexiTags

LexiTags[5] is an acronym for "Lexical Tags", from the fact that the tags are primarily lexical items, or natural language dictionary words. They are disambiguated through the use of an interface which presents the user with a set of choices from WordNet, an electronic lexical database [6]. The main content bearing units in WordNet are *synsets*, which are represented by contextually synonymous word meanings grouped in a single entry. The word *couch* for example is represented by the synset {*sofa, couch, lounge*}. But *couch* is of course an ambiguous word whose alternate meaning appears is a second synset {*frame, redact, cast, put, couch*} as in "Let me couch that statement …. ". There is therefore no ambiguity in WordNet because every synset represents a unique meaning for a word string. LexiTags presents such synsets, and a short gloss, to help users chose their intended meaning.

The use of synsets as tags can combine precise definitions without the need to adopt a set of idiosyncratic keywords. Mappings can be set up between synsets and any other vocabulary, enabling specific keyword markup through a natural language interface.

3   http://www.faviki.com/pages/welcome/

4   http://zigtag.com

5   http://lexitags.dyndns.org/ui/webgui/

The success of mapping efforts can of course vary. While highly technical ontologies could prove difficult, lightweight ontologies and taxonomies like schema.org are not problematical, and in section 3.2 we will see a tool that makes use of exactly such a mapping.

Fig. 1 shows a detail of the main interface of LexiTags, which is mainly a simple list of URLs that have been bookmarked. Clicking on the URL opens a new window with the web site. The user tags appear below the URL. Hovering the mouse over these tags pops up a definition which clarifies the precise sense of that tag. Clicking on a tag will open a new tab which shows bookmarks tagged with the same tag sense.
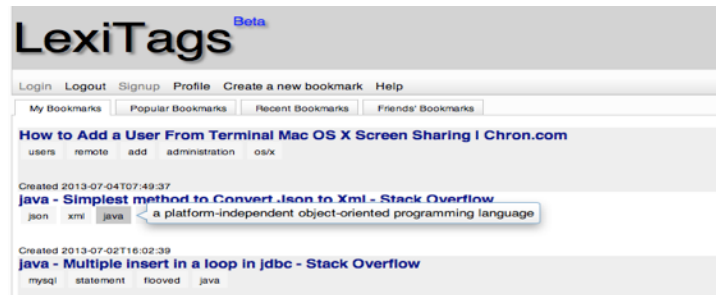

Fig. 1. The main LexiTags interface

Tags for a bookmark are entered freely in the text box near the bottom of the "Edit Bookmark" window (fig. 2) which is popped up through the use of a bookmarklet. The user types one tag at a time into the text box and presses enter which puts each tag in a list above the text box, initially in red colour. Users can simply enter tags as they like, as in most other tagging sites. Finally, users click on each undefined tag to add disambiguation through the final "Editing the tag …" window, shown in figure 3.
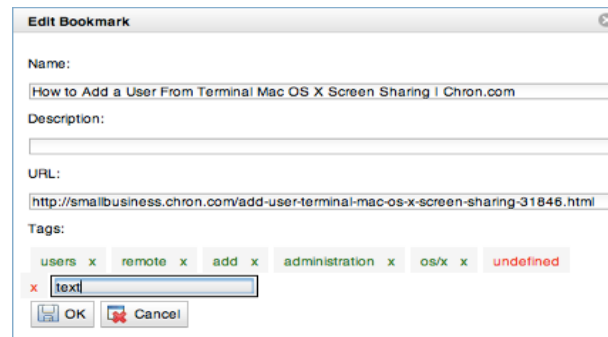
Fig. 2. Window for adding and editing tags

The "Edit tag .." window shows the possible interpretations from WordNet, or DBPedia if WordNet does not have an entry for the tag word. This is the case most often when the tag is the name of a company or a person or a new technology. DBPedia also includes a large number of common abbreviations, such as "NYC". In addition, DBPedia defines mappings to WordNet synsets for many concepts, which helps fill gaps in WordNet. Unfortunately the coverage is not complete, so "NYC" for example is not linked to the synset for New York in WordNet. Users must select the sense that best matches their intent. The choices are ordered by word frequency, and our experience suggests that the intended sense is amongst the first two or three senses. Since the main point of tagging is for future

retrieval, it would make sense if people tended to avoid words in their obscure, low frequency senses. However this is just conjecture at the moment, and we are investigating other methods for optimal ordering. One approach is to weight the rankings according to the aggregate distance of each candidate sense to a context tag provided by already disambiguated tags. Another approach is to personalise the rankings so that each user's own tagging history influences the ranking of the candidate senses.

As each tag is disambiguated by the user, it turns green. Any tag left disambiguated is deleted when the user presses the "OK" button, so every tag in the LexiTags platform is a disambiguated, defined term.
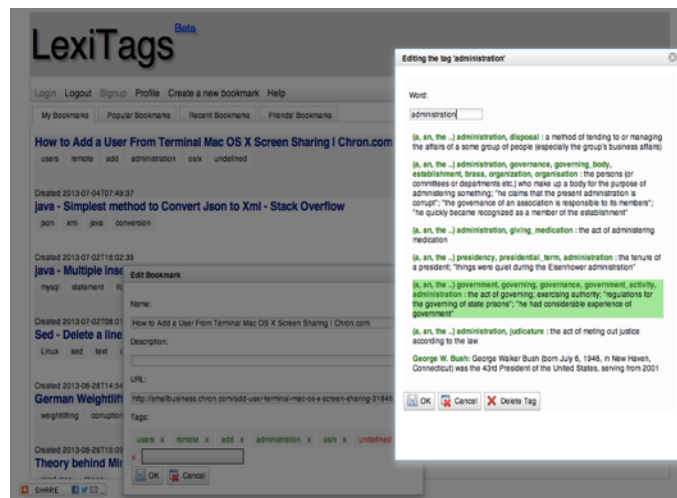
Fig. 3. Disambiguation window

3. The Lexitags ecosystem

If LexiTags were just a bookmarking service with rich tags, there would be little to differentiate it from Zigtags. But the idea was to use the bookmarked sites and their tags as a starting point for a set of tools that extracted value from the tags. As such, LexiTags should be seen as a platform to expose crowdsourced semantic metadata to clients, both for creation and consumption of metadata. In terms of content creation clients, we are developing an iPhone app for tagging photographs with LexiTags, as described in [5]. Due to space limitations we are unable to discuss alternative input applications, but instead describe two applications for consuming the metadata. One creates a content taxonomy, the other produces metadata for the web.

3.1 Content taxonomy

[7] discuss SynsetTagger[6], which was developed to consume LexiTags tags automatically, but has thus far only been demonstrated in manual mode to create lightweight ontologies from user input. SynsetTagger makes use of select WordNet

---

6   http://csabaveres.net/csabaveres.net/Semantic_Apps.html

relations to construct a lightweight ontology by inferring additional nodes from the provided tags. The most important link for nouns is *hyponymy/hypernymy* which are the semantic relations otherwise known as *subordination/superordination*, *subset/superset*, or the IS-A relation. A concept represented by the *synset* {x, x′, . . .} is said to be a *hyponym* of the concept represented by the *synset* {y, y′, . . .} if native speakers of English accept sentences constructed from such frames as "an x is a (kind of) y" [8], [9]. Another relation used by SynsetTagger is *meronymy*, or the *part/whole* relation. A concept {x,x'….} is a *meronym* of a concept {y, y′, . . .} if "an x is a part of y", and it is a *holonym* if "a y is a part of x". There are several other important relations in WordNet, some of which will be mentioned in the case study.

SynsetTagger works by constructing the hypernym chain and pruning nodes if their information content is trivial, which is determined by counting the outward edges from each node and eliminating each node that falls below a threshold value. For example, fig. 4 shows a taxonomy constructed from the input synsets coloured green. The orange coloured inferred hypernyms will be discarded because the have only a single outgoing edge (or are too close to the top level node), and the clear white ones will be kept. The grey nodes are both inferred and asserted, and will also be kept in the final taxonomy. The nodes that are kept are called *informative subsuming nodes*.

Fig. 4. Complete hypernym chain constructed from input synsets

The tool has a number of user configurable parameters which determine the final selection of nodes. Two important ones are the number of outgoing edges, and the distance from the topmost node. The optimal selection is a matter of trial and error with a given input set. The pruning mechanisms are similar to the "*nearly* automated" approach to deriving category hierarchies for printed text [10], but SynsetTagger differs in that it allows users to adjust the parameters and receive immediate visual feedback about their consequences on the constructed taxonomy. It is intended as an interactive tool to give users a sense of control over their content.

3.2 Metadata for the Web

MaDaME (Meta Data Made Easy) is a tool for embedding semantic metadata into web sites. Its development was spurred by the release of the schema.org initiative, which is a *type* schema meant to be used by web masters to add structured metadata to their content. The incentive for web masters to use the schema is that web sites that contain markup will appear with additional details in search results, which enable people to judge the relevance

of the site more accurately and hopefully increase the probability that the site will be visited.

Clearly this is an important development in the effort to crowdsource semantic web content. However, the schema was designed specifically for the use case of search, and both the semantics and the preferred syntax reflect that choice. In terms of semantics, the schema contains some non-traditional concepts to fulfil its intended use. For example there is a general class of *Product* but no general class for *Artifact*. There are also odd property ascriptions from the taxonomy structure, so, for example, *Beach* has *openingHours* and *faxNumber*. In terms of syntax, there is a very strong message that developers should use the relatively new microdata format rather than the more popular RDFa web standard[11]. This is unfortunate because it makes metadata from schema.org incompatible with many other sources of metadata like Facebook's OGP[7].

A strong motivation for MaDaME was to create a tool that would not only help web designers apply schema.org markup to their web sites, but to simultaneously inject ontology terms from other standard sources into their web sites in the RDFa standard. In other words, to maximise the crowdsourcing potential offered by schema.org. This was achieved through mappings between WordNet and schema.org, as well as SUMO [12]. In normal operation users select key words in their web sites, which are then disambiguated using the LexiTags interface. The WordNet synsets are stored, and their most appropriate mapping to schema.org and SUMO computed. These are then inserted into the html source, and made available to the web designer for further refinement.

While MaDaME is currently presented as a standalone tool[8] it can also be used to automatically annotate any web site bookmarked on LexiTags. These annotations can be sent to the maintainers of the web sites with a cover letter explaining the purpose of the markup. Alternatively, the markup could be stored on the LexiTags platform and offered through an API.

4. Results

We present a short evaluation of SynsetTagger on a set of approximately 100 bookmarks for a single user on LexiTags, and then the metadata generated for one web site on MaDaME.

Fig. 5 shows a portion of the taxonomy generated from the semantic tags used on the set of 100 bookmarks. The asserted green tags are used to build the hypernym chain from WordNet. If any node is selected in the interface, the set of connected nodes will also be highlighted, making it easier for users to understand the relationships in the taxonomy. For example the asserted tag *conference* can be seen as a kind of *meeting* which in turn is a kind of *gathering*, and so on. The orange nodes are inferred hypernyms, but they will be rejected in the final taxonomy for one of two reasons; a) they have fewer than the specified number of children, or b) they are closer than the selected cut from the *entity* node.The white nodes are the inferred nodes which will be retained in the final export. Users can therefore manipulate the two parameters until the desired level of generality is reached.

We will see that in general a lower number of requisite children and a lower cut will result in fewer nodes. This may be counter intuitive at first, but the rather straightforward explanation is that more and more asserted tags fail to find a suitable subsuming concept when the criterion for retaining the concepts becomes more stringent.

---

7   http://ogp.me

8   http://csaba.dyndns.ws:3000

Fig. 5. Taxonomy from approx. 100 bookmarked sites

Fig. 6 shows the hyponyms of *entity* that are retained with "children" set at two and "cut" also set at two. The black arrow mark at the right side of each oval shows that the node can be extended to reveal more children. Every tag except for *weather* found a more general subsuming hypernym when the criterion for subsuming nodes is lax. The tags which do not have a subsuming concept are noteworthy because they represent unusual bookmarks which stand apart from the rest. The subsuming concepts themselves are very general in this example, and their utility for browsing the bookmarked resources is questionable.

Fig. 6. Inferred hypernyms with lax criteria

Fig. 7 shows the same set of tags with "children" set at 5. Since the criterion for subsuming concepts is much higher, there are many more tags which do not have a subsumer. For example the two tags *fund* and *date* are both a kind of *measure* in fig. 6. But these are the only two kinds of *measure* in the tag set, so with a criterion of 5 children, *measure* is no longer considered as an informative subsuming node. On the other hand the remaining subsuming concepts have differentiated and are now somewhat more specific. For example *group* is replaced with the more specific *organization* which is one of only three kinds of *group* in the tree. So *group* is discarded but *organization* is kept because there are many different kinds of *organization* in the tree. Overall the constructed taxonomy is much more useful for browsing because the general categories are now more informative, and the outliers are not forced into overly general categories.
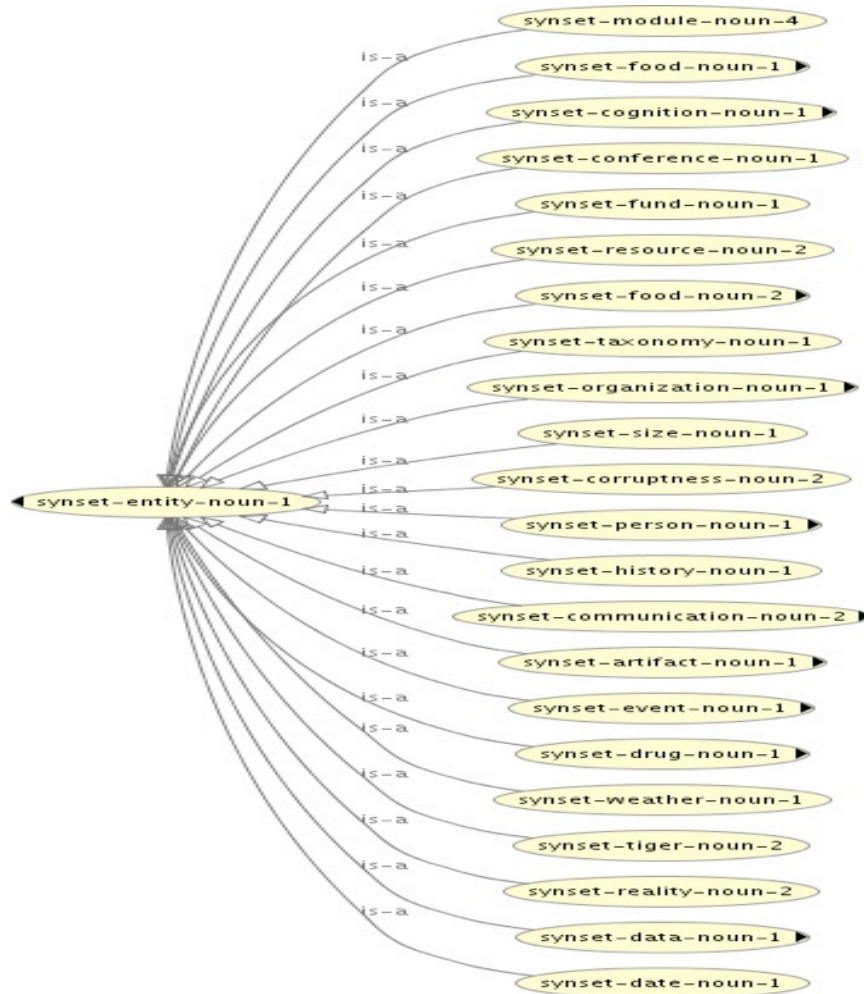
Fig. 7 Inferred hypernyms with stringent criteria

Fig. 8. shows that expanding the *artifact* node reveals some useful sub categories to browse. In general, users can easily configure the parameters to arrive at an optimal taxonomy for their needs.
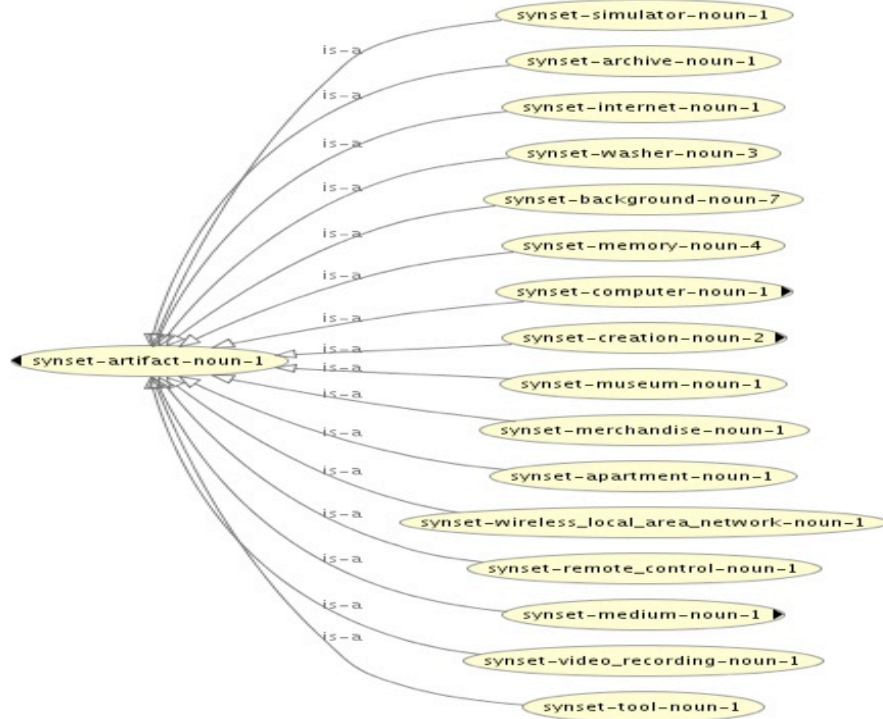
Fig. 8. Different kinds of artefacts in the bookmark set

It is at this point where LexiTags becomes a system that behaves as Web2.0 at the point of insertion, yet as Semantic Web at the point of retrieval. Since the tags have precise definitions they avoid the pitfalls of free form tags like vagueness and ambiguity, and because the defined tags participate in various relations they add value to the asserted tags. They become self organising, with a neat hierarchical structure emerging automatically. Even in this small example of 100 bookmarks and around 400 tags the emerging generalisations form a useful browsing hierarchy. We expect the categorisations to improve with more bookmarks and tags, resulting in a situation where a growing set of bookmarks leads to more organisation rather than complete chaos as seen in traditional free form tagging systems. The users must gain such benefits because they are asked to perform a little more work at the point of insertion. An important point of our work is to enhance LexiTagging services to provide additional benefit to the users, to encourage them to contribute semantic metadata.

The emergent taxonomy can be extended with additional related terms from WordNet. SynsetTagger already has the functionality to add some additional relations to enrich the ontology. Currently these include the two *part-of* relations, and the *domain terms*. For example *movie* has meronyms *episode*, *credits*, *subtitle*, *caption*, *scene*, *shot,* and domain terms *dub, synchronize, film, shoot, take, videotape, tape, reshoot*. These can all be added with the appropriate relations. In addition the synonyms appearing in each synset could also be included. Sometimes this is quite rich, as is the case in our example of *movie*, whose synset consists of {*movie, film, picture, moving picture, moving-picture show,*

motion picture, motion-picture show, picture show, pic, flick}. WordNet also provides a set of *coordinate terms* which are nouns or verbs that have the same hypernym as the target word. Once again *movie* has a number of coordinate terms including *stage dancing*, *attraction*, *performance*, *burlesque*, *play*, and *variety show*.

A final source of data is user tags which are adjectives or verbs, and were not used in the construction of the taxonomy. These tend to be descriptive words that are suitable for use as properties, e.g. *Hungarian*, *fine_tune*, *synchronize*, *semantic*, *amusing*, and *open*. When all of this extra information is added to the taxonomy it results in a greatly enriched ontology which can be used to provide additional services like matching users against one another, or to aid content discovery and recommendation.

As part of the crowdsourcing effort we are planning to enable the export of the generated ontologies and the URLs to which they apply. This would include a list of topics and descriptions where available, as well as relations to other topics of interest. The semantic metadata for each URL would be stored on our servers and made available through an API.

The second form of metadata creation makes use of other vocabularies that have been mapped to WordNet as in the mapping tool MaDaME, which can contribute schema.org and SUMO metadata for any bookmarked URL.

Consider for example one URL bookmarked in the LexiTags data, http://www.imdb.com. This site was tagged with *trailers, information,* and *movies.* When submitted to MaDaME, it can automatically generate markup that can be inserted into the HTML site, or provided as additional data through the aforementioned information API. The mappings generated from the tags are shown in fig. 9. Note that the <span id> is currently assigned to random words at the beginning of the text, and this needs to be inserted into an appropriate location in the original HTML if it is used to mark up the page directly. The example shows that MaDaME currently generates markup from three vocabularies. The original WordNet synset is preserved, as well as the corresponding SUMO class. The SUMO class mappings tend to be quite precise because there is an extensive set of mappings readily available for SUMO[9]. On the other hand the mappings to schema.org are significantly more sparse, because the schema.org types are considerably fewer in number. In cases where no exact match is found a heuristic procedure is used to determine the closest match, and this can result in overly general or erroneous mappings. For example *trailer, preview* is mapped to *schema:Intabgible* whereas a more appropriate mapping might be to the concept *schema:VideoObject* (CreativeWork > MediaObject > VideoObject). On the other hand, many concepts simply do not have a more precise mapping in schema.org, and the mapping of *information, data* to *schema:Intangible* appears to be correct.

9   http://sigma-01.cim3.net:8080/sigma/Browse.jsp?kb=SUMO&lang=EnglishLanguage

```
<span id="madame-NewsDesk-1" class="tagged"
typeof="sumo:Advertising schema:Intangible wn:synset-preview-
noun-1" about="http://csaba.dyndns.ws:3000/load?
q=51dbf00cc6765c3498000002#madame-NewsDesk-1" data-original-
title="">NewsDesk</span>
<span id="madame-movie-1" class="tagged"
typeof="sumo:MotionPicture schema:Movie wn:synset-movie-
noun-1" about="http://csaba.dyndns.ws:3000/load?
q=51dbf00cc6765c3498000002#madame-movie-1" data-original-
title="">movie</span>
<span id="madame-familiar-1" class="tagged"
typeof="sumo:FactualText schema:Intangible wn:synset-data-
noun-1" about="http://csaba.dyndns.ws:3000/load?
q=51dbf00cc6765c3498000002#madame-familiar-1" data-original-
title="">familiar</span>
```

Fig. 9. Automatic schema.org and SUMO annotation for imdb.com

The key point is that metadata from various different namespaces can automatically be made available for different consumers, simply as a side effect of bookmarking. Search engines can use the schema.org markup, while other services can use the SUMO or WordNet classifications. If the bookmarking service were to grow in popularity, it could become a large repository of schema.org markup for a large set of URLs. Search engines could, presumably, use this markup as if it was embedded in the web sites themselves. But MaDaME also provides a user interface that can be used to greatly enhance the schema.org markup using a drop down form as shown in fig. 10. The form includes a text box for all the possible properties of the schema type which is selected for a word in the text. Users could select words in addition to the tags already assigned, and add these to the schema. The figure also shows the extended markup generated by the tool for the word *movie*.
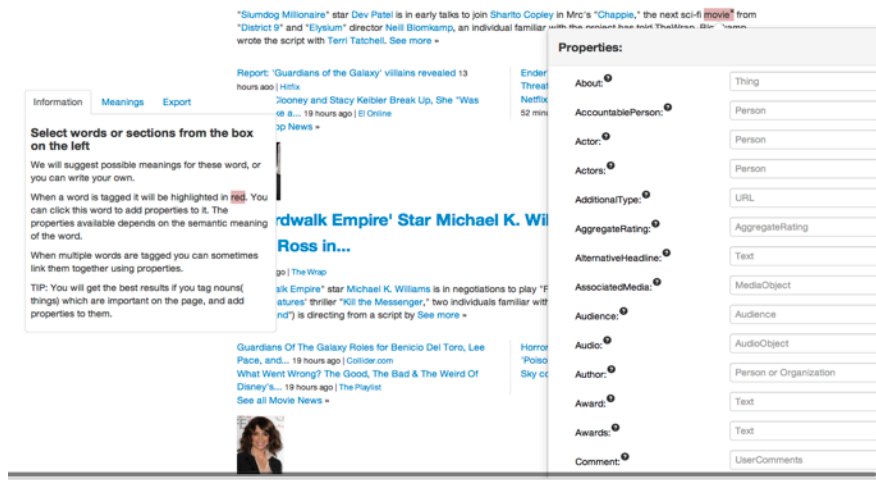
```
<span id="madame-movie-1" class="tagged"
typeof="sumo:MotionPicture schema:Movie wn:synset-movie-
noun-1" about="http://csaba.dyndns.ws:3000/load?
q=51dbfcbdc6765c3498000003#madame-movie-1" data-original-
title="">movie<span class="property" property="schema:about"
data-range="Thing" data-comment="The subject matter of the
content." href="Chappie"></span><span class="property"
property="schema:accountablePerson" data-range="Person" data-
comment="Specifies the Person that is legally accountable for
the CreativeWork." href="Sharlto COpley"></span><span
class="property" property="schema:actor" data-range="Person"
data-comment="A cast member of the movie, TV series, season,
or episode, or video." href="Dev Patel"></span><span
class="property" property="schema:comment" data-
range="UserComments" data-comment="Comments, typically from
users, on this CreativeWork." href="Not yet in production"></
span><span class="property" property="schema:director" data-
range="Person" data-comment="The director of the movie, TV
episode, or series." href="Neill Blomkamp"></span></span>
```

Fig 10. Drop down forms and the extensive schema.org markup they can generate

5. Related Work

WordNet is an extremely highly cited resource in all language related areas of study. The official web site at Princeton University maintains a list of publications[10] based

---

on WordNet, but this is no longer maintained because it was "growing faster than it was possible to maintain".

Within the Semantic Web community WordNet has enjoyed a duality with some researchers criticising its use as an ontology [13]-[15] while others embracing it either as a core taxonomy [16] or as a way to infer semantic relations (e.g. [17], [18]).

[10] used WordNet to automatically infer hierarchical classifications in textually annotated images, and [19] uses it to implement hierarchical faceted classification of recipes and medical journal titles. Both systems use automated extraction and disambiguation of key input terms, which differs from our approach where we ask users to supply these terms. But they use a very similar pruning algorithm to establish the final taxonomic structure.

The idea that free form user tags can be semantically enhanced has received a great deal of attention. Most of the existing work focuses on automatically enriching the tags already present, by exploiting the statistical regularities in the way tags are assigned to resources by users. [20] suggests that the efforts can broadly be classified as (a) extracting semantics of folksonomies by measuring relatedness, clustering, and inferring subsumption relations or (b) semantically enriching folksonomies by linking tags with professional vocabularies and ontologies, for example Wikipedia, and WordNet [21]-[23]. These resources are used in various ways, including to effectively cluster tags, for disambiguation, adding synonyms, and linking to annotated resources and ontology concepts. During this process the terms of the folksonomy are cleaned up and disambiguated, linked to formal definitions and given properties which make them more useful as ontologies.

There are also a few studies in which users are expected to contribute semantics at the time of tagging. [24] studies a corporate blogging platform which included a tagging interface. The tagging interface was linked to a domain ontology, and whenever someone typed a tag that had interpretations in the ontology the interface would present a choice of possible concepts to link the tag to. The ontology would also evolve as users typed new tags which were initially not in the ontology, but the scope of defined tags was limited by the ontology. [25] discuss a sophisticated Firefox plugin, Semdrops, which allowed users to annotate web resources with a complex set of tags including category, property, and attribute tags. These were aggregated in a semantic wiki of the user's choosing. [26] reports on an open source bookmarking application (SemanticScuttle) that has been enhanced with *structurable* tags, which are tags that users can enhance with inclusion and equivalence relations at the time of tagging. [27] describes extreme tagging in which users can tag other tags, to provide disambiguation and other relational information about tags.

These latter approaches require users to learn new ways of tagging, which are often more complex and opaque than free form tags. The benefit of LexiTagging is that the process is minimally different from activities they are already comfortable with. They simply sign up to a bookmarking site, install a bookmarklet and start tagging. The only addition to the workflow is to disambiguate tags, but this process is so similar to looking up definitions in a dictionary that it needs no explanation.

6 Conclusion

The LexiTags platform is a familiar bookmarking platform, like delicious.com, where users can store the URLs of interesting web sites and tag them with meaningful terms that aid in successive recall and discovery. The only modification is that the user tags are simple dictionary words, not disambiguated strings. But this small change gives the resources on the platform a sound semantic grounding which can significantly enhance the functionality of the service. Some examples of benefits to users are automatic content classification and browsing, external content recommendation, enhanced content discovery, and user profile matching. Some of these services represent crowdsourcing solutions to

existing problems which are difficult to fully automate. For example we have shown how the lexitags can be used to infer schema.org and SUMO classifications for each bookmarked web site, which is a task that would otherwise be done manually.

The vision is to create an integrated platform where users begin by simply bookmarking web sites, but then automatically receive the benefits of the enhanced services already described. This gives them the incentive to invest in the added effort to disambiguate their tags. Many of the components are in place, but they need some programming effort to complete the integration. This paper presented the theoretical motivation behind the work, and some preliminary results to show what is possible.

7   References

[1] C. Shirky, "Ontology is Overrated--Categories, Links, and Tags," *http:// www.shirky.com/writings/ontology_overrated.html*, 2007.

[2] A. Mathes, "Folksonomies-cooperative classification and communication through shared metadata," *Computer Mediated Communication*, 2004.

[3] A. Sheth and K. Thirunarayan, *Semantics-empowered Data, Services, and Sensor and Social Webs*. Morgan & Claypool Publishers, 2012.

[4] H. Hedden, "How SEMANTIC TAGGING Increases Findability," *EContent magazine*, 08-Oct-2008.

[5] C. Veres, "LexiTags: An Interlingua for the Social Semantic Web," presented at the Alexandre Passant,Sergio Fernández,John Breslin,Uldis Bojārs, (Eds.) Proceedings of the 4th International Workshop on Social Data on the WebIn conjunction with the International Semantic Web Conference (ISWC2011), Bonn, 2011.

[6] C. Fellbaum, *WordNet: An electronic lexical database*. Cambridge, MA.: MIT Press, 1998.

[7] C. Veres, K. Johansen, and A. Opdahl, "SynsetTagger: A Tool for Generating Ontologies from Semantic Tags," presented at the Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics., 2013.

[8] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, Nov. 1995.

[9] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and G. A. Miller, "Introduction to wordnet: an on-line lexical database," *CSL Report*. International journal of lexicography, 1993.

[10] E. Stoica and M. A. Hearst, *Nearly-automated metadata hierarchy creation*. Association for Computational Linguistics, 2004, pp. 117–120.

[11] P. Mika and T. Potter, "Metadata Statistics for a Large Web Corpus," *LDOW2012, April 16, 2012, Lyon, France*, 16-Apr-2012. [Online]. Available: http:// events.linkeddata.org/ldow2012/papers/ldow2012-inv-paper-1.pdf. [Accessed: 11-Jul-2012].

[12] I. Niles and A. Pease, "Towards a Standard Upper Ontology," presented at the Proceedings of the international conference on Formal Ontology in Information Systems - FOIS '01, New York, New York, USA, 2001, vol. 2001, pp. 2–9.

[13] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari, "Restructuring wordnet's top-level," *AI Magazine*, 2002.

[14] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari, "Sweetening WORDNET with DOLCE," *AI Magazine*, vol. 24, no. 3, p. 13, Sep. 2003.

[15] A. Oltramari, A. Gangemi, and E. al, "Restructuring WordNet's top-level: The OntoClean approach," *LREC2002*, 2002.

[16] F. Suchanek, G. Kasneci, and G. Weikum, "Yago: A Large Ontology from Wikipedia and Wordnet," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, pp. 203–217, 2008.

[17] T. H. Duong, T. N. Ngoc, and G. S. Jo, "A Method for Integration of WordNet-Based Ontologies Using Distance Measures," *KNOWLEDGE-BASED INTELLIGENT INFORMATION AND ENGINEERING SYSTEMS Lecture Notes in Computer Science, 2008, Volume 5177/2008*, Mar. 2008.

[18] J. Kietz and A. Maedche, "A method for semi-automatic ontology acquisition from a corporate intranet," *Workshop "Ontologies and text*, 2000.

[19] E. Stoica and M. Hearst, "Demonstration: Using wordnet to build hierarchical facet categories," presented at the ACM SIGIR Workshop on Faceted Search (August 2006)

[20] F. Limpens, F. Gandon, and M. Buffa, "Linking Folksonomies and Ontologies for Supporting Knowledge Sharing," *Projet ISICIL :Intégration Sémantique de l"Information par des Communautés d"Intelligence en LigneAppel ANR CONTINT 2008 ANR-08-CORD-011-05*, 01-Aug-2009. [Online]. Available: http://isicil.inria.fr/v2/res/docs/livrables/ISICIL-ANR-EA01-FolksonomiesOntologies-0906.pdf. [Accessed: 15-Aug-2011].

[21] L. Specia, "Integrating folksonomies with the semantic web," *The semantic web: research and applications*, 2007.

[22] Angeletou, Sofia; Sabou, Marta; Specia, Lucia and Motta, Enrico (2007). Bridging the gap between folksonomies and the semantic web: an experience report. In: The 4th European Semantic Web Conference 2007 (ESWC 2007), 3-7 Jun 2007, Innsbruck, Austria.

[23] C. Van Damme, M. Hepp , K. Siorpaes "Folksontology: An integrated approach for turning folksonomies into ontologies," *In ESWC workshop. Bridging the Gap between Semantic Web and Web 2.0 (2007)*, 2007.

[24] A. Passant, "Using ontologies to strengthen folksonomies and enrich information retrieval in weblogs," presented at the Proceedings of International Conference on Weblogs, 2007.

[25] D. Torres, A. Diaz, H. Skaf-Molli, and P. Molli, "Semdrops: A Social Semantic Tagging Approach for Emerging Semantic Data," *IEEE/WIC/ACM International Conference on Web Intelligence (WI 2011)*, Aug. 2011.

[26] B. Huynh-Kim Bang, E. Dané, and M. Grandbastein, "Merging semantic and participative approaches for organising teachers' documents," presented at the In J. Luca & E. Weippl (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008 (pp. 4959-4966). Chesapeake, VA: AACE, Vienna, 2008.

[27] V. Tanasescu and O. Streibel, "Extreme tagging: Emergent semantics through the tagging of tags," *Proceedings of the First International Workshop on Emergent Semantics and Ontology Evolution, ESOE 2007, co-located with ISWC 2007 + ASWC 2007, Busan, Korea, November 12th, 2007*, 2007.