Maphub Semantic Tagging Experiment Narrative Report

Executive Summary

Tags assigned by users to shared content can be ambiguous. As a possible solution, we propose **semantic tagging** as a collaborative process in which a user selects and associates Web resources drawn from a knowledge context. We applied this general technique in the specific context of online historical maps and allow users to annotate and tag them. To study the effects of semantic tagging on tag production, the types and categories of obtained tags, and user task load, we conducted an in-lab within-subject experiment with 24 participants who annotated and tagged two distinct maps. We found that semantic tagging implementation does not affect these parameters while providing tagging relationships to well-defined concept definitions. Compared to label-based tagging, our technique also gathers positive and negative tagging relationships. We believe that our findings carry implications for designers who want to adopt semantic tagging in other contexts and systems on the Web.

Use Case Context & Description

Maphub is a Web-Application that allows end-users to annotate digitized high-resolution historic maps. It has been bootstrapped with approximately 6000 historic maps from the Library of Congress Map Division (http://www.loc.gov/rr/geogmap/). The focus of this experiment was on the following use case:

- **UC3: Semantic Tagging**: while the user is creating textual annotations on a map or map region, Maphub automatically proposes resources from the Linked Data Web (e.g., DBpedia), which may be semantically related to the annotation and therefore also to the annotated map. Users can accept or reject link proposals.

The goal of this experiment was to study the effects of Maphub's semantic tagging feature compared to label-based tagging. The underlying motivation can be summarized as follows:

Despite their wide-spread adoption, tagging systems still face a number of problems: a tag can be ambiguous and have many related meanings (polysemy), multiple tags can have the same meaning (synonymy), or the semantics of a tag might range from very specific to very general because people describe resources along a continuum of specificity. These issues are rooted in label-based nature of tags and important for system providers who want to exploit the semantics and contextual information associated with tags for resource discovery. If, for instance, a user tags a resource with "Paris" it is not entirely clear whether this tag means "Paris", the capital of France or "Paris", the city in the United States. Contextual information, such as the translations of the term "Paris" in other world-languages or its geographical location can only be determined after reconciling label-based tags with entries in other data sources.

To solve this problem, we propose that users associate URI-identified Web resources from a knowledge context, such as Wikipedia, as part of their tagging activity. A tagging system could suggest the label "Paris" as a possible tag on the user-interface, but create a link to a Web resource (e.g., http://en.wikipedia.org/wiki/Paris) in the backend. We call this technique **semantic tagging**. Different from label-based tagging, the semantics of a tag is determined by its creator at creation time. Each tag also leads to further contextual information that can be exploited for resource discovery purposes. Explicit user feedback on suggested tags results in a graph of positive and negative tagging relationships that can be used to improve tag recommendation strategies.

We ran an empirical evaluation to compare semantic tagging with other tagging techniques.

Description of Annotation Classes

In the conceptual model for label-based tagging systems introduced by Marlow et al. (2006), which is shown in Figure 1, a user *u* assigns a tag *t* to a resource *r*. Tags are represented as labeled edges that connect users and resources but do not carry or refer to any additional contextual information. Both resources and users may be connected to other nodes, because there may be links between Web pages and users may belong to social networks. Label based tagging systems can allow for multiplicity of tags around resources (bag-model) or deny tag repetitions (set-model).

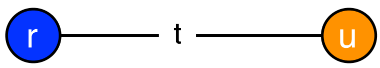

Figure 1. Label-based tagging model.

Semantic tagging, which is shown in Figure 2, extends this model by representing a tag t as qualified relationship between two resources: *rx* is the resource identifying and defining the semantics of a tag (e.g., http://en.wikipedia. org/wiki/Paris), and *ry* is the resource being tagged (e.g., a photo taken in Paris). The former is defined within a knowledge context K and can carry textual labels (e.g., "Paris") and additional context information (e.g., Paris is a city in France). Possible knowledge contexts are online encyclopedias such as Wikipedia, place name registries such as GeoNames, structured Web data sources such as Freebase (Bollacker et al., 2008), domain-specific Web vocabularies or gazetteers (e.g, Mostern and Johnson, 2008), or any other Linked Data source (Bizer et al., 2009) providing suitable concept definitions. An explicit, qualified semantic tagging relationship also implies an *about* relationship between the involved resources, meaning that *rx* is about *ry* if they are connected by a user via a semantic tagging relationship (Haslhofer et al., 2012).

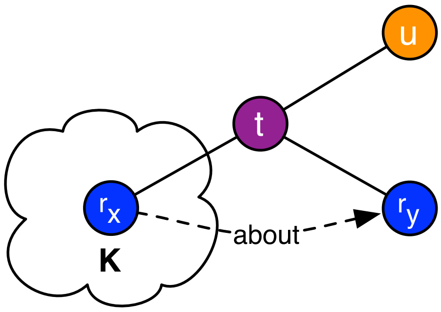

Figure 2. Semantic tagging model.

Since semantic tags can also be represented as first-class URI-identified Web resources, the resulting model is not label- or set-based but graph-based, with different types of nodes (users, resources) being connected to each other. This enables multiplicity and aggregation of tags not only around resources but also around users and user groups, which can be exploited for graph-based tag recommendation and user-based collaborative tag filtering (Jäschke et al., 2007).

Example/Illustrative Annotations

Figure 3 shows an example commentarial map annotation created in Maphub, which includes two semantic tags added by the user. The full set of annotations, including all semantic tags is available at: http://maphub-experiment.herokuapp.com

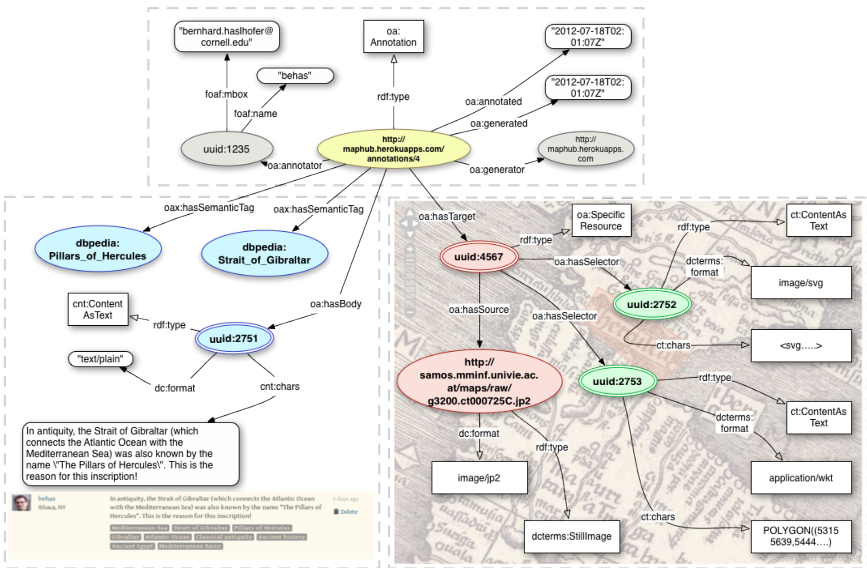

Figure 3. Commentarial Annotation

Summary of Obstacles

We did not face any majpor technical obstacles w.r.t the representation of Semantic Tags using the Open Annotation Model. However, some of the concerns we expressed in our previous report (model verbosity, missing direct links to annotated media object, label-based bodies, etc.) are still present in the current specification.

Technical Lessons Learned

We collected data from 24 participants, each of who created 4 annotations under 4 different tagging conditions. This gives us 96 annotations in total, each carrying zero or more accepted and/or rejected (semantic) tags.

Effect on tag production

Our results show that users added roughly the same number of tags under semantic tagging conditions as under the label-based tagging condition LT. It also shows that suggestive semantic tagging techniques lead to less rejected tags than suggestive tagging based on already added tags. Showing Wikipedia article abstracts as contextual information lead to a higher mean number of accepted tags but did not make a significant difference compared to label-based tagging.

Our results also indicate that users confronted with non-semantic but suggestive tagging techniques added fewer tags at the beginning of the experiment and less than in other conditions throughout the experiment. The evolution of tags added through label-based tagging is roughly linear because this technique does not rely on any other contextual information.

Effect on tag types and categories

An analysis of manually coded tags revealed that from all tags we collected in our experiment 48% were factual and 52% personal for the context of a given map. The distribution of factual and personal tags did not differ by condition.

The distribution of tag categories grouped by condition reveals that *locations* were tagged most frequently, followed by a large fraction of tags classified as *other*. Only some tags (7%) referring to *persons or groups*, hardly any (2%) *event* and only one *temporal* tag has been added. From this we can conclude that users mainly contribute location-specific contextual information to maps when adding tags. As with tag types, the distributions of assigned tag categories did not differ by condition.

Effect on user task load

In a post-test survey we asked participants to (i) express their perceived task load in a NASA TLX questionnaire and (ii) to rank the conditions they had seen according to intuitiveness, influence on annotation text, mental effort, and overall usefulness. Since we administered the TLX rating in an unpaired variant, each factor was rated individually and all factors were rated at once immediately after a participant was confronted with a condition.

From the obtained results we can conclude that (i) the additional effort caused by semantic tagging techniques has no effect on user task load, and (ii) there is no significant difference between semantic tagging with and without provided contextual information with respect to user task load.

Our participants found the semantic tagging user interface showing contextual information most useful and intuitive compared to the other techniques.

Generalizable Results and Conclusions

Using this application, we ran an empirical evaluation to compare semantic tagging with other tagging techniques. Our central findings can be summarized as follows:

- Our semantic tagging implementation does not affect tag production, the types and categories of obtained tags, and user task load while providing tagging relationships to well-defined concept definitions.
- When compared to label-based tagging, our technique also gathers positive and negative tagging relation- ships, which can be useful for improving tag recommendation and resource retrieval.

Even though we applied the semantic tagging technique in the context of historical maps, we believe that our findings carry implications for designers who want to adopt these techniques in other contexts and systems on the Web.

Overall, we believe that our findings carry implications for designers who want to adopt semantic tagging in other scenarios. A major incentive for system providers to implement tagging is to obtain metadata describing the content and context of online resources, which is important for efficient resource discovery but expensive in terms of time and effort when created manually. In traditional, label-based tagging systems providers can add possibly ambiguous label- based tags to their records. With semantic tagging, they obtain references to concepts defined in other Web-based knowledge context. Traditional information retrieval techniques can be enhanced to exploit these relationships and consider additional contextual information.

References

Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked data-the story so far." *International Journal on Semantic Web and Information Systems (IJSWIS)*5.3 (2009): 1-22.

Bollacker, Kurt, et al. "Freebase: a collaboratively created graph database for structuring human knowledge." *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM, 2008.

Haslhofer, Bernhard, et al. "Open annotations on multimedia Web resources. "*Multimedia Tools and Applications* (2012): 1-21.

Jäschke, Robert, et al. "Tag recommendations in folksonomies." *Knowledge Discovery in Databases: PKDD 2007* (2007): 506-514.

Marlow, Cameron, et al. "HT06, tagging paper, taxonomy, Flickr, academic article, to read." *Proceedings of the seventeenth conference on Hypertext and hypermedia*. ACM, 2006.

Mostern, Ruth, and Ian Johnson. "From named place to naming event: creating gazetteers for history." *International Journal of Geographical Information Science* 22.10 (2008): 1091-1108.