Collecting and Sharing Observations with Semantic Support

Sari Vainikainen, Antti Nummiaho, Asta Bäck, Timo Laakko

VTT, POBox 1000, FI-02044 VTT, Espoo, Finland

sari.vainikainen@vtt.fi, antti.nummiaho@vtt.fi, asta.back@vtt.fi, timo.laakko@vtt.fi

Abstract

We present two applications that can be used to store and share ideas, bookmarks and observations from the web and on the move. These applications utilize semantic web technologies both to support users in tagging and to store and integrate data. The core of the system is a social bookmarking application, Tilkut, complemented with a mobile application TagIt, which can be used to send photo and text entries from a mobile device. Tag suggestions are given from external ontologies, and from earlier tags. TagIt stores its data in an RDF database, which is also used in integrating these applications. The ontology for the RDF database combines existing social media ontologies, and its key structures are presented. The paper shares our experiences from linking and using external ontologies, and its special challenges on mobile applications.

Introduction

Social bookmarking introduced by del.icio.us was among the first popular Web 2.0 applications, and it has remained popular ever since. Social bookmarking relies on the use of tags - any user-given words that describe the bookmarks. Tags play a key role in linking resources and people to each others. Ease of creating tags has been an important factor in driving its popularity, but there are limitations, such as the flat structure, use of different words to describe the same thing, words with several different meanings (polysemy), experts using more specific terms than others, and different forms of the same word (singulars, plurals, typos). Also the context is important: tagging for personal use sets different requirements from tagging for wide findability and attention.

Semantics can be connected to social bookmarking at two different levels: tag level and system level. At system level, the data storage can be based on ontologies. At tag level, a tag becomes a semantic tag, when it is linked to a URI in a semantic knowledge base. Semantic knowledge bases are services like GeoNames[1], DBpedia[2] and Freebase[3], which offer semantic knowledge as ontologies, conceptual knowledge and structured knowledge.

We have utilized the KOKO[4] Ontology and GeoNames for creating semantic tags. The Finnish KOKO ontology is a large national collaborative ontology that includes an upper ontology (YSO) and domain ontologies like cultural, health related and geographical (Hyvönen & al 2008). GeoNames is a geographical database including names of places in various languages, place categories and coordinates in WGS84.

This paper presents our work with semantically enhanced social bookmarking. A web application, Tilkut, and a mobile application, TagIt, both with support for semantic tagging, and our social media ontology (OSMO) are presented. The ontology has been built by utilizing existing ontologies to interconnect data about users, their activities, networks and content from different applications. As the first test of its applicability, the ontology was used to exchange data between these two applications. We share the experiences of creating the semantic aspects and discuss the related challenges and opportunities.

Related Work

There are several online bookmarking services such as del.icio.us, Connotea and Clipmarks but only few offer semantic support for handling bookmarks. In Twine[5], users do not make semantic annotations, but the system aims at understanding the semantics of the bookmarked web pages and using this to support finding related items and people. ZigTag[6] lets the user select from tags with definitions. Faviki[7] lets its users use Wikipedia terms as tags. Terms are available in several languages.

Mobile location bookmarks are an example of mobile semantic annotations. A generic approach for contextual bookmarking comes from Henze et. al. (2007). They propose a system to link the physical and the digital worlds

[1] http://www.GeoNames.org/

[2] http://wiki.dbpedia.org/About

[3] http://www.freebase.com/

[4] http://www.yso.fi/

[5] http://www.twine.com/technology

[6] http://zigtag.com/

[7] http://www.faviki.com/

so that users can request additional digital information of real world objects. DBpedia mobile uses DBpedia dataset to demonstrate how the Geospatial Semantic Web can be explored using a mobile device (Becker & Bizer 2008). CityFlocks is a mobile system that lets visitors and new residents tap into the knowledge and experiences of local residents (Blinadzic & al. 2008).

Ames and Naaman (2007) describe a qualitative study of ZoneTag/Flickr users' tagging patterns and motivations for photo annotation. They propose a taxonomy of motivations for tagging along sociality and function. They showed that it is possible to motivate users to tag content at point-of-capture annotation. The user tests in (Bäck & al 2008) showed that people's attitudes towards sharing location tags vary a lot because of privacy issues. Location sharing was regarded the more useful and acceptable, the more familiar the community was. Ease of use is especially critical in mobile location tagging.

Tilkut: Semantic Social Bookmarking

Tilkut[8] is a social bookmarking and collaboration application. In addition to bookmarks, also notes and observations without a web link can be stored with tags.

Tilkut consists of a web application and bookmarklet for saving the link via a form. Some metadata is picked automatically. The user may change it and add more. Tags are entered into tag categories, which are topic, type, products, companies, places, project, person, importance, event and miscellaneous. Semantic tags are suggested from GeoNames to places and from KOKO to other tag categories (Fig. 1). When a user starts to write a tag, terms from the ontology vocabulary starting with the same characters are suggested. The user is free to discard the suggestions and write own tags. The user's previously used tags in the same tag category can be used easily.

Tagging with Tilkut was tested with users with our first prototype (Näkki & al 2008). The tests showed that users have very different needs. Some put high value on precise and systematic tagging, others value ease and speed. Based on these tests, the tag categories were slightly modified and groups were added. Consistent tagging in groups is supported by giving the user the opportunity to see which tags have been used in the groups and they can be used easily.

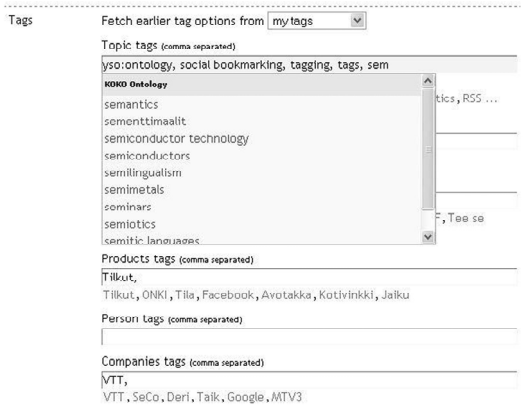

Figure 1. Adding tags in Tilkut into tag categories and with the help of semantic tag suggestions.

[8] http://owela.vtt.fi/tilkut/

TagIt Mobile Application

TagIt is a Java ME application that supports user-generated text and picture messages and uses ontologies to add semantics to messages (Fig. 2). It was developed independently from Tilkut. The ability to send clips to Tilkut was implemented to combine clips of real world with web clips. At this point, the support for sharing clips with other users and groups, and defining a clip public or private were added to TagIt so it has the same main features as Tilkut.

TagIt uses the same two semantic knowledge bases as Tilkut, i.e., GeoNames for adding location information (place tags) and KOKO for adding keywords that describe the content (topic tags). Also in TagIt, users own words can be used as tags. TagIt accesses GeoNames and KOKO directly. This was possible as they do not require the use of API keys that would be tied to a certain domain.

The application uses several methods for making the inputting of tags as easy as possible. The user's previous tags are listed for quick selection. When the user starts to type a new tag, the application accesses the semantic knowledge base in the background and retrieves suggestions starting with the same characters. After the search is complete, the suggestions are shown to the user. This way the user can often select the intended tag at some point instead of having to key it in entirely. When the place tags are added, the application also shows the country of the suggested location and lets the user check a map where the location is marked.

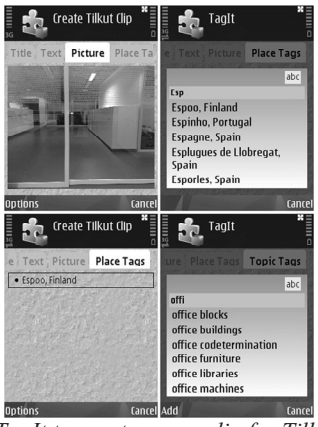

Figure 2.Using TagIt to create a new clip for Tilkut.

OSMO Ontology and Integration

Tilkut and TagIt were integrated by using an RDF database to store and transfer the common data (Fig. 3). The definition and use of the RDF database served not only to

integrate the applications but to evaluate the usability of existing social media ontologies as building blocks of a comprehensive social media ontology.

Our social metadata ontology (OSMO) supports describing social media content, activities, users, and sites. The starting point was to utilize existing ontologies as much as possible. We use SIOC[9] (Semantically Interlinked Online-communities) ontology for describing content, users and communication in social media services, and FOAF[10] (The Friend of a Friend) ontology together with the sioc:User and sioc:Usergroup classes from SIOC ontology for managing information relating to users and usergroups, their links and activities.

We combined Richard Newman's Tag ontology, and SKOS and MOAT ontologies for describing tags. Newman's Tag ontology[11] defines the key concepts like tags, resources and their relations. SKOS[12] (Simple Knowledge Organization System) is used for describing thesauruses and taxonomies semantically. In the tag ontologies, the Tag class is defined as a subclass of the SKOS's Concept, so the SKOS properties like broader, narrower and related can be used. The MOAT[13] (Meaning Of A Tag) ontology provides a way to describe different meanings for a tag. Meaning and Tag are core objects of the MOAT. The Tag object extends the Tag class of the Tag ontology. The Meaning object has exactly one meaning described with a URI which may refer to a semantic knowledge base (Passant & al 2008). Examples

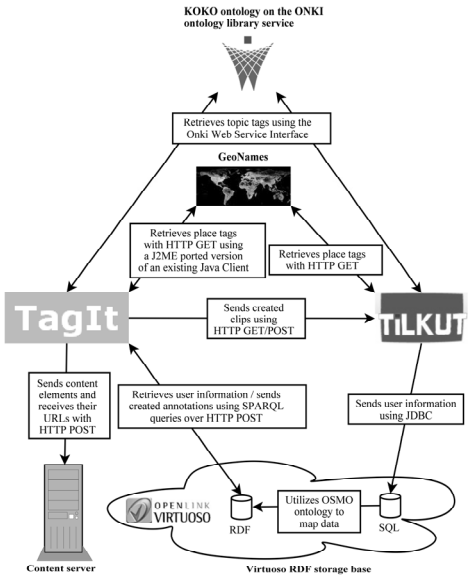

Figure 3. Architecture for the integrated Tilkut and TagIt applications. The open-source edition of Openlink Virtuoso database was used in the implementation.

[9] http://www.sioc-project.org/

[10] http://xmlns.com/foaf/spec/

[11] http://www.holygoat.co.uk/owl/redwood/0.1/tags/

[12] http://www.w3.org/2004/02/skos/

[13] http://moat-project.org/

of co-using SIOC, FOAF, SKOS and MOAT can be found in (Bojärs & al. 2008).

Tilkut supports hierarchical tag management; category, tag categories and tags. It's why we defined the Tag category and Category classes as subclasses of the skos:ConceptScheme. We use the MOAT ontology for describing tags, as well as the properties of tags:Tag and skos:Concept. This is possible, because the ontologies are interlinked. We need to store information of which user has used which tag, in which category and when. This information can be regarded as the context of tagging, and we use the tags:RestrictedTagging class to describe this. The information is used when the user browses bookmarks and takes different views on them with the help of the tag categories. Each user is able to select which tag categories she uses, the selected tag categories are linked to the user's bookmark category with the dcterm:hasPart property.

As explained earlier, users are not forced to use suggested tags, but they can add any words as tags. This means that there may be a tag with a clear meaning expressed with the help of the semantic web resources in KOKO and the same label without exact knowledge about the meaning. The meaning of a tag is expressed with help of the moat:Meaning class. Fig. 4 shows the relationships between the utilized classes.

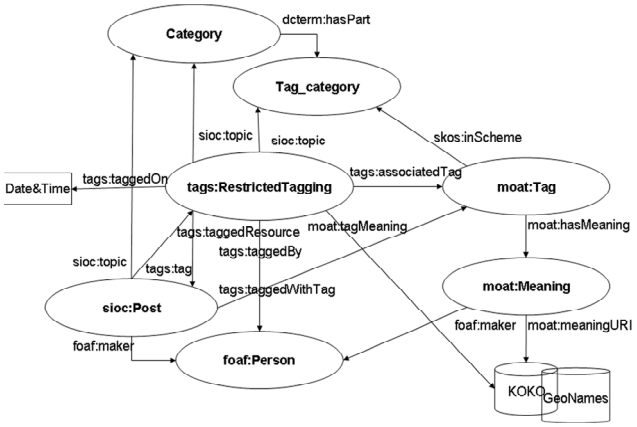

Figure 4. Interlinking the Tag, MOAT, SIOC and SKOS ontologies with the Category and Tag_category classes defined for the TILKUT application.

Lessons Learned

Using the KOKO ontology in a web application was easy because of the available ONKI selector user interface widget, but the exact meaning of the suggested term is not always clear, because the ontology is a collection of merged ontologies with some overlapping concepts. A more extensive web service API is under development, which should provide more information for term selection.

GeoNames location database may offer several suggestions with the same label, so information of the type of the place (e.g. city, lake, park) should be showed to help selecting the correct one.

There is no unified way to access semantic knowledge bases and an additional challenge is to access them from a mobile application. Based on our experiences, a relatively easy way to implement the communication from a mobile application is to use a web service interface, if there is one. The typical second choice is to implement the communication via HTTP requests and responses. This usually means more work, particularly if access is needed to multiple semantic knowledge bases not using the same protocol. For example, Freebase and DBpedia can both be accessed with HTTP GET requests, but the requests for Freebase must conform to MQL (Metaweb Query Language) while DBpedia uses SPARQL queries. The last option is to build one's own server that acts as a proxy between the mobile application and the semantic knowledge base.

Ajax (Asynchronous JavaScript And XML) techniques to show possible results after each input character can be applied also in mobile, as implemented in TagIt. The challenge is to be fast enough, where the critical factor is the response time of the semantic knowledge base. One way to make sure that the mobile application does not generate too many requests is to wait for a second or two after the user has input a character to see whether the user wants to input another character and only send the request after the user has stopped writing the tag.

Discussion

This paper presents different ways of adding semantics to tags and tagging data as a whole. The use of tag categories with a predefined meaning already adds some semantics. A more precise way is to use semantically defined tags. The whole tagging data gets a semantic structure when it is stored in RDF using our social media ontology.

From the user point of view, tags and tagging are not the aim, but means to an end, such as storing valuable ideas and resources, and sharing them with others. Our approach lets the user choose between freedom and structure. Freedom reduces the cumulative value of the stored data and sets additional requirements on utilizing it. For example, the use of categories to group tags helps in finding relevant resources, but if not all users use the category, some relevant resources cannot be found that way. The support for tag categories and tags with semantically defined meaning set special requirements on the ontology.

In Tilkut and TagIt we have shown how users can be supported to easily add semantic tags based on tag suggestions from semantic knowledge bases both in a web and mobile application. We presented also challenges of this approach from developers' and users' viewpoints. The APIs of different semantic knowledge bases vary complicating the development. Features like the possibility to search for concepts and to define how many results are wanted, would be very useful for mobile applications.

There are many semantic knowledge bases, and the challenge is to find the most suitable one. Even in different contexts, e.g., Tilkut in work and private use, different semantic databases might be relevant. It would be good to be able to choose and switch between ontologies. This sets additional challenges for utilizing semantic tags, like how concepts in different ontologies relate to each others.

Automatic methods of text and data mining combined with the available extensive semantic knowledge bases can already be utilized for creating tags. They should be used where possible to make relevant suggestions for the user to accept. The user's main task should be to create the tags that tell about the personal or group relevance.

In the future, we intend to focus at using the OSMO ontology and created annotations to add more intelligence. There are opportunities relating to creating and utilizing profiles, creating aggregations and making recommendations both to single users and groups.

References

Ames M. and Naaman M. 2007. *Why We Tag: Motivations for Annotation in Mobile and Online Media*. CHI 2007. April 28-May 3, 2007, San Jose, CA, USA.

Becker C. and Bizer, C. (2008) DBpedia Mobile: *A Location Enabled Linked Data Browser*. LDOW2008, April 22, 2008, Beijing, China.

Bilandzic M., Foth, M. and De Luca A. (2008) *CityFlocks: Designing Social Navigation for Urban Mobile Information Systems*. DIS '08, February 25 – 27, 2008, Cape Town, South Africa.

Bojãrs, U., Passant, A., Cyganiak, R., Breslin, J. (2008). *Weaving SIOC into the Web of Linked Data*. LDOW2008, April 22, 2008, Beijing, China.

Bäck, A.; Melin, M., Näkki, P., Vainikainen, S. Sarvas, R.; Seppälä, L.; Vihavainen, S. *Tags and tagging. Creating meanings, organizing, and socializing with metadata*. VTT, Espoo. 86 p. + app. 4 p. VTT Tiedotteita - Research Notes : 2449 ISBN 978-951-38-7238-0

Henze N., Lim M., Lorenz A., Mueller M., Righetti X., Rukzio E., Zimmermann A., Magnenat-Thalmann N. , Boll S, and Thalmann S. (2007) *Contextual Bookmarks. In: Pr. of the Joint Workshop Mobile Interaction with the Real World* (MIRW 2007), 2007, p. 51-54. Singapore, 2007.

Hyvönen E., Viljanen K., Tuominen J., Seppälä K., Kauppinen T., Frosterus M., Sinkkilä R., Kurki J., Alm O., Mäkelä E. and Laitio, J. (2008) *National Ontology Infrastructure Service ONKI*. Oct 1, 2008.

Näkki, P., Vainikainen, S., Bäck A. (2008) *Experiences of Semantic Tagging with Tilkut*. MindTrek'08, October 69, 2008, Tampere, Finland.

Passant A., Laublet P., (2008) Meaning Of A Tag: *A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data*. LDOW2008, April 22, 2008, Beijing, China.