Combining Tags and the Semantic Web for Linked Tagging Data

Haklae Kim[1], Sungkwon Yang[2], Jiwoong Jung[3], Kwangsub Kim[4], John G. Breslin[1], Stefan Decker[1], and Hong Gee Kim[2]

[1] Digital Enterprise Research Institute, National University of Ireland
IDA Business Park,Lower Dangan, Galway, Ireland
{first name.last name}@deri.org
[2] Biomedical Knowledge Engineering Lab, Seoul National University
28-22 Yeonkun-Dong, Chongno-Ku, Seoul 110-749, Korea
{hgkim, sungkwon.yang}@snu.ac.kr
[3] NCSoft Openmaru Studio
602-245 Namhyun-dong, Kwanak-gu, Seoul 151-801, Korea
open80@gmail.com
[4] NHN Corp
Bundang First Tower, 266-1 Seohyeon-dong, Bundang-gu, Seongnam
Gyeonggi 463-824, Korea
kwangsub.kim@gmail.com

Abstract. We describe an open tagging platform that aims to make tagging data open, more universal, and apply it across different tagging sites. We implement the web-based application - int.ere.st - to realize this goal. int.ere.st collects information related to tagging behaviors from Web 2.0 sites and offer direct access using Semantic Web technologies such as Linked Data. The application is available at http://int.ere.st/.

1 Introduction

Although social tagging and folksonomies have a lot of advantages (visualization, navigation, etc.) to offer users who tag content items in social media sites, critical drawbacks with current tagging systems are that 1) there is no formal conceptualization in order to represent tagging data in a consistent way [1], and 2) there is no interoperability support for exchanging tagging data among different applications or people [2]. The simplicity and ease-of-use of tagging leads to a lack of precision with keyword ambiguity caused by misspelling certain words, singular vs. plural, synonyms, morphologies, or too-personalized tags [3–5]. There are many different manners of using tags, and one may not be able to understand what a given tag is about. These limitations come from a lack of standards for tag structures and little semantics for specifying the exact meaning. Aside from these problems, current tagging systems do not provide a uniform way to share and reuse tagging data amongst users or communities. There is no consistent method for reusing one's personal set of tags among people or communities. In individual perspective, users participate in diverse social media sites by contributing to tagging activities. Although they are able to collect the tagging data resulted from these

activities, the real challenge is to integrate and combine this data into a comprehensive personal view. On the side, users are part of different communities and projects, and interact with the members of these communities by sharing or exchanging tagging data between them. In this setting, a new issue arises, i.e. the reuse of the data across multiple communities.

Therefore, it is not easy to meaningfully search, compare, or merge similar collective tagging data from different applications [6]. With the usage of tagging systems increasing daily, these limitations will become critical. To overcome the limitations of current tagging systems, we need to look at an open platform for tagging similar to OpenSocial [5] that provides a common set of APIs for social networking applications across multiple web sites. The goal of int.ere.st realizes the open tagging platform.

2 Semantic Web technologies for social tagging

Semantic Web technologies in general allow us to expose human knowledge to machines in order to performing automatic data linking and data integration [1]. In particular, ontology as enabling technology for the Semantic Web enables knowledge exchange amongst different users and applications by providing reusable constructs. That is, these technologies can be improved current tagging environment, not only by representing semantically tagging data, but also by allowing users to share and exchange the data across sites, services, or applications.

As more users contribute to social software applications, they may want to be able to reuse their data across heterogeneous systems and to move their data into wherever they want to use. Tag ontologies could be a solution that aims to publish semantic tag metadata that makes it easy to interlink, discover and consume data on Semantic Web climate. One of the advantages of tag ontologies is that isolated tagging data can be easily made mobile and integrated across applications. Tags, user, and their relations in particular application can be represented in a form of ontology such as RDF or OWL, and these data can be be accessible and movable on the Web as linked data. This can be considered as a starting point of sharing, exchanging for separate tagging activities on different platforms.

We use SCOT (Social Semantic Cloud of Tags) ontology [6] to describe tagging information. This ontology can describe the structure and the semantics of tagging data and also offer social interoperability of the data among different sources [1].

3 int.ere.st Overview

The int.ere.st website aims at publishing the open semantic web database for tagging data, including a large number of interlinks to several data sets on tagging applications. Tagging data of individual users, communities, or corporates existed in distributed environments can be imported to int.ere.st via mash-up services. Then, the data is transformed into the specific format - SCOT - as a machine-processable way and the data is automatically updated when information from the host site is changed.

---

[5] http://code.google.com/apis/opensocial

[6] http://scot-project.org

Our database now contains information about 15,000 users and 270,000 items ranging from multimedia, social bookmark, research to blog posts. We expect that a number of resources will be increased daily via the RSS aggregator and the RDF crawler and be published by a form of triple.

4 int.ere.st as tagging platform

int.ere.st is the first open tagging platform for the Semantic Web that aims to make tagging data open, more universal, and apply it across any number of social tagging sites. In order to allowing users and developers to support social capabilities underlying tagging data, the platform consists of core components:

– open data formats: aims to specify tagging data in a machine-processable way
– methods: import, share, and search functions
– open API: allows external users to direct access

As illustrated in Figure 1, tagging information can be collected from various sites. Thus gathering these data is highly dependent on service providers' policies and the methods for doing this task is also varied. For collecting data, we implement two ways:

– RSS reader and crawler: automatically aggregates RSS feeds and extracts URIs
– mash-up importer: collects data of individual users who has an account of a particular services such as Flickr, Delicious, YouTube

All tagging data from Web 2.0 sites is transformed and published in SCOT and then these resources can be machine-processable. The semantic metadata can be shared via several ways such as RDF vocabularies and open APIs. The open API is to allow users to get a specific information from the sites. At present, we offer very simple functions such as get my tag cloud, search this tag, and add tag set, etc.

5 Linked Tagging Data

All information in int.ere.st is published as linked data using the D2R Server that a tool for mapping relational databases to RDF accessible through SPARQL. In particular, the linking of tagging entities can be utilized with SCOT and queries for a particular information can be made using SPARQL. One important issue is that tagging information by nature has aggregative and collective features. Thus all information has to updated automatically when an event such as add tags or resources is happened in a host site. As we have a number of sources, it is difficult to synchronize metadata with raw sources. By using D2R we avoid this issue because metadata in our system does not need to synchronize and expose as new one.

int.ere.st implements a HTTP content negotiation mechanism, users can get either HTML or RDF representation of a particular data and also the SPARQL interface (http://dev.snu.ac.kr:8080/snorql/) allows users to query semantic tag metadata directly..
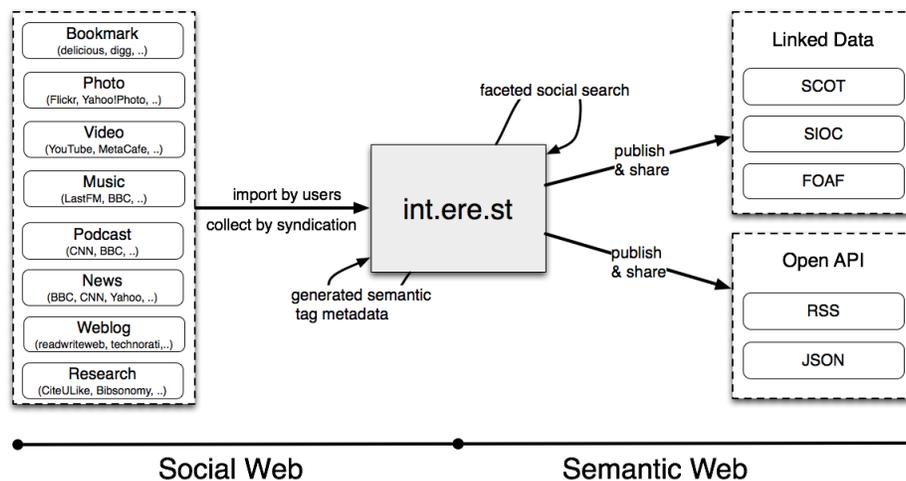
Fig. 1. Functional architecture of int.ere.st

6 Faceted social search

int.ere.st aggregates various resources from a variety of applications, services or sites. From visualization purpose, a tag cloud is often used as visual interfaces for navigating in tagging systems, a flat visualization however is not sufficient to provide effective and efficient information retrieval. Basically we also use the tag cloud as a default visualization for individual tagging data. However, in order to improving search capabilities we provide facet search interface that classifies a set of media, a set of sites. The major benefits of facet approach include "strong reduction of the mental work", and better support for exploration, discovery and iterative query refinement [7]. The choice of facets in int.ere.st is based on features of the popular user-generated web sites. Table 1 shows facets we provide at present.

Table 1. int.ere.st facets definition by resource types

Facets	Popular sites
News	BBC, CNN, Yahoo
Bookmark	Delicious, Digg, Simpy
Blog	ReadWriteWeb, TechCrunch, TechnologyReview
Podcast	CNN, BBC, TechnologyReview
Video	Yahoo, YouTube, Metacafe
Research	CiteULike, Bibsonomy
Photo	smugmug, zoto, vi.sualize.us
Music	LastFm, BBC, iTunes, Yahoo

We provide the public SPARQL end-point (http://dev.snu.ac.kr:8080/) which enables users to query the RDF resources with SPARQL queries. A query can be extended for different purposes to get data. For instance, most simple type of a query is based on combination of tags such as co-occurring tags, while int.ere.st allows users to make much difficult queries like "a user who uses a particular tag in delicious and flickr" or "users who uses a set of tags in YouTube". int.ere.st provides not only a tag based search, but also social search which aims to find out users from their tagging practices.

7 Summary

Using int.ere.st, users will have the opportunity to get a live, hands-on experience of combinations the Social Web and the Semantic Web. The users can manage a collection of their tagging data across different sources or applications and they can search resources by multidimensional facets. In addition, the tagging data is published in SCOT as Linked Data and we offer some functions as open APIs to encourage the share and exchange of tagging data. In a future we plan to interlink our data to Linked Data practices such as Dbpedia, Revyu.com etc.

Acknowledgments

References

1. Kim, H., Scerri, S., Breslin, J.G., Decker, S., Kim, H.: The state of the art in tag ontologies: A semantic model for tagging and folksonomies, Proceedings of the International Conference on Dublin Core and Metadata Applications (2008)
2. Gruber, T.: Collective knowledge systems: Where the social web meets the semantic web. Journal of Web Semantics **6** (2008) pp. 4–13
3. Golder, S.A., Huberman, B.A.: The structure of collaborative tagging systems. Journal of Information Science **32** (2006) pp. 198–208
4. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: WWW '07: Proceedings of the 16th international conference on World Wide Web, New York, NY, USA, ACM (2007) pp. 211–220
5. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Ht06, tagging paper, taxonomy, flickr, academic article, to read. In: HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia, New York, NY, USA, ACM (2006) pp. 31–40
6. TagCommons: Ontologies vs. formats vs. schema vs. apis (2007) available at: http://tagcommons.org/2007/03/02/ontologies-vs-formats-vs-schema-vs-apis/ (accessed 24 September 2008).
7. Rosati, L., Resmini, A., Quintarelli, E.: Facetag: Integrating bottom-up and top-down classification. In Tummarello, G., Bouquet, P., Signore, O., eds.: SWAP. Volume 201 of CEUR Workshop Proceedings., CEUR-WS.org (2006)