StYLiD: Structure Your Own Linked Data

Aman Shakya and **Hideaki Takeda**
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku,
Tokyo, Japan 101-8430
shakya_aman@nii.ac.jp, takeda@nii.ac.jp

Vilas Wuwongse
Asian Institute of Technology
Klong Luang, Pathumthani,
Thailand 12120
vw@cs.ait.ac.th

Abstract

Information sharing can be effective with structured data. The Semantic Web is mainly aimed at creating standard ontologies. However, users have different preferences and evolving requirements. It is not practical to attempt perfect schema definitions with strict constraints. Creating structured formats should be a collaborative and evolutionary process. Social software motivates wide participation by providing easy interface. We propose a system called StYLiD for sharing wide variety of structured information. Users freely define their own structured concepts. The system consolidates different versions defined by different users. Combined search and browsing can be done. Popular concepts gradually emerge from the concept cloud and stabilize. Concept definitions are flexible. The suggestive range for attribute values does not constrain the contributors. StYLiD generates unique dereferenceable URIs so that data items can form a linked data web. Structured data is embedded using RDFa.

Introduction

Information sharing on the Web has become a basic need in communities. People want to share a wide variety of information. Structuring helps in handling different types of data effectively. With structured data, we can define the machine understandable semantics so that processing can be automated. Search and browsing becomes more effective. Interoperability and integration of data from various sources is possible with standard or compatible formats.

Efforts for the Semantic Web have been mainly directed towards creating standard ontologies. However, there are not many ontologies to cover the wide variety of information. Even if ontologies exist, it is difficult to search for the appropriate one. Further, understanding and using it is not easy. It is not always feasible to build new systems or ontologies for emerging needs. It is not possible to accommodate all requirements and exceptions that surface in future. Users may need different types of data depending upon the purpose. Thus, creating structured formats should be a widely collaborative process. However, to have large scale collaboration, the system should be easy to use and understand for general users. On the other hand, social software has been successful in drawing huge user participation. Tagging is successful because it is very simple and does not impose any hard constraints. However, these systems do not provide much semantic structure and interoperability.

Recently, the combination of social software with Semantic Web technology is gaining significant attention. However, the informal approach of the social web allows inconsistencies and inaccuracies(Schaffert 2006). We need tolerant mechanisms to round them up. We propose a system called StYLiD[1] (an acronym for Structure Your own Linked Data) for collaborative and flexible structured data sharing.

Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

The StYLiD Platform

StYLiD allows users to create, share, reuse and refine structured concepts. Users input information freely. Similar concepts are consolidated later. It has been built upon the Pligg content management system using PHP and MySQL.

Use Case Scenario

Suppose a user wants to share some structured data, for e.g., announcement of a talk program. He may register on StYLiD and define his own "talk program" concept on the fly with a list of attributes, for e.g., date, venue, speaker, etc. Then he can start sharing data using this concept. Any other user can also contribute data. A cloud view of concepts allows browsing contents of different types. The user can navigate through linked data entries. He may search data by the concept name and a set of attribute name, value pairs. RDFa aware browsers would be able detect structured data, offer suitable operations and copy it to desktop applications.

Sharing Structured Concepts and Data

The users may freely define their own *concepts* by specifying the concept name, some description (optional) and a set of attributes. Each attribute has a name, description (optional) and a set of concepts as the suggested value range (optional). Any user may enter *instance* data. An attribute can have single or multiple values. Each value may be a literal or a resource URI. The system embeds structured data using RDFa format. It also outputs RDF separately.

The system allows different users to define concepts having the same name. The user can modify an existing concept to make own version instead of defining from scratch.

[1] http://dutar.ex.nii.ac.jp/stylid

Users can update their own concept definitions and add attributes. However, to alter attributes a new version of the concept should be defined to keep existing data intact. Thus, the same user can also have different versions of a concept.

The system offers every user a personal *concept collection*. Concepts created or adapted by the user are automatically added to this collection. Other useful concepts can also be added. Thus, the users need not be overwhelmed by the huge number of concepts.

Consolidation of User Defined Concepts

Concepts having the same name defined by different users are grouped together. This forms a single virtual concept which subsumes all the grouped concepts. This can be used to retrieve all instances though users have different definitions for the concept.

All the concepts contributed by different users are visualized together as a **concept cloud**. When the user hovers over any concept, the attributes and description are shown so that it can be understood at a glance. Better concept definitions will satisfy more users and will have more instances. Popularity of concepts is visually highlighted by increasing size. Stable definitions will gradually emerge from the vast cloud of concepts as more instances are contributed.

A consolidated concept can be expanded into a *sub-cloud* which shows all the versions. Further, in the sub-cloud, multiple versions defined by the same user are sub-grouped together. The sizes of the different versions in a sub-cloud add up to form the size of the consolidated concept. Clicking on a consolidated concept shows instances of all versions of the concept. Similarly, we can see all instances of the multiple versions defined by a user.

Flexible Definitions and Relaxed Data Entry

Creating perfect concept definitions with strict constraints is not easy and practical. It is difficult to think of all attributes and possible value ranges. While defining a concept A, if an attribute takes value of type B, concept B must already be defined. If concept B has an attribute taking values of type C, then concept C must be defined first, and so on. Similarly, it may be difficult to enter strict data as mandated by a schema. Resource URIs for attribute values may not exist or the user may not be able to find it at the time. The system tries to avoid such difficulties by allowing flexible and relaxed definitions. The concept definitions may be updated incrementally. The suggestive range of attribute values does not impose strict constraints. Rather it assists the user by suggesting values. The range may be updated by including more concepts or narrowing it down. The user may type in literal values for attributes. If the user knows a resource URI for the value, it may be entered to link to that resource. The resource may also be entered later and the entry can be edited to specify the link. Users may input single or multiple values for any attribute.

Open System for Creating Linked Data

The system generates unique dereferenceable URIs for each concept, attribute and instance. A concept is identified by its name, creator and the version number. Similarly, consolidated virtual concepts are also assigned URIs that can be uniquely referenced. An attribute is uniquely identified by the concept and the attribute name. An instance is uniquely identified by the system generated ID. The URI of a concept or an instance dereferences to a page describing its details in both human and machine readable forms.

Data instances can be linked to each other by entering the resource URIs as attribute values. This is manifested as simple hyperlinked entries for the user while creating a linked data web. We can also link to URIs from other systems and others can link in using the unique dereferenceable URIs. StYLiD is an open system that does not lock data into itself. The structured information snippets in RDFa may also be posted elsewhere on the Web. The system also provides a SPARQL query interface for open external access.

Related Work

Freebase(http://www.freebase.com/) allows users to define their own structured types. However, it keeps the types and domains defined by different users separate. It is difficult to contribute types and instances because of strict constraints. We cannot link to external data to form a linked data web. The myOntology project(Siorpaes & Hepp 2007) is based on wikis for community-driven lightweight ontology building. However, it does not demonstrate sharing of structured data. Freebase, myOntology and semantic wikis are all based on wiki which assumes everyone to settle with a single version for each resource. StYLiD offers the flexibility to maintain own preferences. Revyu(http://revyu.com/) allows people to review anything using unique dereferenceable URIs. However, the structure of different things is not modeled.

Conclusions and Ongoing Work

We proposed StYLiD as a single platform for sharing a wide variety of structured data. Users can define their own concepts. Relaxing constraints would encourage free user contribution to better meet their requirements. The task of consolidating several user defined concepts is handled by the system facilitating the emergence of popular and stable generalized definitions. The open system supports creation of a linked data web. Thus, using informal social software it produces formal machine understandable structured data.

Currently, concepts are only consolidated by names and viewed uniformly through searching and browsing. Work is ongoing for more intelligent consolidation of similar concepts and mapping the attributes of consolidated concepts.

References

Schaffert, S. 2006. Semantic social software: Semantically enabled social software or socially enabled semantic web? In *Proceedings of the SEMANTICS 2006 conference*, 99–112. Vienna, Austria: OCG.

Siorpaes, K., and Hepp, M. 2007. myOntology: The marriage of ontology engineering and collective intelligence. In *Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007)*, 127–138.