Community-Driven Linked Data Authoring and Production of Consolidated Linked Data

Aman Shakya, National Institute of Informatics, Japan
Hideaki Takeda, National Institute of Informatics and the University of Tokyo, Japan
Vilas Wuwongse, Asian Institute of Technology, Thailand

ABSTRACT

User-generated content can help the growth of linked data. However, we lack interfaces enabling ordinary people to author linked data. Secondly, people have multiple perspectives on the same concept and different contexts. Thirdly, not enough ontologies exist to model various data. Therefore, we propose an approach to enable people to share various data through an easy-to-use social platform. Users define their own concepts and multiple conceptualizations are allowed. These are consolidated using semi-automatic schema alignment techniques supported by the community. Further, concepts are grouped semi-automatically by similarity. As a result of consolidation and grouping, informal lightweight ontologies emerge gradually. We have implemented a social software, called StYLiD, to realize our approach. It can serve as a platform motivating people to bookmark and share different things. It may also drive vertical portals for specific communities with integrated data from multiple sources. Experimental observations support the validity of our approach.

Keywords: Concept Consolidation, Data Integration, Emergent Semantics, Freebase, Structured Data, Linked Data, Social Semantic Web, StYLiD,

INTRODUCTION

Linked data is a method of exposing, sharing and connecting data on the Semantic Web. It provides the mechanisms for publishing and interlinking structured data into a Web of Data. This forms a data commons where people and organizations can post and consume data about anything. Due to the network effect, usefulness of data increases the more it is linked with other data. Organizations benefit by being in this global data network, accessible to both people and machines. Linked data can be fully realized with existing technologies maintaining compatibility with legacy applications while exposing data from them. Thus, linked data is a significant practical movement toward the vision of the Semantic Web (Berners-Lee, 2006; Bizer, Cyganiak, & Heath, 2007).

However, some issues still remain which need to be addressed for wider adoption of linked data. Firstly, it is not obvious how ordinary
people, without any technical expertise, can publish and share linked data directly. Linked data research can benefit from the combination of Semantic Web and social Web techniques. A lot of data on the Web comes from the people. However, there is a lack of human interfaces to publish linked data explicitly. People still share unstructured data, and it is hard to derive semantic structure and links from such contents.

Secondly, the fact that there may be multiple perspectives on the same concept, different aspects or contexts to be considered, is often ignored. In the distributed web, different parties may have different schemas or conceptualizations for the same type of data because of different requirements, data formats or preferences. Thus, organizations usually need to integrate their data at the schema level. However, today data is mainly being linked at the instance level only (Jhingran, 2008) though knowledge of schema is very important for information exchange and integration between systems and querying data sources. Therefore, we should also link data at the schema level to explicitly encode the knowledge of relations among multiple conceptualizations. Currently, it is not obvious how to link or relate such multiple concept schemas in the linked data web.

Lastly, the state of the art lacks structures that can represent and organize the wide range of concepts needed by the community. There are still not enough ontologies or vocabularies for describing linked data about various things (Siorpaes & Hepp, 2007a; Van Damme, Hepp, & Siorpaes, 2007). There is a long tail of information domains for which people have information to share (Huynh, Karger, & Miller, 2007). Developing individual solutions for the long tail is infeasible because data modeling is difficult. It is not always practical for different parties to commit to a single data model or common vocabulary. It may be possible to achieve some level of consensus but the process of collaborative interaction with common understanding is itself difficult and time consuming.

Considering the above issues we propose the following as our main contributions.

- **Social linked data authoring:** We attempt to enable ordinary users to publish structured linked data directly through simple authoring interfaces. We have implemented a linked data authoring social software for sharing a wide variety of data in the community.
- **Multiple conceptualizations:** Users may freely define their own concept schemas and share different types of structured linked data. We propose allowing different people to have multiple conceptualizations.
- **Concept consolidation:** At the same time, these multiple concept schemas are consolidated by mapping and linking them at the schema level. This is done semi-automatically, supported by the community, using data integration principles with schema alignment techniques. We propose concept consolidation as a new way of building up conceptualizations from the community. This is a loose collaborative approach requiring minimum understanding and allowing different parties to maintain individual requirements.
- **Emergence of lightweight ontologies:** Besides community-based formation of conceptualizations by consolidation, concepts can evolve and gradually emerge out by popularity. Further, similar concept schemas can be grouped and organized semi-automatically. Together these processes enable emergence of informal lightweight ontologies.

The rest of the article is structured as follows. We will start with an overview of our approach and the architecture in the next section. Next, we will discuss about user motivation and some possible application scenarios. In the following section, we will describe about social linked data authoring. Then we will describe about consolidated linked data. Next, we will discuss about the emergence of lightweight ontologies. Following this, we will show some experimental observations on an existing col-
Related Content

**AL-QuIn: An Onto-Relational Learning System for Semantic Web Mining**
[www.irma-international.org/article/quin-onto-relational-learning-system/62560/](www.irma-international.org/article/quin-onto-relational-learning-system/62560/)

**A Context-Based Approach for Supporting Knowledge Work with Semantic Portals**
[www.irma-international.org/article/context-based-approach-supporting-knowledge/2811/](www.irma-international.org/article/context-based-approach-supporting-knowledge/2811/)

**Improving Curated Web-Data Quality with Structured Harvesting and Assessment**

**Improving the Information Security of Collaborative Web Portals via Fine-Grained Role-Based Access Control**
[www.irma-international.org/chapter/improving-information-security-collaborative-web/39184/](www.irma-international.org/chapter/improving-information-security-collaborative-web/39184/)
Towards Process Mediation in Semantic Service Oriented Architecture (SSOA)
www.irma-international.org/chapter/towards-process-mediation-semantic-service/35757/