Chapter 3.22
Social Support for Ontological Mediation and Data Integration

Gianluca Correndo
University of Southampton, UK

Harith Alani
University of Southampton, UK

Manuel Salvadores
University of Southampton, UK

ABSTRACT

Most organisations store their data in several databases with no flexible mechanism for integration and access, and with no common vocabulary in place. Maintaining local vocabularies while realising distributed access is a challenge that most organisations face regularly. For several years, the Semantic Web community has been developing algorithms for mapping data models (ontologies). Nevertheless, ontology mapping remains to be a great challenge, and humans are always expected to verify the results of existing automatic mapping tools. The spread of social web demonstrate the possibility of using collaborative techniques for reaching consensus and fostering user participation. While a number of prototypes for collaborative ontology construction are being developed, collaborative ontology mapping is not yet well investigated. In this article, the authors describe an approach that combines off-the-shelf ontology mapping tools with social software techniques to enable users to collaborate on mapping ontologies.

INTRODUCTION

Most public and private organisations own a vast capital of information, stored in many distributed databases, using various departmental vocabularies, and are usually subject to frequent change and update. Realising a persistent integration and access to such data is usually a major challenge, especially when there is a need to integrate with external data sources, such as in coalition forces scenarios, or in government.

The UK, in its e-Government agenda, has identified data integration and sharing as crucial issues, and recommended the Semantic Web as an enabler technology (Alani et al. 2007). B2B is another example of cross organization scenario that would benefit from data integration. Traders

can in fact achieve a higher level of exposure of their businesses by sharing structured data and SW research initiatives are helping in delivering ontologies for defining offerings, services and products (Hepp, 2008).

In the academic domain, the Linked Data initiative (Bizer et al. 2007) has been very successful so far in guiding the exposure and sharing of data in semantic formats. Over 17 billions of triples are currently available as Linked Data, covering a wide range of domains, such as Wikipedia, FOAF files, census data, scientific publications, social websites data, and much more (Hausenblas et al. 2008). However, this growing community has little support for mapping the data it is publishing and sharing.

Enhancing information sharing and reuse requires providing communities with common spaces, tools and technologies for the dissemination and integration of structured information from different sources. There are many cases where organisations invested in building monolithic data models to subsume their entire information asset. However, such top-down models are very costly and time consuming to build. Furthermore, they are often hard to deploy and impose on data owners for various reasons, such as difference in coverage, inflexibility of the model, and its disproportionate scale (Alani et al., 2008).

The Semantic Web (SW) brings hope to easing the above problems by providing the technology to develop and share ontologies to represent data sources, and to map them together to facilitate integration and access, without the need to adhere to a common terminology. The Semantic Web offers hope to solve a number of problems that have been haunting the digital information world for many decades. Problems such as lack of shared understanding, differences in terminology, lack of machine understandable information, difficulty in integrating distributed information sources, high reuse costs, are all quite common problems.

Using the web as a common framework for data publishing helps in providing communities with a worldwide shared information space, and adopting standards from the SW (Berners-Lee, 2007) enabled the process of data exposure and linking. Ontologies provide a formal interpretation of data semantics that can be used for supporting information exploitation.

Defining an ontology or a knowledge base and mapping it properly to related resources is usually a costly and time consuming task. If ontologies are meant to reflect the views of a specific community over a specific domain and support their knowledge sharing tasks, then the community itself should be empowered to express, agree, bridge and formalise their definition by social means in order to support such tasks (Shadbolt et al. 2006).

There is currently an increasing interest in exploring new methods for constructing such knowledge in more social and collaborative ways, and several tools have recently been developed for this purpose (Noy, Chugh, and Alani, 2008). These tools tend to focus on supporting communities to create instance data or to build domain ontologies collaboratively. However, generic tools for collaboratively mapping these shared resources are not widely available.

In this article we describe OntoMediate, a system for supporting communities in sharing, mapping, and extending their ontologies and knowledge bases. More specifically, OntoMediate allows the following: **align** local ontologies to shared ones; **exploit** social interaction and collaboration to improve alignment quality; **reuse** user ontology alignment information for enhancing future automated alignments and **query** heterogeneous data sources.

In the following section we will briefly explain ontology mapping and existing automatic techniques for producing mapping results. In section 3 we will describe a number of tools for collaborative knowledge construction, highlighting their goals and main characteristics. Section 4 is dedicated to describing OntoMediate; our tool for community-driven ontology mapping.