Collaborative Ontology Mapping and Data Sharing

Gianluca Correndo, Harith Alani
University of Southampton,
Electronic and Computer Science Department
{gc3, ha}@ecs.soton.ac.uk,
SO17 1BJ, United Kingdom

Abstract

The Semantic Web aims to create a web of data where contents can be easily discovered and integrated using metadata. Many ontologies have been proposed over the years in different domains, thus producing a semantic heterogeneity that is difficult to manage. Various automated ontology mapping techniques and tools have been developed to facilitate the bridging and integration of distributed data repositories. Nevertheless, such tools are still in need of human supervision to ensure accuracy. The spread of Web 2.0 approaches demonstrate the possibility and the added value of using collaborative techniques for improving data sharing and consensus reaching. In this paper, we describe our prototype for collaborative ontology mapping and data sharing. The possibility to exploit ontology alignments for querying data is a key capability for data sharing in a networked ontology environment.

1 Introduction

The envision of a Web of data is one of the most prominent targets of Semantic Web (SW). To help reach this goal, knowledge repositories need to publish semantic representations of their data models to enable other machines to understand and query their content. Due to the highly distributed and decoupled nature of the information present on the web, the vision of the SW is moving towards a scenario where the task of creating and maintaining ontologies, that formalise data semantics, is going to be handed to the community that actually uses them [13]. Such an effort must be supported by tools and methodologies that allow latent models to emerge as a product of a collaborative effort and dialogue. To this end, much research and development has focused on building tools and capabilities for ontology and KB construction. However, support for distributed teams to remotely and continuously collaborate on mapping ontologies and knowledge repositories is still underdeveloped.

In this paper we describe our prototype for data integration based on collaborative and social ontology mapping, our system allows to: **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.

2 Related Work

Investigations into enhancing user knowledge through collaboration and sharing goes back to the early nineties [12]. The Semantic Web has taken this approach further by providing the tools and languages to construct networked semantic representational layers to increase understandability, integration, and reuse of information. The rise of Web 2.0 approaches has then demonstrated the effectiveness and popularity of collaborative knowledge construction and sharing environments that adopted lighter version of ontologies, where the emphasis is put on the easiness of sharing knowledge rather than creating or adopting static formal ontologies [2]. Harnessing Web 2.0 features to facilitate the construction, curation, and sharing of knowledge is currently pursued by different communities.

Collaborative Protègè [14] was recently developed as an extension to Protègè to support users to edit ontologies collaboratively, by providing them with services for proposing and tracking changes, casting votes, and discussing issues, thus infusing classical ontology editing with a number of popular social interaction features. Another ontology editor with collaborative support is Hozo [10], which focuses on managing ontology modules and their change conflicts.

Other approaches use social tagging as the main driver for enacting collaborative lightweight ontology building [7, 16]. Similarly, other tools are focussing on editing instance data by adopting a wiki philosophy, like OntoWiki [1], or by implementing a framework API, like DBin [15]. Most of the tools listed above focus on supporting users to collaboratively construct ontologies or to collaboratively populate an ontology with instance data. Unlike these tools, however, our proposed system, OntoMediate, extends the collaborative notion to support the task of *ontology mapping*, where users can collaborate and interact to map their existing ontologies and maintain a quality mapping asset within the community. Such alignments are then used to achieve ontological mediation for accessing locally hosted data.

An approach related to OntoMediate, that addresses ontology mapping within communities, is the Zhadanova and Shvaiko [17] method. The authors proposed to use similarity of user and group profiles as a driver for suggesting ontology alignments reuse. The focus of that work was on building such profiles to personalise reuse of ontology mappings. In OntoMediate, we are exploring the use of collaborative features (discussions, voting, change proposals) to facilitate the curation and reuse of mappings by the community, to facilitate a social and dynamic integration of distributed knowledge bases. Our approach is novel in the way it addresses the task of aligning ontologies, by extending and enhancing automatic mapping tools with a full community support. In our approach, alignments are seen as a resource, built and shared by a community. The community is able to investigate, argue, and correct individual mappings, using various supporting services provided by OntoMediate.

3 OntoMediate Features for Community Support

In the OntoMediate project [3] we are studying how social interactions, collaboration and user feedback can be used in a community, in order to facilitate the alignment of ontologies and to share mapping results.

3.1 OntoMediate System

The prototype has been designed to be extendible, allowing off-the-shelf tools to be integrated and is composed of three main subsystems: ontologies and datasets manager; ontology alignment environment; social interaction environment.

Ontologies and Datasets Manager. This part of the system allows users to register (as well as unregister) the datasets they intend to share with the community and the ontologies that describe their data vocabulary. The ontologies that are loaded onto the system need to be aligned with one or more shared ontologies in order to enable access to its data by the community.

Ontology Alignment Environment. Our system provides an API for automated ontology alignment tools to be plugged in. Examples of tools already integrated with our system are CROSI mapping system [9] and Falcon OA [8]. These tools allow the system to support the alignment task by proposing to the user some initial candidate mappings. The system allows the user to link his/her local ontologies to shared and agreed ontological models that can be accessed by anyone in the community.

Social Interaction Environment. This functionality allows community members that deal with similar data - and therefore have a mutual interest to maintain a good quality alignment asset - to socially interact with each other. The system provides a way to browse and review the local alignments toward shared concepts, allowing users to spot errors, provide feedback, propose and discuss alternative solutions and conflicting interpretation (see Figure 1). Aim of the social interaction system is to exploit community feedback in order to enhance the overall quality of the ontology alignment and achieve agreement on semantics of concepts by means of community acceptance.

In OntoMediate system, one of the aims is to reuse user input to increase the quality of data integration and ease future alignments toward the same target ontology. One way for achieving this goal is to take into consideration lexical information originated by users' ontologies. In fact, lexical labels can be adopted by the shared model as *rdfs:label* that can be considered in future automatic alignment tasks in an attempt to improve performance and accuracy of automatic mapping tools. The working assumption is that allowing the system to learn the community lexicon will improve future automatic ontology alignments that usually rely on such information.

3.2 Data Integration support

Because the information could be described differently in different ontologies (e.g. using different systems of measurement) we developed a service for translation of RDF data and queries between aligned ontologies. So far the features described above allow a community to maintain a network of ontologies, data sources and an alignment asset that enable access to data sets that are described with local ontologies. In order to integrate data from sources described by different ontologies, the collaboratively maintained alignments must be exploited for translating queries and data.

Different approaches have been proposed over the

Figure 1. Ontology evolution forum

years by the semantic web community for dealing with this task by using alignments for creating:

- **rules** for the translation of data between two aligned ontologies, namely the source and the target ontology
- **queries** for the creation of data compliant to a target ontology addressing data sources compliant to a source ontology

In OntoMediate the ontological mediation is achieved by means of query translation. Similarly to the approach adopted by Euzenat et al. [6] the alignments managed by the users and the topology of the data network are both exploited to translate queries and results from one ontology to another. Within the OntoMediate scenario, the shared ontologies are used to define the SPARQL queries that translate between the terms of one or more local ontologies for accessing the user data. Moreover, since the results need to be described with the same ontology used for the query in order to be understood by the issuer of the query, the translation mechanism developed allows to retranslate the results back into the source ontology.

Currently, ontological relations are used to define equivalences among concepts and object properties. Local entities that are not equivalent to an entity belonging to a shared ontology, shall be proposed and discussed as an extension to the shared model. Alignments among datatype properties allow to define $n : 1$ relations and complex value transformations, by means of scripts, needed to bridge different encoding systems. A typical example where such complex transformations are needed is when different information systems decide to encode point coordinates with two different expressions(e.g. is_1 with cartesian and is_2 with cylindrical coordinates). The notion of *point* is easily mappable between the two ontologies (i.e. $is_1 : Point \equiv is_2 : Point$) but the information described within the two concepts are not always directly mappable. In order to be able to translate a point from a cartesian system (x, y, z) to a cylindrical system (ρ, φ, z) we need to be able to describe value transformations (i.e. $\rho = f(x, y) = \sqrt{x^2 + y^2}$). For representing such alignments, the usual representation by quadruple [5] $\langle e, e', R, n \rangle$ has been extended: the source entity e is replaced by an ordered set of entities E and a scripting function s is associated with the alignment. The representation of an alignment in our system is therefore a quintuple: $\langle E, e', R, n, s \rangle$.

Our translation algorithm visits and rewrites the algebraic description of a SPARQL query [4]. This implementation exploits the features provided by Jena API for: rewriting the RDF triple patterns present in the original query; obtaining a query that matches the target ontology; retrieving the required instance set; and rearranging the result set to fit the source ontology.

4 Comparison

This section provides a brief comparison of the features provided by OntoMediate and other tools for collaborative knowledge construction and sharing [11] (the comparison is limited to the tools presented at the CKC 2007 workshop[1]). Table 1 compares the features provided by some of the latest tools for collaborative knowledge construction and sharing. All the tools manage ontological structures except Bibsonomy that does not handle properties and instances and SOBOLEO that does not handle properties. Only Protégé and OntoMediate manage ontology alignments but only OntoMediate provides query translation capabilities. Bib-

[1]The comparison is based on the features that the tools had in the versions that participated in the CKC 2007 challenge.

Sonomy, Hozo and OntoMediate allow users to manage personal spaces where the users are free to share only part of their ontology and instances. Among the social features compared, OntoMediate is the only tool that provides both argumentation, discussion and user notification features.

Table 1. Features comparison

	BibSonomy	DBin	Hozo	OntoWiki	Collab. Protégé	SOBOLEO	OntoMediate
Hierarchy of concepts	✓	✓	✓	✓	✓	✓	✓
Properties		✓	✓	✓	✓		✓
Instances		✓	✓	✓	✓	✓	✓
Ontology mapping					✓*		✓
Query translation							✓
Ratings or voting				✓	✓		✓
Personal space	✓		✓				✓
History of changes		✓	✓	✓	✓		
Discussion		✓			✓	✓	✓
Argumentation					✓		✓
User notification		✓					✓
Web browser interface	✓			✓		✓	✓

5 Summary and Future Work

This paper presented a prototype for supporting ontology mapping and data sharing with community interactions, where users can collaborate on aligning their ontologies, and manually-driven alignments can be stored and reused later. The alignments can then be exported or used to translate queries into different ontologies for merging results from different data sources. Next, we plan to run experiments to test the validity of the social approach, and the usability of the services and features that it provides. We will also implement services for handling the discovery and exploitation of instance mappings. Such service will be of paramount importance for the data integration task since it is likely that information relative to the same topic shall be scattered in different data sets.

6 Acknowledgements

This work was funded by a grant awarded to General Dynamics UK Ltd. and the University of Southampton as part of the Data and Information Fusion Defence Technology Centre (DIF DTC) initiative. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the UK Ministry of Defence, or the UK Government.

References

[1] S. Auer, S. Dietzold, J. Lehmann, and T. Riechert. OntoWiki: A tool for social, semantic collaboration. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC) at WWW 2007*, Banff, Canada, 2007.

[2] G. Correndo and H. Alani. Survey of tools for collaborative knowledge construction and sharing. In *Workshop on Collective Intelligence on Semantic Web (CISW 2007)*, November 2007.

[3] G. Correndo, Y. Kalfoglou, P. Smart, and H. Alani. A community based approach for managing ontology alignments. In *16th International Conference on Knowledge Engineering and Knowledge Management Knowledge Patterns (EKAW 2008)*, 2008. To appear.

[4] R. Cyganiak. A relational algebra for SPARQL. Technical report, HP Labs, 2005.

[5] J. Euzenat. Alignment infrastructure for ontology mediation and other applications. In *Proceedings of the First International workshop on Mediation in semantic web services*, pages 81–95, 2005.

[6] J. Euzenat, A. Polleres, and F. Scharffe. Processing ontology alignments with SPARQL. In *International Workshop on Ontology Alignment and Visualization - OnAV'08*, 3 2008.

[7] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. BibSonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, 2006.

[8] N. Jian, W. Hu, G. Cheng, and Y. Qu. Falcon-AO: Aligning ontologies with falcon. In *Workshop on Integrating Ontologies (K-CAP 2005)*, pages 85–91, 2005.

[9] Y. Kalfoglou, B. Hu, D. Reynolds, and N. Shadbolt. Capturing, representing and operationalising semantic integration (CROSI) project - final report, October 2005.

[10] K. Kozaki, E. Sunagawa, Y. Kitamura, and R. Mizoguchi. Distributed and collaborative construction of ontologies using hozo. In *Proc. WWW 2007 Workshop on Social and Collaborative Construction of Structured Knowledge*, Banff, Canada, May 2007.

[11] N. F. Noy, A. Chugh, and H. Alani. The CKC challenge: Exploring tools for collaborative knowledge construction. *IEEE Intelligent Systems*, 23(1), 2008.

[12] R. Patil, R. Fikes, P. F. Patel-Schneider, D. McKay, T. W. Finin, T. R. Gruber, and R. Neches. The DARPA knowledge sharing effort: A progress report. In *KR*, pages 777–788, 1992.

[13] N. Shadbolt, T. Berners-Lee, and W. Hall. The semantic web revisited. *Intelligent Systems, IEEE*, 21(3):96–101, 2006.

[14] T. Tudorache and N. Noy. Collaborative Protégé. In *Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007*, Banff, Canada, 2007.

[15] G. Tummarello, C. Morbidoni, and M. Nucci. Enabling semantic web communities with DBin: An overview. In *Proc. 5th Int. Semantic Web Conf., ISWC 2006, Athens, GA, USA*, 2006.

[16] V. Zacharias and S. Braun. SOBOLEO – social bookmarking and lightweight engineering of ontologies. In *Proc. WWW 2007 Workshop on Social and Collaborative Construction of Structured Knowledge*, Banff, Canada, May 2007.

[17] A. V. Zhdanova and P. Shvaiko. Community-driven ontology matching. In *ESWC*, pages 34–49, 2006.