A community based approach to managing ontology alignments

How to cite:

For guidance on citations see FAQs

© The Authors
Version: Accepted Manuscript
Link(s) to article on publisher’s website:

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
A Community Based Approach for Managing Ontology Alignments

Gianluca Correndo, Harith Alani, Paul Smart

University of Southampton,
Electronic and Computer Science Department
[gc3, ha, ps02v]@ecs.soton.ac.uk,
WWW home page:http://ecs.soton.ac.uk
SO17 1BJ, United Kingdom

Abstract. The Semantic Web is rapidly becoming a defacto distributed repository for semantically represented data, thus leveraging on the added on value of the network effect. Various ontology mapping techniques and tools have been devised to facilitate the bridging and integration of distributed data repositories. Nevertheless, ontology mapping can benefit from human supervision to increase accuracy of results. The spread of Web 2.0 approaches demonstrate the possibility of using collaborative techniques for reaching consensus. While a number of prototypes for collaborative ontology construction are being developed, collaborative ontology mapping is not yet well investigated. In this paper, we describe a prototype that combines off-the-shelf ontology mapping tools with social software techniques to enable users to collaborate on mapping ontologies.

1 Introduction

The transformation of the Web from a mere collection of documents to a queryable Knowledge Base (KB) is one of the most prominent targets of Semantic Web (SW) [1]. 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. 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 building and updating ontologies and knowledge repositories is still underdeveloped.

Defining an ontology for representing data semantics is usually a costly and time consuming task. Furthermore, knowledge evolves over time which adds to maintenance cost. That is why more and more often successful proposals for information sharing involve user’s feedback exploiting a network effect. If an ontology is meant to reflect the views of a specific community and support their knowledge sharing tasks, then the community itself should be empowered to express, formalise, share and maintain a set of ontologies for supporting such tasks [2]. Some ontologies need to be agreed upon by the user community, and this agreement process must be supported by tools and methodologies to allow users to express their views and opinions freely.
The rise of social Web 2.0 applications has demonstrated how general Web users can actively contribute and share all sorts of data and information, such as images, videos, bookmarks, opinions, diaries and experiences. Adopting a similar approach on the SW means supporting users to dynamically and collaboratively build ontologies, add semantics to data, discuss and share views and suggestions, etc. Good and colleagues [3] showed how SW users can successfully collaborate to negotiate and build good quality ontologies when provided with a tool that supports such activities. User-contributed content can also be beneficial for engineering ontology mapping activities, most of which rely on automated linguistic and statistical methods that make use of lexicographic clues and structural information but rarely take into account user input [4]. In this paper we describe a prototype and its underlying approach for facilitating gradual ontology mapping by supporting social collaboration and reuse of mapping results. More specifically, our approach allows the following:

– **Alignment of local ontologies to shared ones**: users can align local models, used for bridging data sources, to shared ontologies by using a number of automated ontology mapping tools. These tools are flexibly plugged into our system;

– **Social interaction and collaboration**: users can discuss ontology alignments and propose changes through a number of social services, such as discussion and voting facilities;

– **Reuse of ontology alignment information**: users can add to, and correct, the alignments suggested by automated ontology mapping tools, or suggested by other users. User feedback and mapping information are logged by the system and reused to improve the accuracy of future alignments on similar concepts;

2 Related Work

The need to make explicit and publish the semantics of the data is becoming increasingly central since more information systems are becoming largely decoupled and separately managed. To this end, 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 [2]. In accordance with this vision, the models for making data semantics explicit and exchangeable can be the fruit of a collaborative effort by the community members whom will share the responsibility of ontologies creation and maintenance. 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.

Our work taps on the intersection of different but overlapping areas in ontology engineering: collaborative construction and management using social networking tools, data web and sharing of ontology fragments. We briefly highlight the main contenders in these areas and elaborate on their relationship with our work.
Historically speaking, investigations into enhancing user knowledge through collaboration and sharing goes back to the early nineties [5]. Ontolingua [6] is an early proposal in this area, which provides some basic support for users to reuse and extend shared ontologies. Another example is the model discussed by Euzenat in [7], where users can build their local ontologies, get them approved by the community, and get support by a discussion protocol which conveys users’ rationales for changes in a formal schema. 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 [5-9]. Harnessing Web 2.0 features to facilitate the construction, curation, and sharing of knowledge is currently pursued by different communities. Collaborative Protégé [10] was recently developed as an extension to Protégé 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 [11], which focusses on managing ontology modules and their change conflicts. Good and colleagues demonstrated how good quality ontologies can be built quickly in a collaborative fashion [3]. Other approaches use social tagging as the main driver for enacting collaborative lightweight ontology building (e.g. [12-13]). Similarly, other tools are focussing on editing instance data, like OntoWiki [14] and DBin [15] which are prime examples of tools for community-driven knowledge creation.

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. An approach similar to OntoMediate, that addresses ontology mapping within communities, is the Zhadanova and Shvaiko [16] 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 ontological mappings by the community, to facilitate a social and dynamic integration of distributed knowledge bases. The use of collaboration for achieving consensus on terms’ semantics is largely justified because of the social nature of ontologies. In order to mediate possibly conflicting concept’s description, user feedback must be taken into account and discussion within the community must be fostered. Our approach is novel in the way it addresses the task of aligning ontologies, by ex-
tending 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 the individual mappings, using various supporting services provided in OntoMediate.

3 The OntoMediate Approach

In the OntoMediate project we are studying how social interactions, collaboration and user feedback can be used in a community in order to ease the task of ontology alignment and ontology mapping sharing. Focus of our research is how to ease the integration of data sources using ontologies and ontology alignments in order to provide an agreed semantics to integrated data.

The implemented prototype is a Web application developed with J2EE and AJAX technologies. The system manages OWL ontologies that are parsed using the Jena API. The system has been designed to be extended via its APIs and is composed of three main subsystems:

– Ontologies and datasets manager;
– Ontology alignment environment;
– Social interaction environment.

3.1 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 querying of the published data by the community. The system currently supports different storage types for the ontologies and/or datasets:

– URL: only the URL is stored and the ontology is accessed (read only) remotely;
– Cached file: the ontology file is uploaded to the system and stored in a file server;
– Jena RDBMS: the ontology file is uploaded to the system and stored in a relational database using the Jena database back-end;
– SPARQL endpoint: the document is remotely accessed using the SPARQL protocol.

Once an ontology is registered with the system, the owner (or everyone if the ontology has been shared within the community) can browse it by using a flexible frame-like interface. The ontology browser displays the hierarchy of concepts, as well as detailed information for the focused concept (selected concept). The detailed information includes: labels, superconcepts, subconcepts, equivalent concepts, concept description (from the rdfs:comment annotations), properties and their constraints.

1 http://jena.sourceforge.net
2 http://www.w3.org/TR/rdf-sparql-protocol/
3.2 Ontology Alignment Environment

The full automation of ontology alignment is not an easy task \([18]\). The factors that affect the computation and accuracy of ontology alignments are so delicate that we cannot afford not to take into account user input as a contributing factor of paramount importance. It is for this reason that, implementing an environment for aligning ontologies, great attention has been made to the usability issues that could affect this task \([19]\).

Our system provides an API for automated ontology alignment tools to be plugged in and also maintains data structures to store parameters needed by a particular tool to execute (e.g. threshold values or available tool options). The API allows for easy integration of new alignment tools, when they become available, by means of wrappers - some tools have been already integrated with our system (e.g. CROSI mapping system \([20]\), INRIA Align \([21]\) and Falcon OA \([22]\)). These tools allow the system to support the alignment task by proposing to the user some initial candidate mappings. The results from different tools can be merged and the decision of which combination of tools to use can be parameterised together with the configuration used to invoke each tool. The merge of results from different tools is achieved by a weighted mean of each contribution and it is implemented as a normal alignment tool plugged into the system (i.e. different merging algorithms can be coded and plugged in).

Once the automated mapping has been executed, the results are displayed in a proper interface for reviewing and for searching further alignments. The ontology alignment interface is split into two main panels, the left panel for the source ontology and the right panel for the target ontology, whereas the bottom space is used for summarising the mappings found for the focused source concept. The interface has two view modalities: Hierarchical and Detailed.

In the Hierarchical view the two taxonomies are centered on the source concepts that have been mapped to a target concept, both of which are highlighted. The user can browse both taxonomies and create new mappings by dragging a source concept and dropping it into a destination concept. When the user focuses on a mapping, he/she can switch to a detailed view and the description of the source and target concept are shown side by side.

In the Detailed view, the user can map the properties using the same drag & drop facility used for mapping the concepts. The users can also explicitly reject some automatically proposed mappings. This choice will be recorded by the system and will be used to filter future mappings towards this target concept, thus increase future ontology alignment precision. Alternative interface designs for ontology mapping, such as the one presented in \([23]\), will be considered for future version of the system.

3.3 Social Interaction Environment

This functionality allows users of a community that deal with similar data - and therefore have a mutual interest to maintain good quality alignments - to socially interact with each other. The aim of the social interaction 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. This subsystem displays to the user three views: **Ontology** view; **Mappings** view and **Forum** view.

![Discussion environment - Ontology View - Post](image)

**Fig. 1.** Discussion environment - Ontology View - Post

The **Ontology** view (see Figure 1, top-left corner) displays an enhanced taxonomy browser for the selected shared ontology. The enhancements concern the user activities affecting the shared concepts, visualising additional information (e.g. number of incoming mapping per concept are reported in brackets like the number of post exchanged in the forum discussing such mappings). Moreover, the interface allows to inspect the set of labels used for equivalent concepts (i.e. the ones provided with the alignments) in local ontologies (see the **Additional labels** text field in Figure 1). The user or administrator can edit such labels and add them to the shared concept to enrich the concept description with users’ contributions. The new mapping, and the edited/added labels, will be logged in a database to be reused later to improve the **recall** of future ontology alignment tasks (section 4.2).

When the user selects a concept that has some user mappings associated with it, he/she can switch to the **Mappings** view that displays information about the local mappings for the focused concept. The user can then inspect a summarised description (i.e. subconcepts, superconcepts, properties etc.) of the local concepts and decide if they are relevant to the focused target concept or initiate a discussion thread in order to change them. The change proposal is composed of a thread post, that describes in natural language the content of the proposal,
and a formal description of the operation to discuss. The proposed change can affect a number of alignments and may lead, if the proposal is accepted, to the relocation of such alignments to a different target concept. If the target concept referenced in the change operation is not yet present in the ontology, a new one will be created within the hierarchy in accordance with the input given by the users in the forum. The possibility to create new concepts to host user alignments provides a way to reshape (even if only by additions) the target ontology in function of the (meta)data provided by users.

The system provides a forum for the discussion of the users’ proposals (see Figure 1 bottom-right corner). Every time a user proposes a change using the mappings view, a new thread is created in the forum and other users are free to debate the proposal, reply the proposal with a new one or simply agree or disagree with it. The user’s vote is computed for update the proposal statistics (i.e. number of votes, percentage of approvals and disapproval) that is promptly displayed along the proposal.

The new action item associated with a target concept is notified to every interested user by means of RSS feeds whose the interested users can subscribe to. Once a proposal has reached a critical mass (e.g. when the majority of users affected by the change have expressed their opinion) it will be endorsed, or submitted to the administrator in order to judge it and reach a final decision.

4 Working Example

In order to better explain our approach and show how users’ feedback can be used in order to improve the ontology matching task, we report on a small example in the chemical domain and the findings of a working experiment. In this example, two users want to share information on hazardous chemical compounds. They each create an ontology that reflect the nature and structure of their data sources (in our example the users deal with data about Landmines and Hazardous Components, see Table 1).

<table>
<thead>
<tr>
<th>Name</th>
<th>Domain</th>
<th>n°Concepts</th>
<th>Main Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shared Ontology</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td>Chemistry</td>
<td>130</td>
<td>Element, Compound, Explosive</td>
</tr>
<tr>
<td><strong>Local Ontologies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landmine</td>
<td>Explosive devices</td>
<td>830</td>
<td>Country, Explosive Device, Material</td>
</tr>
<tr>
<td>Hazardous Components</td>
<td>Hazardous materials and</td>
<td>89</td>
<td>Explosive, Flammable, Container</td>
</tr>
</tbody>
</table>

Table 1. Domain ontologies used in the experiment
4.1 Alignment task

This tiny community is provided with a shared domain ontology where a set of entities and relationships relevant to the chemical domain is defined (see Table 1). The two users need to align their local ontologies to the shared one in order to exchange information and integrate their data. To fulfill this task, the users use off-the-shelf automatic tools with the Ontology Alignment environment (see section 3.2). The automatic ontology alignment tools provide an initial set of alignments that the users can revise, using the system interface explicitly stating the correct alignments and the incorrect ones. With the same interface, the users can then browse the two ontologies and provide manual alignments if required.

At the moment only equivalence relation is supported for expressing alignments but the adoption of more expressive primitives is under study. In this scenario the local ontologies act as “contexts” of their respective data sources (following the nomenclature used by Bouquet et al. [24]) while the shared ontology is meant to provide an ontological formalisation of the domain to enable the actual data integration. They are the objects that catalyse the consensus process.

4.2 Reuse of information from mappings

The alignments provided by the alignment task will be reused to improve automatic future alignments toward the same target ontology. Lexical labels from users’ ontologies 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. Within the chosen domain (i.e. hazardous chemical compounds, but the assumption holds in other domains), different labels can represent the same concept (e.g. the explosive HMX is also known as Octogen or Cyclotetramethylene-tetranitramine, see Table 2 for a summary of the labels logged from the alignment activity). The working assumption is that, gathering all the labels related to a concept from local representations, and learning which alignments must be avoided in the future (e.g. rejected by users), can help to increase the performance of automated alignments. As an example, assuming the two users of this example have subsequently aligned their ontologies, the labels collected from the first alignment (see Table 2) can be used for improving the performances of the second. Manual mappings discovered by the first user (e.g. Black Powder ⊑ Gun Powder or Nitromethane ⊑ Nitrocarmine) can in fact helping the discovery of target concepts that would be missed otherwise by automatic tools. Such additional user’s labels can in fact bring, if integrated in the shared model, to an increase in automated tools precision and recall for subsequent alignments.

4.3 Social interaction

Browsing the definition of the shared ontology, the users can revise each other’s alignments to check that the definition of the local concepts is relevant to the targeted shared concept. The self curation of the shared alignments is an important premise of the approach; users that are interested in integrating their data
Table 2. Alignments based on past users activity

<table>
<thead>
<tr>
<th>Source concept</th>
<th>Target concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Powder</td>
<td>Black Iron Oxide</td>
</tr>
<tr>
<td>Magnesium</td>
<td>Nitromethane</td>
</tr>
<tr>
<td>Red P</td>
<td>White P</td>
</tr>
<tr>
<td>Black Iron Oxide</td>
<td>Magnetite</td>
</tr>
<tr>
<td>Nitromethane</td>
<td>Nitrocarbol</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discovered by system and proposed to user</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Powder ≡ Gun Powder</td>
</tr>
<tr>
<td>Magnesium ≡ Mg</td>
</tr>
<tr>
<td>Red P ≡ Red Phosphorus</td>
</tr>
<tr>
<td>Black Iron Oxide ≡ Magnetite</td>
</tr>
<tr>
<td>Nitromethane ≡ Nitrocarbol</td>
</tr>
<tr>
<td>White P ≡ White Phosphorus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learnt from user input to be wrong and rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Iron Oxide ≡ Iron Oxide</td>
</tr>
<tr>
<td>Nitromethane ≡ Nitroethane</td>
</tr>
</tbody>
</table>

or in querying the integrated knowledge base have a main concern in browsing such alignments, providing feedback and starting corrective operations whenever needed.

Automated ontology alignment tools usually fail to catch the difference among lexically similar concepts such as Nitromethane and Nitroethane. Despite their lexical and chemical similarity, it is very important to distinguish the two (the first can be used as an explosive while the second can not). For this reason, once a user has found the incorrect alignment (i.e. Nitromethane ≡ Nitroethane) inspecting the local concept definition, he/she can select the faulty alignment and initiate a change process. Along with the incorrect mapping, the user can provide the URI of the suggested correct target concept (i.e. Nitrocarbol, a synonym of Nitromethane) and issue a change proposal. If no suitable concept can be found in the target ontology the user can suggest the creation of a new one providing its location in the targeted hierarchy. The proposal will be posted in the forum dedicated to the maintenance of the shared concept alignment asset. The community can be alerted of the change proposal by RSS feed subscription (every target concept has a feed where new posts are published, and every interested user can register to the feed) and inspect the change proposal, discuss it on the forum, replying to the post or just expressing dis/agreement with the content of such proposal.

4.4 Alignment asset management

Once the two ontologies have been aligned with the shared model, they can be exploited for assuring a meaning preserving information exchange between the components of the community. The discussion fostered in the social environment and the constant supervision by the users upon the ontology alignments help in maintaining agreement and awareness on terms’ semantics within the community.

5 Discussion

Collaborative ontology mapping has a great potential in enhancing performance and in sharing results of automatic mapping tools. The system presented in this paper supports users in their ontology mapping activities and logs their feedback to further enhance the output of automated ontology mapping tools. Moreover
it provides social features for community driven mapping revisioning and limited support for shared ontology evolution.

Ontology mapping is inherently difficult, and can be influenced by various issues. For example, some mappings can be user or context dependent, in which case a mapping that has been approved by some users may not necessarily suit others. **Mapping popularity** can be used to weight each ontology alignment. The degree of popularity of a specific alignment can be taken into account when displaying alignment suggestions to the user. Storing user profiles to personalize mappings has been proposed elsewhere [16].

When reusing mapping results, it is important to prevent error propagation. It is important to build a user interface in such a way to discourage blind reuse of mappings. OntoMediate allows the community to flag, discuss, and democratically change incorrect mappings, but this is of course dependent on users spotting erroneous mappings. If a mapping is reverted, it will be important to readjust its popularity accordingly.

In addition, mappings that receive repeated change proposals or become subject to long and intense discussions may be regarded as **controversial or debatable mappings**. Such mappings may also need to be handled with care when used or reused suggesting administrators to create appropriate ontological description to better characterize those particular local concepts.

OntoMediate uses off the shelf automatic ontology mapping tools, and hence the complexity of its mappings are largely based on those of the mapping tools. The current implementation of OntoMediate allows users to manually map entities expressing simple one to one mapping. More complex mappings, such as mapping a union of classes or linking properties by means of transforming functions, is not currently supported. However, it has been reported that when engineering ontologies collaboratively, complex OWL constructs are often not required [9].

Ontology mapping is a not an easy task, and hence users will not expected to link their ontologies without a clear added value. The ultimate goal of OntoMediate is to facilitate distributed querying and integration of knowledge bases in a community. Therefore, in addition to displaying concept mappings, it will be important to also display some information about the knowledge that each mapped ontology brings to the table. Showing what data a specific mapping or a whole ontology is bringing to the community might encourage others to (a) see the general value of this mapping and hence offer their expertise and help to map the new ontology correctly, and (b) map their ontologies to others if they have not already done so (e.g. to link their data to the new repository).

The approach we focused on in OntoMediate is based on a small to medium size community, sharing interests and goals that can benefit from integrating their data. In OntoMediate, it is presumed that an overall administrator can act as the ultimate curator of the system. For such an approach to scale up to the Web as a whole, the wisdom of the community will have to be the final ruler. Wikipedia is a fine example of how this can work, and the Linked Data initiative is a first step to creating a wide network of linked semantic data [25]. However, demonstrating added value will be more difficult once the community
is too large and diverse, and hence it will probably breakup into sub communities with similar requirements.

6 Summary and Future Work

This paper presented a prototype for supporting ontology mapping with community interactions, where users can collaborate on aligning their ontologies, and manually-driven alignments can be stored and reused later. Our initial experiment showed good potential of increasing both precision and recall in ontology mapping when reusing past mapping results. Next, we plan to run much larger experiments to further test the validity of the approach, and the usability of the services and features that it provides. We have lately implemented services that exploits the managed alignments for translating queries and data. In the near future we will also implement services to allow users to submit formula to mediate between concepts or data that might not be directly mappable (e.g. when the concepts are culture-dependent, or when data property values are function of different other values). Additionally, we will next focus on building the capability to allow users to perceive, and query, the integrated KBs, thus increasing added value. The ontology alignments and the social network will be exploited to focus the search task. We will make the system available to the public online in the next few weeks.

7 Acknowledgements

This work was partially 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