Towards Ontology Mapping for Ordinary People

Colm Conroy[1], ,
(Co)Supervisors: Declan O'Sullivan[1], David Lewis[1]

[1] Knowledge and Data Engineering Group,
Trinity College Dublin
{coconroy,declan.osullivan,dave.lewis}@cs.tcd.ie

1 Introduction

The reasons for the lack of uptake of the semantic web amongst ordinary users can be attributed to technology perception, comprehensibility and ease of use. It is perceived that the creation of ontologies is a top-down and complex process, whereas in reality ontologies can emerge bottom-up and be simple. Ontology technology is based on formal logics that are not understandable for ordinary people. Finally there is significant overhead for a user in the creation of metadata for information resources in accordance with ontologies. To address these three problems, we believe that the interfaces to ontology tools will need to be engineered in such a way as the tools disappear into the background from the ordinary person's perspective.

There is a common diversity between the semantic models of interest between people. If users of the semantic web choose to model their interest with a personal ontology, their ontology will need to be mapped to the models used in the various diverse communities by the person himself or herself. The automatic and efficient matching between the personal ontology and the models used by others (collaborative tags and/or community ontologies) can be achieved through the application of a variety of matching techniques [1]. Fully automatic derivation of mappings is considered impossible as yet [2], and the majority of state of the art tools in the ontology mapping area [3] and the community ontology creation area [4] rely on a classic presentation of the class hierarchy of two ontologies side by side and some means for the user to express the mappings. These approaches predominately assume that the mapping is being undertaken by an expert: who does not require a personalized interface; whose explicit task is to generate a "one size fits all" full mapping (to be used in common by several applications); and who undertakes the task during a small number of long sessions. The number of user trials that have taken place have also been small [5] and those that have, have focused purely on the effectiveness and do not address usability issues (an exception recently being that of [6]). In contrast, we propose that the user who will benefit from mappings (through usage by their applications), will undertake themselves partial targeted mappings, gradually and over time, using techniques that address usability issues, support personalization and enable control of the mapping interactions.

2 State of Art

There is an emergence of focusing on better support for users within the ontology mapping area with cognitive support for user [7], a community-driven matching [8], explanation of matches to users [9], and developing a formal model for ontology mapping [10]. One of the key problems which have seen little research from a cognitive perspective is how to display the match information in a manner that is natural for the user. The visualization of ontology (schema) mapping can be categorised into four different categories: *tree-type*, *object-type*, *instance based spreadsheet*, and *hyperbolic*. *Tree-type interfaces* are the most common for mapping tools and represent the semantic models side by side in a tree form, e.g. COMA++ [3], and mappings can be represented in two different ways with lines drawn between matching terms or via a mapping table that contains all the mappings. *Object type interfaces* represent the ontology in an object type or UML structure, with mappings between the two schemas being drawn using lines (and symbols to represent the type of mapping), e.g. SMART [11] '=' is used for equals, ')' for subset and so on. *Spreadsheet type interfaces*, e.g. Webscripter [12], uses spreadsheet functionality to display global mapping tables containing instance data from different users. Finally a *hyperbolic interface* is one where the source and target models are represented by hyperbolic graphs in different frames, e.g. Schema Mapper [13]. AlViz [14] is a tab plug-in for Protege which provides visualization techniques to facilitate user understanding of alignment results. There are other types of ontology tools which use different types of interfaces; GINO [15] is a guided input natural language ontology editor that allows users to edit and query ontologies in a language akin to English. OntoViz [16] is a protégé plugin which represents ontologies via a direct graph with concepts in boxes and relations defined with lines.

3 Research Question/Contribution

The research question we are addressing is *what kind of interactions will be acceptable, efficient and effective for an ordinary user to achieve semantic mappings gradually and over time between information models of interest to the user*. In particular our work will provide:

- *Design of a mapping framework in support of ordinary people*: There is a need to make the ontology mapping process as unintrusive and as natural as possible, as it is important not to interrupt ordinary users during their daily life, so that they do not see mapping as inconvenient work but more as something that will be beneficial to them, and where they can clearly see the benefits.
- *Determine the most appropriate user interaction in the process of constructing mappings*: There has been little or no research by the current state of the art completed regarding user interaction within mapping systems. While visualization is a key factor in this problem another key issue is determining the most appropriate (different) way(s) for (different) people to construct mappings, e.g. draw lines graphically, 'yes/no' answer on questionnaire, etc…

- *Determine the appropriate ways to engage the user over time*: There is a need to engage the user in the mapping process over time rather than one long session. This need comes from the realization that by reducing the mapping process into piece wise comparison will make the mapping process easier to comprehend and also will be able to give feedback to the user of the mapping choices they make outside of the mapping sessions. Another issue will be does collaborative knowledge sharing (via groups) assist ordinary people in the mapping task

An outcome of this research will be a process (Figure 1), methodology and tools.

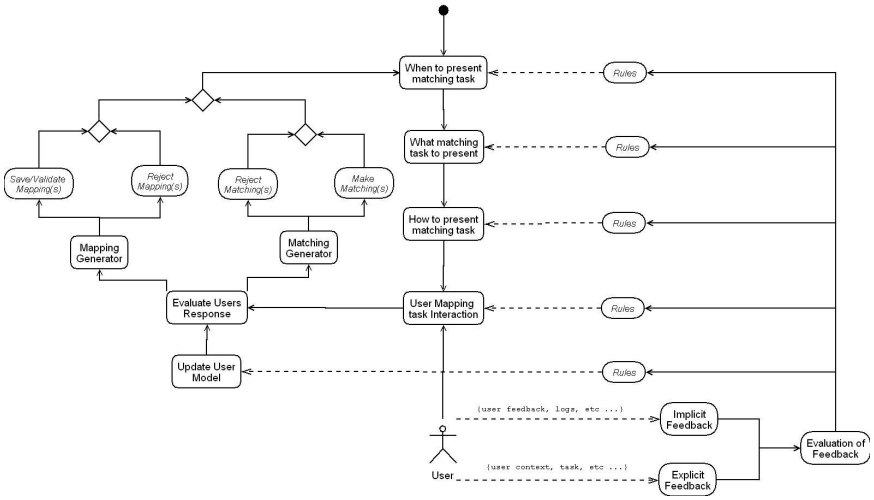

Figure 1: Mapping Process

4 Evaluation

In our initial experiment undertaken early in 2007 we aimed to determine the most practical way of visually displaying the mapping information for different groups of users. Our hypothesis was *using a Question & Answer natural language interface to visually display ontological information helps in making mapping more familiar and accessible and also reduces the complexity of the mapping process for users*. The intention of our experimentation was to investigate the effect and usability of a natural language prototype tool (NL) on three groups of users and to determine whether it made the mapping task more user friendly for one group over another. In addition we wanted to contrast our tool against a current state of the art mapping tool. We chose COMA++ as our State of the Art tree type graph mapping tool. The three different groups of users were: Ontologically aware, Technology aware and Non-Technology Aware (Casual Web User). The paper [17] goes into detail about the experiment, some key conclusions drawn were:

- On the positive side, results suggested casual web users can map effectively and efficiently even compared to ontology aware users. Using Natural Language seemed to help people read and understand the information and the Q&A approach helped in navigating through the mapping task.
- On the negative side, casual web users found it very restrictive to be limited to a narrow range of mapping terminology, e.g. "corresponds" and "similar to" when answering mapping questions. In addition, some users were unclear about the benefit in engaging in the mapping task.

In our current experiment (finishing May 2008) we are focusing on whether it is valuable to embed the mapping process within the user environment, designing a user-centric mapping process, and addressing the negative concerns garnered from the previous experiment by allowing the user to be more expressive by allowing them to 'tag' the mapping relation. Our hypothesis is *the mapping task can be simplified and become unintrusive by embedding the mapping process within the user environment and by using a 'tagging' approach paradigm.* By using the power of "Web 2.0" through a Firefox extension within our new 'tagging' prototype, we aim to engage the user and display matching collections at appropriate times within their own work environment, see [18] for more details. We use online questionnaires, interviews, and a log of each user's actions to evaluate the impact of the 'tagging' prototype. In particular through the use of our implementation over the coming months we aim to investigate whether casual web users will be able to use tagging to turn matches into expressive mappings in a straightforward, practical and natural manner. We will also investigate whether embedding the mapping interface inside a browser extension will allow the mapping process to take place over time within a casual web persons work environment in an unobtrusive, sensible, and normal way.

Our next experiment is going to be investigating the effects of collaborative knowledge sharing via groups (from June 2008 to October 2008). A key factor of this experiment will be to investigate whether the knowledge shared by ontological aware users can be beneficial to casual web users. Another issue is which users are the most beneficial for validating each matching pair question, i.e. will users with musical background be better than ontological aware user in relation to music based matching's. Another aspect is whether it is better to divide the users into different groups and what type of characteristics determine which group(s) each user is put into. A final feature of investigation will be if categorising the 'user-defined' tags collaboratively, whether globally or within groups, is of any help to users.

In our first experiment we looked at displaying ontological information with natural language and to analysis the difference between a graph types interfaces. The natural language used was far from ideal and although the results showed the benefits of using this type of interface we intend to revisit this objective to make tests with different visual types of interfaces (from November 2008 to April 2009). We intend to investigate different visual types of natural language, such as representing the concepts with shallow text generation as discussed in [19], and other different visual types. Although finding a one-size-fits-all interface for everybody may be impossible, we believe there can be a basic interface which can be suitable to most

casual web users while also allowing these users to change the visualization to their needs, i.e. some may rather graph based to natural language and vice versa, etc...

A key problem is reducing mapping tasks to being *unintrusive* to casual web user. We intend to explore in what context and at what time should a mapping task be performed by the user (from April 2009 to June 2009). Finally the thesis write up will occur from July 2009 till November 2009.

References

1. Shvaiko P. and Euzenat J., "A Survey of Schema-based Matching Approaches", Journal of Data Semantics, Long version in DIT Technical Report DIT-04-87, 2004.
2. Noy N., "Semantic Integration: A Survey of Ontology-Based Approaches", Special Issue on Semantic Integration, SIGMOD Record, Volume 33, Issue 4, pages 65-70, 2004.
3. Aumüller D., Do H., Massmann S., Rahm E., "Schema and Ontology Matching with COMA++", Proc. SIGMOD 2005 (Software Demonstration), Baltimore, June 2005.
4. Zhdanova A., "Towards Community-Driven Ontology Matching", Proceedings of K-CAP, Banff, Canada, 2005.
5. Jameson A., "Usability and the Semantic Web", 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Springer LNCS Vol. 4011
6. Falconer S. M., Noy N. F., Storey M., "Towards understanding the needs of cognitive support for ontology mapping", ISWC 2006, 2006, Springer LCNS Vol. 4273
7. Falconer S. M., Storey M., "A Cognitive Support Framework for Ontology Mapping", In Proceedings of ISWC+ASWC, pages 114-127, 2007
8. Zhdanova A., Shvaiko P., "Community-driven ontology matching", In Proceedings of ESWC, pages 34-49, 2006
9. Shvaiko P., Giunchiglia F., Pinheiro da Silva P., McGuinness D., "Web Explanations for Semantic Heterogeneity Discovery", In Proceedings of ESWC, pages 303-317, 2005
10. Mocan A., Cimpian E., Kerrigan M., "Formal Model for Ontology Mapping Creation", In Proceedings of ISWC, pages 459-472, 2006
11. Morishima A., Okawara T., "SMART: a tool for semantic-driven creation of complex XML mappings", SIGMOD '05, Baltimore, MD, USA, ACM Press, NY, 2005
12. Yan, B., Frank, M., Szekely, P., "WebScripter: Grass-roots Ontology Alignment via End-User Report Creation.", ISWC 2003, LNCS, vol. 2870, Springer, Heidelberg, 2003
13. Raghavan, A., Rangarajan, D., Shen, R., Goncalves, "Schema Mapper: A Visualization Tool for DL Integration.", In Proceedings JCDL2005, Denver, p. 414, 2005.
14. Lanzenberger M., Sampson J., "Alviz – a tool for visual ontology alignment", In Proceedings of the 10th International Conference Information Visualization, 2006
15. Bernstein A., Kaufmann E., "GINO – A Guided Input Natural Language Ontology Editor", ISWC, 2006
16. OntoViz, http://protege.stanford.edu/plugins/ontoviz/
17. Conroy C., O'Sullivan D., Lewis D., "A Tagging Approach to Ontology Mapping", The 2nd International Workshop on Ontology Mapping, ISWC, Korea, 2007 (poster)
18. Conroy C., O'Sullivan D., Lewis D., "Ontology Mapping Through Tagging", The International Workshop on Ontology Alignment and Visualization, CISIS, Barcelona, 2008
19. Kalyanpur A., Hashmi N., Golbeck J., Parsia B., "Lifecycle of a Casual Web Ontology Development Process", Proceedings of the WWW2004, May 18, 2004