A 'Tagging' Approach to Ontology Mapping

Colm Conroy[1], Declan O'Sullivan[1], Dave Lewis[1]

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

Abstract. 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. To address these three problems, we believe that the interfaces to ontology management tools will need to be engineered in such a way as they disappear into the background from the ordinary person's perspective. The majority of the state of the art approaches to ontology mapping relies on the user being ontologically aware. In contrast, this paper reports upon an approach to use 'tagging' as a means of ontology mapping to support ordinary people.

1 Introduction

The promise of ontologies, which is a core technology of the semantic web, is in the sharing of an understanding of a domain that can be communicated between people and application systems [fensel 2003]. Although the semantic web approach is establishing itself in certain domains, as yet no viral uptake within the mainstream internet environment has occurred [berners-lee 2006]. Instead recently collaborative tagging schemes (referred to as "folksonomies") have started to emerge within the mainstream internet community. With collaborative tagging, people publicly annotate resources with keywords that describe those resources (called tags). In contrast, ontologies provide a means to associate terms with concepts and relate concepts together, thereby providing a means to discover ambiguity, navigate over relationships, and cope with synonyms and homonyms. 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. To address this problem, we believe that the interfaces to ontology management tools will need to be engineered in such a way as they disappear into the background from the ordinary person's perspective. 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 [shvaiko 2004]. The research challenge lies in how to derive ontology mappings from the candidate matches. Fully automatic derivation of mappings is considered impossible as yet [noy 2004], and the majority of state of the art tools in the ontology mapping area [kalaynpur 2004], [aumuller 2005] and the community ontology creation area [zhdanova 2005] 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. Two key objectives we address is making ontology mapping **natural** and **transparent** as possible for the casual web user, which can be stated as:

- To represent match information in a clear and intuitive way for the user;
- To make the turning of the match information into a mapping expression intuitive for the user; which is also unobtrusive and takes place over time.

2 Initial Experiment & Findings

In our initial experiment undertaken in early 2007 we aimed to make the semantic mapping process as user friendly as possible so that ordinary people could undertake mapping of semantic models without having to be ontologically aware. With current ontology mapping systems, the ontologies are explicit, the mappings are difficult to read, hard to navigate, and are prone to data overload with too much clutter. In our solution we chose to take a Q&A approach to reduce the complexity involved. Using this approach we reduce the mapping process into piece wise comparisons. We used a fix set of question templates that use the words "similar" and "correspond" in comparing the terms of the match. We used a natural language approach to display each concept as it was our contention that it should help a non technical user in understanding the information better, since ordinary people are more accustomed to reading sentences rather than looking at graphical structures. For each concept its parents, properties and siblings are displayed. We categorised each match within one of the following groups: **valid/mostly/skipped/to-be-validated/rejected**. The main purpose of our initial experiment was to test the usability of our prototype natural language mapping tool (NL) and contrast it with a current state of the art ontology mapping tool, which we choose as the graph type mapping tool COMA++. We split the user test group into three distinct groups: ontology aware, technology aware and non-technical aware. Two different domains were used for the ontologies to help keep the experiment grounded which were music[1] and university[2]. 8 different users were asked to use both tools in different orders which allowed the experiment to be done in 4 different orders, twice per each group. For the experiment we allowed the users to use each tool and after they were finished we had a short interview with them where we asked some questions to gauge the usability of each tool. Some conclusions that we drew from this experiment were:

- Natural language can be used for representing ontological terms to ordinary people as it seems to help people read and understand the information.
- The question & answer approach helped user to navigate through the process of validating the mappings.
- Results showed that ordinary users can validate mappings effectively and efficiently even compared to ontology aware users.
- Some were still confused as to the task of the experiment which showed there is still a need to reduce the complexity even more.
- When answering mapping questions with the NL tool a lot of the wrong answers came from people getting confused with the mean of some questions, i.e. it may not be the same but it corresponds, and on the use of ontological terms like Thing and Agent.

1 http://maciej.janik/test and http://www.kanzaki.com/ns/music
2 http://annotation.semanticweb.org/iswc/iswc.owl and http://swrc.ontoware.org/ontology

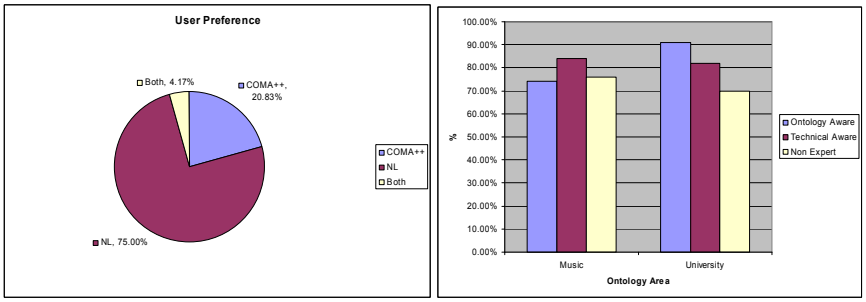

Fig 1: *Experiment results* showed that a two thirds majority preferred the NL tool (left). NL prototype overall accuracy across the 3 groups were very comparative (right).

3 Turning matches into mappings using a tagging approach

One of the main outcomes of our initial experiment was the realization that ordinary users found it very restrictive to be limited to a narrow range of mapping terminology, e.g. “corresponds” and “similar to”. It is our contention that ordinary people do not like the ‘yes’ or ‘no’ strict approach but rather they would prefer the ability to ‘tag’ things with their own terms to give them their own meaning. Sites like flickr[3] have become popular because they allow the user to become more expressive by ‘tagging’ items with their own terms, it is our opinion that by allowing users to map ontologies in this fashion would allow the mapping process to become easier for the user and would lead to expressive mappings. In our current work we are allowing the user to 'tag' the mapping relationship with multiple user defined 'tags'. Each tag is then annotated based on the categories in Fig 2. A decision is then made as to which category the mapping should belong to overall. The decision making rules are configurable, for example majority rules. The ‘corresponds’ category refers to when a mapping relation is tagged with unknown ‘tags’ and the matching cannot be assigned to one of the other categories. 'Validation' questions will be used to sort the ‘correspond tags’ into other categories, based on other users’ experiences. In our current implementation our mapping tool is a Firefox browser extension[4] which will be used to display the 'tag' question by making a transparent screen over the current page the user is browsing. The information source which we plan to use in our next experiment will be information from RSS feeds and podcast feeds. Each user will have ontologies representing their interests, e.g. music, sport etc. The system will then be used by the user to map between their interests and the RSS feeds. For the experimentation we will use the same groups as in the initial experiment with different and more people tested. The experiment will occur over a period of a month with feedback captured at random times. After the experiment is concluded we will have the users fill in an online questionnaire form to gauge the user’s reaction/performance.

3 http://www.flickr.com/
4 https://addons.mozilla.org/firefox/extensions/

- Equivalent
  - The same
  - Subclass
- Equivalent Sometimes
  - Super class
  - One of
  - Union
  - Intersection
- Different
  - Different from
  - Complement of
  - Disjoint
- Corresponds

Fig 2: *The 'Tag' Categories* (left): The 4 categories have corresponding concept relationships as subcategories. *Conceptual 'tag' interface* (right): It uses same NL to represent the ontological terms as in initial experiment.

4 Design of Mapping Process to make more transparent

In the design of our tagging approach (Fig 3) we aim to make the mapping process more transparent and occur over multiple sessions so the user will see mapping not as inconvenient work but more as being part of there daily life.

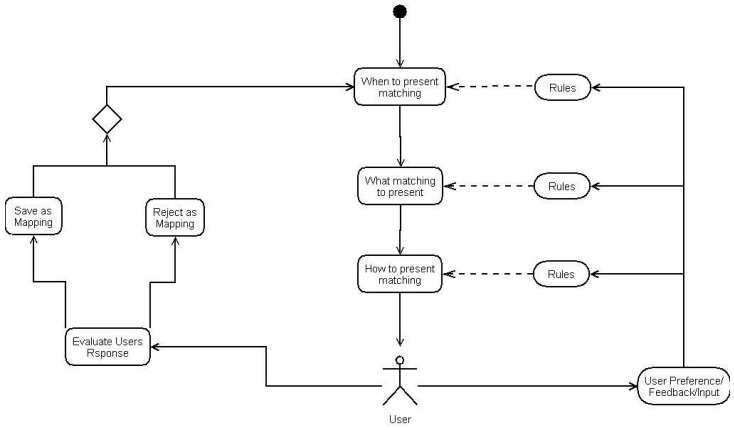

Fig 3: *User Mapping Interaction*

- **When to present matching pair:** Involves calculating the time to present a matching pair for the user, making the process occur over multiple sessions.
  - ”Just in time”, e.g. if the user is submitting a query but needs to map their own ontology to another one for the query to work.
  - When the user is perceived as being idle.
  - After a specific time period, e.g. every hour when available.
  - Threshold on number of matching pairs asked, e.g. 3 per hour.

- **What matching pair to present:** This step involves deciding which matching present is the next one to present to the user
    - Priority based on expected need (based on user interaction patterns).
    - Specific matching API, matching strategy and threshold percentage.
    - A user specified matching, e.g. the matching API misses a matching pair which the user implicitly/explicitly implies is a mapping.
- **How to present the matching pairs**: This step involves deciding the best way to display the matching pair to the user.
    - Visual type, i.e. natural language, graphically, etc...
    - Different forms of representation, i.e. NL bullet point or paragraph..
    - Filtering away information whether necessary or unnecessary.

In our next experiment we plan to address the first two points.

5 Final Remarks

It is hoped through our proposed tagging experiments we will be able to show that the mapping process can be undertaken over multiple sessions rather than one sit down session and that this method will be demonstrably better for the user in making the process as transparent as possible. We are also hoping to show that through our Firefox browser extension that the combination extension of the natural language interface with tagging will enable users to turn matches into expressive mappings in a natural manner. In summary we hope to show that a combination of natural language information presentation, 'tagging' for mapping expressivity, and process to manage the mapping generation over time, will reduce the complexity of semantic mapping, help the user be more expressive in the mappings generated, which will lead to higher user engagement in undertaking mapping and clearer understanding by the user of the benefits for the user in terms of greater access to internet resources

6 References

[aumuller 2005] Aumüller D., Do H., Massmann S., Rahm E., "Schema and Ontology Matching with COMA++", Proc. SIGMOD 2005 (Software Demonstration), Baltimore, June 2005.

[berners-lee 2006] Berners-Lee T., Shadbolt N., Hall W., The Semantic Web Revisited. IEEE Intelligent Systems Vol. 21, Issue 3, pp. 96-101, 2006.

[fensel 2003] Fensel D., "Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce", 2nd edition, Springer-Verlag, Berlin, 2003.

[kalyanpur 2004] Kalyanpur A., Sirin E., Parsia B., Hendler J., "Hypermedia inspired ontology engineering environment: Swoop", Proceedings of 3rd International Semantic Web Conference (ISWC-2004), Japan (Poster).

[noy 2004] Noy N., "Semantic Integration: A Survey of Ontology-Based Approaches", Special Issue on Semantic Integration, SIGMOD Record, Volume 33, Issue 4, pages 65-70, December 2004.

[shvaiko 2004] Shvaiko P. and Euzenat J., "A Survey of Schema-based Matching Approaches", Submitted to Journal of Data Semantics, Long version in DIT Technical Report DIT-04-87, 2004.

[zhadonova 2005] Zhdanova A., Towards Community-Driven Ontology Matching, Proceedings of K-CAP 2005, Banff, Canada, October 2-5, 2005.