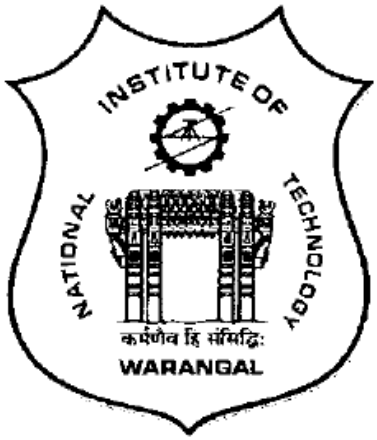

REVISITING ONTOLOGY ALIGNMENT IN WEB 2.0

Asankhaya Sharma
asankhaya@yahoo.com

Technical Report
July 2009

Revisiting Ontology Alignment in Web 2.0

Abstract – *In this paper we look at the some of the new challenges and issues for ontology matching and alignment with the rise and popularity of web 2.0. Heterogeneous data sources, composite applications and mashups create new challenges and opportunities for applying semantic technologies. Web sites have increasingly become more interactive and responsive with use of technologies like HTML5, JS and Ajax. In this changing environment increasingly developers are using non-relational data models (NoSQL, JSON etc.) to store and retrieve data. Ontology matching and schema integration in a schema free data model remains a challenge. We review some of the interesting aspects of the web application development and enumerate some new challenges for ontology alignment. Future web sites will look more like applications and data used by diverse applications need to be described and understood to support interoperability. Thus ontology alignment has a vital role to play for data interoperability in the web 2.0.*

1. Introduction

Ontologies have been used to describe a concept to help in integration of heterogeneous information sources. In the context of the semantic web, ontologies are used to aid in describing entities and concepts. Since the same concept can be described in multiple ways it leads to the semantic heterogeneity problem. Ontology matching and alignment is meant to alleviate this heterogeneity in information sources but matching and combining different related concepts. Over the last few years several tools and systems [1-4] have been built which implement different algorithms for ontology matching. Traditionally these systems are used for information integration from different sources. The structural differences between ontologies from varied sources can be handled by alignment and matching.

Ontology matching can sometimes be done statically based on schema definitions. But dynamic and responsive web applications use data in a manner which forces us to consider dynamic ontology matching. Alignments can also be elaborated manual guided by cognitive factors [5] based on visualizations. Ontology matching helps in reducing heterogeneity in data sources of information and helps in creating a consistent system free from redundancies. This has wide applications in data warehousing, data management and the semantic web. Large ontologies require careful algorithms [8] that consider performance and efficiency to enable scalable processing of alignments for ontologies.

Large body of research and investigation into the problem of ontology alignment needs to be revisited in the context of web 2.0. With the changing web application and data requirements ontology matching faces new challenges. Web sites are no longer static web pages serving content and images. They are increasing responsive, adaptive and dynamic. In these conditions existing approaches for ontology alignment need to be modified and we need new fresh perspectives for solving this problem for the future. The use of newer data formats many of which are schema less makes it harder to use existing matching techniques when dealing with current web applications. In this paper we elaborate some of the key challenges and issues for ontology alignment facing application developers today.

The paper is organized as follows, in section 2 we look at the use of ontologies in the context of web 2.0, in section 3 we elaborate on the new challenges and issues web 2.0 brings for ontology alignment, in section 4 we describe some related work and finally we conclude in section 5 with some pointers for future.

2. Ontologies and Web 2.0

The promise of ontologies was nothing less than a fully functional semantic web. However as researchers were focused on adding meaning to the web, existing technologies for web application development underwent a major transition. The move from static HTML pages to Ajax driven JavaScript websites marked the advent of web 2.0. Interactive and responsive design is the hallmark of a good web application today. With static HTML pages it was easy to associate ontologies and reason about concepts. But in many web applications today HTML is generated dynamically on the fly based on events fired by user interaction. In such a scenario it becomes hard to associate ontologies with concepts.

Data management is no longer done using explicit schema based relational models. Web developers use document object models like JSON to manipulate and store their databases. The success of the whole NoSQL movement rests on the ease of use of such schema less representation of data. Ontology matching becomes much harder with no fixed schema, it may be necessary to give up on the deterministic alignment and consider only dynamic approximate matching algorithms. Semantic web services based on fixed and static ontologies may no longer be possible. In such a context it becomes important to reassess the use of ontology alignment and matching for information integration.

As an example consider social networks like Facebook, Twitter and LinkedIn. They use different representations for the same concept of a “Person” and yet a walled garden approach prevents one from building any useful application that can integrate information about the “Person” ontology from these diverse social networks. Increasingly questions about data integration and ontology matching are becoming economical and business oriented. Companies like Facebook are taking steps to ensure that the data within their networks cannot be exported and used in an easy format by others.

With these new social-economic trends and the changing technology landscape of web applications it may so happen that we have great ontology matching and alignment algorithms but not get access to the data to build semantic web services. Web 2.0 is also heavily driven by the needs of users; the economic incentives for social networks prevent them from releasing their data to others. On the other hand research and development on semantic web services has not yet yielded a killer application which would force these companies to rethink their strategy.

The growing number of these new web 2.0 companies and their ever increasing users are getting accustomed to walled gardens. Soon the inability to interoperate between different web applications may be mistaken by for feature of the platform. This set of challenges and problems also bring new opportunities for people working on ontology matching and alignment. If we can design systems that can integrate information from various sources in the current environment we can hope to make a bigger impact with real users.

3. New Challenges and Issues

The rise of Web 2.0 brings up an interesting set of challenges and issues for ontology alignment. We enumerate a few of them below based on our studies.

- The ontology matching and alignment systems need to work on a larger scale. With the number of uses on Facebook reaching about 1 billion, soon better and fast graph algorithms would be needed to analyze and integrate such a large dataset. Existing ontology matching tools have not demonstrated that they can be applied at such a large scale.
- Use of schema less data models necessitates techniques which are independent of underlying schema. Existing ontology matching algorithms need to be modified to discover schema during the matching either partly or entirely. Most current algorithms assume existence of diverse but fixed structural schemas to do matching and alignment with.
- Dynamic and constantly changing nature of web 2.0 applications means that there is inherit uncertainty in the ontology matching process. Probabilistic methods based on approximate sampling should be explored to overcome this challenge. Existing approaches that are based on deterministic assumptions will not perform well in situations that are non-deterministic due to the presence of large number of user driven events and responses.
- Because of the user driven nature of web 2.0 applications, approaches based on visualization can help in alignment. It remains a challenge to come up with succinct representation of large graph modes that come up in ontology matching often. A hierarchical model of ontologies based on different levels can help in alignment with only a portion of the entire graph been visualized at any given time.
- Crowd sourcing can be used to aid in the process of matching and alignment. The task of the ontology alignment can be divided into several smaller micro tasks which are distributed over the internet using a framework like Amazon Mechanical Turk. These tasks can be done by solvers around the globe and then integrated to give the final result. Such an approach will make use of the possibilities created by web 2.0 itself (crowd sourcing) to solve the ontology alignment problem.
- In order for ontology alignment to be useful for end users, there is a need to come up with some killer applications that use the results of the ontology matching. It can help show the benefits of ontology alignment to end users and force the networks to make some changes. By giving incentive to the end users we can persuade the networks to support better mechanisms for ontology alignment.

The above list is by no means exhaustive but covers most of the interesting challenges for ontology alignment in the light of recent developments in web 2.0. For the field of ontology matching to be relevant in future it has to reinvent itself and abandon the grand goals of solving the semantic web problem. It is much better to focus on problems facing real users on web today (like data interoperability) and develop some nice applications which show the benefits of such information integration. Existing methods and techniques which were focused on static algorithms guided by schema and structural assumptions may need to be modified and adapted for web 2.0.

4. Related Work

In a prior work [1] we presented a schema based ontology matching algorithm for alignment. The algorithm was based on ranking the nodes of various structural schema elements and then calculating a similarity measure between various nodes. The ranking of nodes was then used to discover similar ontological elements. With the increase in representation free data models such algorithms cannot be expected to perform well in practice. The rise of new kind of web 2.0 applications has made the entire corpus based schema mapping [9] methods irrelevant for ontology matching. However methods that are based on more general data representation like graphs [2] are still quite useful. Graph based representation of data sets can handle schema less data models and still give useful results for ontology alignment. Many of these methods have to be adopted and extended to work for large data sets based on millions (even billions) of nodes.

Some early work already is showing signs of tackling some of the challenges we mentioned in section 3. The user of approximate [6] but semantic preserving matching can help to handle some of the issues we listed earlier. Researchers are already taking advantage of the web itself to create new and innovative matching measures. In [7] the authors use the google distance to support ontology matching and alignment. There is also some interest in the ability to make ontology scale for larger data sets. Exploiting the underlying multi core and parallel architecture of some of the new processors in market for ontology alignment still remains to be seen. However there are promising initial results [8] for divide and conquer type approaches that deal with ontology alignment.

Most related work currently doesn't take into account the recent challenges in the changing landscape of web 2.0. There is a need to refocus the problem of ontology alignment and develop solutions that help end users. In this respect we hope this paper can bring some fresh perspectives for the people building ontology matching tools. Part of the new challenge is the ontology matching is no longer only a technological problem; there are interesting socio-economic trends in play. We believe this presents with new opportunities for someone who can solve these challenges. Going forward with the rise of cloud based computing data interoperability will be a much larger problem. If ontology matching can help in solving data interoperability it can be a impactful tool in web 2.0.

5. Conclusions

In this paper we reviewed from the social-economic and technological trends coming of age with web 2.0. We explored how these trends affect the field of ontology matching and alignment. We presented a set of the challenges for ontology matching with the rise of web 2.0 based applications. In the next few years we expect more focus on these challenges and issues by people working on integration of information from diverse sources.

Some of the challenges listed are hard and difficult but even small improvements can help bring the benefits of ontology matching and alignment to end users. For that reason we suggest that we should focus on identifying a niche and building applications that demonstrate the real benefit of ontology matching.

References

[1] Asankhaya Sharma and DVLN Somayajulu. Ontology Matching and Schema Integration using Node Ranking. In SWWS 2006.

[2] Asankhaya Sharma. Ontology Matching using Weighted Graphs. In ICDIM 2006.

[3] J. Euzenat and P. Shvaiko. Ontology matching. Springer, 2007.

[4] C. Domshlak, A. Gal, and H. Roitman. Rank aggregation for automatic schema matching. In IEEE Transactions on Knowledge and Data Engineering, 2007.

[5] S. Falconer and M. Storey. A cognitive support framework for ontology mapping. In Proceedings of ISWC/ASWC, 2007.

[6] Giunchiglia, F. McNeill, M. Yatskevich, J. Pane, P. Besana, and P. Shvaiko. Approximate structure preserving semantic matching. In Proceedings of ODBASE, 2008.

[7] R. Gligorov, Z. Aleksovski, W. ten Kate, and F. van Harmelen. Using google distance to weight approximate ontology matches. In Proceedings of WWW, 2007.

[8] W. Hu, Y. Qu, and G. Cheng. Matching large ontologies: A divide-and-conquer approach. In Data and Knowledge Engineering, 2008.

[9] J. Madhavan, P. Bernstein, A. Doan, and A. Halevy. Corpus-based schema matching. In Proceedings of ICDE, 2005.