Bootstrapping Domain Ontologies for Rapid Semantic Annotation of User-Friendly Semantic Web Content

Yuri A. Tijerino, President & CTO (yuri@ontolligence.com)
Ontolligence Corp.

1. Introduction

In attempting to develop tools, mechanisms and content for the Semantic Web we have to keep in mind that the requirement for machine understandability [2] is not a constraint and should not preclude usability requirements by end users. In essence, the Semantic Web should remain as distributed, self-evolving, ad-hoc, and easy to build, navigate and maintain as the World Wide Web (Web) is today. However, Semantic Web ontology languages such as XML, RDF, RDF-S, DAML+OIL, OWL, OWL-S and others require specialized expertise to understand and use. This raises three interesting problems: 1) How to bootstrap enough domain ontologies that are capable of representing the intricacies of Web information in the form of knowledge in the Semantic Web, 2) how to bootstrap enough Semantic Web content by using those ontologies to annotate Web content in a resilient manner, and 3) how to provide easy to use Semantic Web applications that are easy to use and understand by the average Web user.

Ontolligence Corp., attempts to address these problems head-on. We create tools, techniques and processes that make it possible to quickly create Semantic Web ontologies from sample ontologies and from domain specific sample Web pages. At the same time, we create automated and semi-automated tools that quickly annotate Web pages with Semantic Web ontologies in a manner that makes the pages understandable by computational mechanisms. Furthermore, we implement applications that enable both users and computational mechanisms to interactively collaborate in performing problem solving tasks that reap the benefits of Semantic Web content.

2. Technical Objectives

The technical objectives of Ontolligence Corp., a spin off of Brigham Young University's Data Extraction Group [4], are threefold: First, we are investigating how to technologically expand our existing mechanisms [3] to incorporate support for OWL data-extraction ontologies in a commercially viable manner. Currently we support conversion of DAML ontologies into OSM data-extraction ontologies [6]. The technical objective here is to make it easier for companies and organizations to make a transition to the Semantic Web while supporting the OWL standard.

Second, we are investigating the construction of a mechanism that supports automatic or semi-automatic generation of data-extraction ontologies in the OWL standard. Currently this is a tedious manual process that does not fit

well in commercial applications due to a high cost to benefit ratio. We have already experimented with the automated generation of OSM data-extraction ontologies [5] and are working on new techniques which takes advantage of structured data found in the Web such as tables [8, 7], to make the process of ontology generation more cost effective. The technical objective here is to make it possible for companies and organizations to create their own internal conceptual domain models (i.e. ontologies) in a timely manner without the need for specialized and costly ontology engineering expertise, which is one of the main factors preventing industry-wide investment in this area.

Third, we are developing user friendly mechanisms that allow Semantic Web users to interact with computational mechanisms to browse, search, reason and perform problem solving on the Semantic Web. We have obtained positive experimental data that indicates that it is possible to allow agents to collaborate with other agents without requiring them to share the same ontology [1]. The technical objective here is to allow humans to communicate seamlessly with agents through simple, ontology-generated Web forms to specify problems, resolve conflicts and clarify requests [9].

We are interested in discussing with, learning from and collaborating with other researchers and partners about current cutting-edge research, techniques, approaches and mechanisms that addess these three particular areas.

References

1. Al-Muhammed, M., 'Dynamic Matchmaking between Messages and Services in Multi-Agent Systems'. Technical report, Brigham Young University, Provo, Utah.
2. Berners-Lee, T., J. Hendler, and O. Lassila: 2001, 'The Semantic Web'. *Scientific American* **36**(25).
3. Chartrand, T.: 2003, 'Ontology-Based Extraction of RDF Data from the World Wide Web'. Master's thesis, Brigham Young University, Provo, Utah.
4. DEG: 2004, 'Homepage for BYU Data Extraction Research Group'. URL: http://www.deg.byu.edu/.
5. Ding, Y.: 2003, 'Semiautomatic Generation of Relilient Data-Extraction Ontologies'. Master's thesis, Brigham Young University, Provo, Utah.
6. Embley, D., D. Campbell, Y. Jiang, Y.-K. Ng, R. Smith, S. Liddle, and D. Quass: 1998, 'A Conceptual-Modeling Approach to Extracting Data from the Web'. In: *Proceedings of the 17th International Conference on Conceptual Modeling (ER'98)*. Singapore, pp. 78–91.
7. Tijerino, Y., D. Embley, D. Lonsdale, and G. Nagy, 'Ontology Generation from Tables'. *Journal of World Wide Web Internet and Web Information Systems*. Submitted.
8. Tijerino, Y., D. Embley, D. Lonsdale, and G. Nagy: 2003, 'Ontology Generation from Tables'. In: *Proceedings of the 4th International Conference on Web Information Systems Engineering*. Rome, Italy. 242–249.
9. Tijerino, Y. and M. A.-M. Embley: 2004, 'Toward a Flexible Human-Agent Collaboration Framework with Mediating Domain Ontologies for the Semantic Web'. In: *Proceedings of ISWC'04 Workshop on Meaning Coordination and Negotiation*. Hiroshima, Japan.