

The Role of Ontologies in Linked Data, Big Data and Semantic Web Applications

Mike Bennett* Kenneth Baclawski†

April 15, 2018

Abstract

Since the beginnings of the Semantic Web, ontologies have played key roles in the design and deployment of new semantic technologies. Yet over the years, the level of collaboration between the Semantic Web and Applied Ontology communities has been much less than expected. Within Big Data applications, ontologies appear to have had little impact. These communities, along with the Linked Data community, all share the need for a common semantic understanding and a formal representation of the domains being studied, but they have taken very different approaches to deal with the challenges of large scale applications and linking of vast heterogeneous data. Because of this situation, the Ontology Summit 2014 focused on building bridges between these four communities. It was felt that identifying and overcoming ontology engineering bottlenecks is critical for all of these communities. This special issue is an effort to continue the process that began in 2014. The papers in this issue are concerned with the various aspects of the barriers identified at the Ontology Summit, and propose approaches for addressing them.

Keywords: semantic web, linked data, big data

The Semantic Web and Linked Data communities acknowledge the role that ontologies play in designing and employing their respective technologies. Yet, collaboration between these communities and the Applied Ontology community has been much less than expected. A more striking situation is that ontologies appear to have had little impact in Big Data applications, in spite of the clear need for better understanding of the meaning of the data and the results of data mining. In an attempt to address these concerns, the Ontology Summit 2014 focused on building bridges between the Semantic Web, Linked Data, Big Data, and Applied Ontology communities. While these communities all share the need for a common semantic understanding and a formal representation of the domains being studied, they have taken very different approaches to deal with the challenges of large scale applications and linking of vast heterogeneous data. The Ontology Summit 2014 brought together representatives from all of these communities to better understand the barriers and challenges that hinder the use

*Hypercube Ltd., London UK

†Northeastern University, Boston, MA USA; corresponding author email: kenbaclawski@gmail.com

and reuse of ontologies by the Semantic Web, Linked Data and Big Data communities. Fig. 1 is a graphic depiction of the main barriers and challenges that were identified at the summit, expressed as gaps between and among the approaches used in these communities. The summit sponsored a wide variety of events, including a four-month online discussion forum, four conference tracks, six hackathons, an online community library and an ontology repository. The summit culminated in a two-day symposium, and issued a communiqué that summarized the results of the summit (Grüninger et al., 2014).

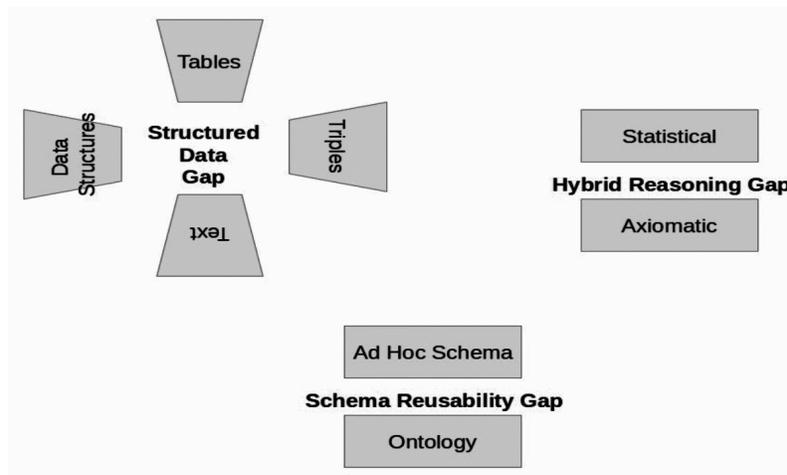


Figure 1: The graphic depiction of the gaps addressed by the Ontology Summit 2014

The four conference tracks of the Ontology Summit 2014 each focused on different aspects of the summit topic:

1. Use and Reuse of Semantic Content – Experiences in Knowledge Sharing: Lessons from research and experience in Big Data, Linked Data and Semantic Web Applications
2. Making use of Ontologies: Tools, Services, and Techniques
3. Overcoming Ontology Engineering Bottlenecks
4. Tackling the Variety Problem in Big Data

The summit hackathons explored solutions to problems that span two or three of the Data Modeling, Semantic Web and Application Ontology domains and that address at least one of the gaps in Fig. 1. The hackathons engaged in both software coding and data preparation to provide cross-domain experience for the hackathon teams members. The six hackathons represented a wide variety of teams and topics within the overall theme of the summit.

1. Reference data for Anime and Manga: Semantic Linking and Publishing of Diverse Data-Sets, led by Victor Agroskin

2. Ontology Design Patterns and Semantic Abstractions in Ontology Integration, led by Mike Bennett and Gary Berg-Cross
3. Optimized SPARQL performance management via native API, led by Victor Chernov
4. Ontohub consolidation, led by Till Mossakowski and Oliver Kutz
5. Semantic Annotation of the Ontolog Community Environment (SAOCE), led by Kenneth Baclawski
6. An ontological catalogue of ontology and metadata vocabulary characteristics relevant to suitability for semantic web and big data applications, led by Amanda Vizedom

To follow up on the success of Ontology Summit 2014 and to maintain its momentum, the Semantic Web for Applied Ontology (SWAO) special interest group was founded within the International Association for Ontology and its Applications (IAOA). This special issue is one of the initiatives of SWAO. The topics of the special issue expand on the topics of the Ontology Summit 2014 to include:

- The role that ontologies play (or can play) in Linked Data, Big Data and Semantic Web Applications
- Engineering of ontologies to address integration and domain-specific modeling concerns
- Sharing and reuse of ontologies within and across application or domain areas
- Tooling and techniques in support of ontology development for Linked Data, Big Data and Semantic Web Applications (provided that a contribution to the practice of ontological analysis and conceptual modeling is clearly established)

The papers that have been selected for this special issue include papers dealing with the main tracks, and reports on the progress of subsequent work on three of the six hackathon projects.

The lead article is “Choosing Ontologies for Reuse” by Megan Katsumi and Michael Grüninger. This article deals with the challenge involved in choosing among the different ontologies that are candidates for a particular application. Katsumi and Grüninger address this challenge by introducing a novel preference relationship between ontologies. They also develop a procedure for determining whether this relationship holds. Their definition and procedure allows developers to make well-founded decisions concerning whether to reuse ontologies for their application. In terms of Fig. 1, this paper develops a technique for helping to close the schema reusability gap. Their technique is a significant advance in ontology development that addresses four of the seven key limitations for ontology reuse that were identified by the Ontology Summit 2014:

- *Mismatches and Misunderstandings* when attempting to reuse an ontology that is not, in fact, suitable.

- *Finding Mr. Right Ontology* can be a challenge if one can only use keyword matching.
- *This Ontology Doesn't Fit...* occurs when modifications are required to make an existing ontology usable.
- *Just Do It Yourself* may be necessary when existing ontologies are too difficult to reuse.

Mike Bennett has provided another perspective on the ontology reusability problem in his article, "Framing Ontology Distinctions: An Exploration." In this article he points out that there are some fundamental distinctions between ontologies that are easily missed by ontology developers, yet can have a significant effect on reusability. Some ontologies are fundamentally describing phenomena in the real or an imagined world, while others are more concerned with logical and computational principles. What can make this distinction subtle is that an ontology of one kind can appear to be essentially identical to an ontology of the other kind. This is important for the Semantic Web because the use of the Semantic Web languages tends to predispose ontology developers toward the latter kind of ontology, which can limit their reusability. To explain the distinction between ontologies, Mike Bennett has refined the well known "Semiotic Triangle" (Ogden & Richards, 1923; Ullmann, 1972) to a "Semiotic Rhombus," by splitting the notion of a "concept" into the intension and extension of the concept. The framework developed in this article could help with the task of communicating an appreciation of conceptual semantics, classification theory and terminology among ontology developers and users that have largely been dominated by technological considerations.

The summit was particularly concerned that ontologies appear to have had little impact in Big Data applications. To address this, Track D focused on Big Data (Baclawski and Thessen, 2014). The specific aspect of Big Data that was recognized as having the greatest potential to benefit from a greater use of ontologies is Variety aspect. The article by Baclawski, Chan, Gawlick, Ghoneimy, Gross, Liu and Zhang, "Framework for Ontology-Driven Decision Making," takes as its premise that the process employed in Big Data makes use of statistical techniques for discovery and hypothesis testing. Consequently, the process may be regarded as a form of scientific discovery. A fundamental feature of scientific discovery is that it is an iterative loop in which one observes nature, formulates hypotheses to account for the observations and then tests them. The processes in Big Data automate the scientific process. The article by Baclawski and his colleagues introduces a formal ontology-based framework for this process to develop a bridge between Big Data processes and the Semantic Web, Linked Data and Applied Ontology. Their framework addresses both the Structured Data Gap and the Hybrid Reasoning Gap in Fig. 1.

The number of ontologies and ontology languages have been increasing. As a result, finding ontologies that can be reused for an application domain can itself be an obstacle to reuse. Having discovered that an ontology exists that may be reusable, one may also be faced with still other obstacles. The ontology may no longer be available or may be in a different language. Navigating through a large ontology can also be daunting. Search engines alone are not a solution to these obstacles to reuse. The

article by Codescu, Kuksa, Kutz, Mossakowski and Neuhaus, “Ontohub: A semantic repository for heterogeneous ontologies” reports on recent progress in the development of Ontohub, which was the subject of one of the hackathons at the Ontology Summit 2014. The Ontohub repository supports a wide variety of languages for defining ontologies, thereby addressing many of the most profound obstacles to reuse. Ontohub is the first, and currently the only, ontology repository that meets a substantial amount of the requirements of the Open Ontology Repository initiative which was developed in response to the Ontology Summit 2008, “Toward An Open Ontology Repository.” In addition to addressing the aims of an earlier ontology summit and helping to bridge the ontology reusability gap, Ontohub is one of the most impressive examples of a system that addresses the structured data gap, given that it can map between at least two dozen formats and languages, using at least a dozen kinds of mappings.

Convincing non-ontologists to make use of ontologies is a perennial issue. The article by Westerinen and Tauber, “Ontology Development by Domain Experts (Without Using the ‘O’ Word)” gives a thorough survey of how one can adapt commonly used tools, such as spreadsheets, for ontology development. Since experience with such tools are so widespread, use of such tools greatly reduces the barriers to ontology development by non-ontologists; indeed, by individuals who might not be aware nor care that they are developing ontologies. This work addresses the structured data gap, since the concepts and data that the domain experts deal with are generally in the form of natural language text and database tables. The article explains how one can bridge the gap between the domain experts and their languages and tools and the languages and tools of ontologists.

Another kind of tool commonly used by non-ontologists is the wiki. The wiki allows a user to update web pages using nothing more than a browser. Wikis are one of the most powerful tools for geographically distributed collaboration. The most commonly used wiki software is MediaWiki, the software behind Wikipedia. MediaWiki has a large number of plugins for adding new features. Semantic MediaWiki (SMW) is a suite of plugins that add semantic features such as tagging and queries to the wiki. While SMW is relatively limited as an ontology language, it allows non-ontologists to develop ontologies, usually without every being aware that they are doing so. The article by Baclawski, “Semantic Annotation of Collaborative Work Environments” reports on the SAOCE hackathon at the Ontology Summit 2014. At this hackathon, SMW was used for developing semantic tools for the Ontolog Community wiki. While specific to a particular community, the tools are applicable to other semantic ecosystems. Consequently, this article is related to another ontology summit; namely, the Ontology Summit 2016, “Framing the Conversation: Ontologies within Semantic Interoperability Ecosystems.” The article reports on progress that has been made since the time of the summit.

Another article by Westerinen and Tauber, “Integrating GoodRelations in a Domain-Specific Ontology” gives an experience report of their use of ontologies for a significant real world application domain. This work deals with two of the gaps being addressed by the Ontology Summit 2014: the reuse gap and the structured data gap. The article actually deals with three ontologies: GoodRelations, Schema.org and a domain-specific ontology for manufacturers and retailers of climbing gear. The integration of the GoodRelations and Schema.org ontologies illustrates the difficulties en-

countered when existing ontologies are being reused. The development of the domain-specific ontology illustrates how domain experts and domain-specific knowledge can be expressed in terms of ontologies, rules and queries.

The hackathon that won the IAOA Ontology Summit 2014 Hackathon Prize addressed the schema reusability gap by selecting an example problem and then attempting to construct an ontology by reusing existing ontologies and ontology design patterns. The article by Mike Bennett, “Ontology Design Patterns and Semantic Abstractions in Ontology Integration,” reports on this prize-winning hackathon. The example problem they addressed was risk assessment for travel. The main outcomes of the hackathon were the lessons learned by their attempt. For example, they found that there are two very different notions of “event” that are not distinguished in common usage in English, yet the two notions are not compatible with one another. One event concept is an actual occurrence with a time, place and other characteristics. Another event concept is an abstract notion of something that could occur but, hopefully, will not, such as an accident. The article discusses the challenges involved in integrating incompatible notions such as the two event notions in a single ontology.

Acknowledgments

We wish to acknowledge the help of the members of the Semantic Web for Applied Ontology (SWAO). This article was published by IOS Press in *Applied Ontology*, 12(3-4), 189–194. The final publication is available at IOS Press through <http://dx.doi.org/10.3233/AO-170185>

References

- Baclawski, K., & Thessen, A. (2014, February 13). Tackling the variety problem in Big Data. In *Ontology Summit 2014 Big Data and Semantic Web meet Applied Ontology*. (Retrieved December 9, 2017 from <http://bit.ly/2xefFQT>)
- Grüniger, M., Obrst, L., Baclawski, K., Bennett, M., Brickley, D., Berg-Cross, G., ... Yim, P. (2014). Ontology Summit 2014 Communiqué: The Semantic Web and Big Data meet Applied Ontology. *Applied Ontology*, 9(2), 155-170.
- Ogden, C., & Richards, I. (1923). *The meaning of meaning: A study of the influence of language upon thought and of the science of symbolism*. Orlando, FL: Harcourt Brace Jovanovich.
- Ullmann, S. (1972). *Semantics: An introduction to the science of meaning*. Oxford: Basil Blackwell.