Intelligence Analysis and the Semantic Web: an Alternative Semantic Paradigm

Brock Stitts

***Abstract*— Intelligence analysis involves gathering data from multiple and diverse sources. The Internet provides a monstrously large set of diverse sources. It is so large and diverse, in fact, that the project of manually gathering data from all the potentially useful sources is not feasible. This is where the Semantic Web comes into play. With the Semantic Web, web pages are given a machine understandable content such that web agents can search the internet and perform tasks autonomously. A key property of this machine understandable content is that it must provide for semantic interoperability between the various web pages. The Semantic Web, as its chief advocate, Sir Tim Berners-Lee, admits remains "largely unrealized." The thesis presented here is that by going back to the foundations of semantics, we can generate a new hypothesis as to how the Semantic Web can be realized. In particular, centering on activities (or services) instead of a trying to build a global upper ontology will more effectively cope with semantic interoperability issues and thus will help realize the Semantic Web.**

***Index Terms*—intelligence analysis, semantic web, ontology, semantics**

I. Introduction

Applied Systems Intelligence, Inc. (ASI) has developed a methodology for intelligence analysis which involves evaluation of a threat via a parameterized Bayesian belief network (BBN). "Feeding" this BBN to build a threat analysis involves actively seeking evidence to confirm or deny parameterized hypotheses. An outstanding data source for this analysis would be the Semantic Web. With it, web pages are given a machine understandable content so that web agents can search the internet and perform tasks, such as retrieving evidence, autonomously. A key property of this machine understandable content must be to provide for semantic interoperability between the various web pages. The Semantic Web, as its chief advocate, Sir Tim Berners-Lee, admits remains this "largely unrealized."[1] The thesis presented here is that by going back to the foundations of semantics, we can generate a new hypothesis as to how the Semantic Web can be realized. First, we begin with a brief discussion of semantics.

II. Two Views on Semantics

- Meaning is denotation: words are defined by reference to the objects or things which they designate in the external world or by the thoughts, ideas, or mental representations that one might associate with them
- Meaning is use: words are defined by how they are used in effective, ordinary communication.[2]

If one inquires as to how the denotation gets set up between a word and its object, one finds that the answer is that it is by virtue of using the word in particular contexts that it receives its denotation. In other words, communication happens within the context of some human activity. It is this activity that gives words their meaning. The philosopher Ludwig Wittgenstein considers the following simple scenario (the so-called "builder's language" introduced in section two of the Philosophical Investigations):

"The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones, in the order in which A needs them. For this purpose they use a language consisting of the words "block", "pillar" "slab", "beam". A calls them out; — B brings the stone which he has learnt to bring at such-and-such a call."[3]

This is a simple illustration of the basic functioning of language. The words are used as "moves" in a kind of "game." Wittgenstein coined the term "language game" based on this and other examples. In general, the meaning of the parts (the words and objects of the activity) is derived from the whole (the activity). Likewise, the activity is defined in terms of its parts. This circle is referred to as the "hermeneutic circle." Another way of saying this is:

> "It (the hermeneutic circle) refers to the idea that one's understanding of the text as a whole is established by reference to the individual parts and one's understanding of each individual part by reference to the whole."[4]

Instead of seeing words as the "semantic atoms" out of which sentences are built, the semantic unit is a language game (or activity). Much further argumentation can be provided to support this view, but providing this support is the topic of another paper. Instead, we assume it to be accurate, and generate a new approach to building the Semantic Web based on it.

III. An Alternative Semantic Paradigm

Underlying the approaches of much symbolic artificial intelligence (AI) is the use of set theoretic concepts. In such approaches, the world consists of a set of individuals. These individuals have properties. For an individual to have a property corresponds to its being a member of some set. With such a viewpoint, assertions about individuals are not relative to some context. For the approach presented here, individuals and their properties are relative. In particular, they are relative to an activity. The individuals and their properties are components of an activity. While these individuals and properties may be used in other activities, there is no guarantee of synonymy across them. It is the hypothesis here that the assumption of synonymy across language games leads to much erroneous reasoning. In general, the long chains of inference found in some traditional AI systems will be problematic because they will cut across multiple activities and so will contain invalid inferences. Metaphorically, they will be using apples to infer things about oranges.

A. Application to the Semantic Web

As noted above, semantic interoperability between web services (or agents) is a prerequisite of the Semantic Web. The general idea on how to do this is to create metadata that accompanies web pages. This metadata would contain the semantic contents of the web page. The representation of the metadata would use the web ontology language (OWL). The assumption by Berners-Lee is that the web agents would use an inference engine to reason about this semantic content.[5] The approach here reverses the implicit denotational semantics of Berners-Lee's approach; instead, a web agent knows the meaning of the name and parameters of a service if it knows how to use the service. The semantics of a language game are contained in the game itself. With the Semantic Web, however, different language games must interact. The problem of creating the Semantic Web is then essentially a matching problem. A web agent would try to find an appropriate web service to accomplish whatever task it needed to perform. To do this, it must match up its service request with a web service that can fulfill that request. This matching problem is difficult because any solution must also solve the semantic interoperability problem. This problem comes about in two ways. First, the requester and provider may use different symbols that mean the same thing. The second, and more difficult problem, occurs when they use the same symbol but mean different things by that symbol. To make matters worse, both problems can occur with a single match.

This matching problem has no easy solution. What we outline here are a proposed set of techniques to solve it.

- Use Google-style page ranking as part of the matching algorithm. This is clearly effective to some degree, but one need only attempt using Google to perform Berners-Lee's example of the Semantic Web in action[6] to see why Google only is not sufficient. The goal of this step is really just to generate a set of candidate agents.
- Use case based reasoning (CBR) methods. If one thinks of a web service as a "solution" and a web agent as having a "problem" it is trying to solve, we see that there is a strong analogy between CBR and the matching problem.[7]
- Perform verification. If a web agent has an "answer key" for selected "problems," it can use this key to verify that it has used a web service appropriately. Likewise, if the web service provides a sample usage set, this can also be used for verification. The importance of this step cannot be understated. This is a key part of cognition and scientific reasoning. In cognition, the subject generates expectations based on his or her understanding of a situation. If these expectations are met, that understanding is verified.
- Rather than just providing a service's name, input parameters, and output parameters, provide for instructions (in the form of metadata) on how, why, when, and who should use the service. Although these "instructions" would be prone to ambiguity just as all symbols are, they provide a richer data set to use in matching.

Just as the Web gradually grew as content providers built more content, the approach here would lead to a gradual growth of the Semantic Web. In fact, every piece of this solution would evolve over time. Clearly much work needs to be done to flesh out the details. ASI is currently at work doing this so as to extend its intelligence analysis capabilities.

IV. Conclusion

If the thesis approach presented here is correct, much of the work in deriving an upper ontology will not be all that productive. With the IEEE suggested upper merged ontology (SUMO), for example, there are bound to be numerous cases where its logical axioms are ambiguous; they apply in some contexts but not others. Rather than solving the problem of how to keep chains of reasoning consistent, the approach here is not to perform them. The Semantic Web has two components: the Web and semantics. Semantics for natural languages are captured in dictionaries. However, dictionaries are descriptive. Neologisms are generated when new situations arise that call for them, and are created by a wide variety of language users. Likewise, the web is built "bottom up" by its numerous content providers. Having a committee to define language syntax is workable, but this does not hold for semantics. The semantics of a language is the set of uses of that language. How to use and grow that language is best left to the users of the language.

[1] Nigel Shadbolt, Wendy Hall, Tim Berners-Lee (2006). "The Semantic Web Revisited". *IEEE Intelligent Systems*. Retrieved on 2007-04-13.

[2] See http://en.wikipedia.org/wiki/Philosophical_Investigations
[3] See http://en.wikipedia.org/wiki/Language-game
[4] See http://en.wikipedia.org/wiki/Hermeneutic_circle
[5] See http://www.sciam.com/article.cfm?id=the-semantic-web&print=true
[6] See http://www.sciam.com/article.cfm?id=the-semantic-web&print=true
[7] See http://en.wikipedia.org/wiki/Case-based_reasoning