# Current State of the Semantic Web

AUTHOR: SCOTT STREIT, CHIEF TECHNOLOGY OFFICER

**Executive Summary**
The amount of information on the Web and within organizations has grown at an astounding rate; however, the technology used to accommodate and process this information has not. Computers still cannot automate this information or perform complex tasks. HTML can only provide a structure for text-based information in a document; it cannot do anything with it.

Differing technology environments and protocols make it impossible for computers to cooperate with each another. Additionally, while information grows exponentially, and search engines are returning thousands of results, users are required to wade through countless insignificant results in order to locate the specific information they are looking for.

In order for search functions to meet the demands of busy users with diverse, time-sensitive needs, the technology must be capable of:

- Retrieving large amounts of textual data quickly.
- Allowing users to add annotations so that a reasoning capability exists.
- Making text retrieval more specific.
- Allowing conclusions to be drawn by data on the Web and across organizations.

The first generation of data management took the form of file systems, followed by relational databases. The Semantic Web represents that next generation because it solves problems that were previously difficult or impossible to solve. Not an ontology itself, but rather a language about ontologies, the Semantic Web succeeds where fixed ontologies fail. It is specifically relevant to complex problems with ever changing requirements because it assumes that the relationship of data to itself is higher in value than strictly flat data. Simply put, the Semantic Web thrives in chaos.

In order to understand the Semantic Web it is important to recognize the type of problem that is poorly suited to this technology. A poorly suited problem is an easily defined fixed ontology, an example of which is a General Ledger. Double ledger book keeping has existed for over 3,000 years, and the current technologies of relational databases with fixed viewing capabilities is more than sufficient for this type of problem. If one was to use the Semantic Web for General Ledger issues, we would increase the complexity for no overall value.

The Semantic Web appears in at least four areas. The first area is an interchange format between systems. The fact that the Semantic Web describes any number of ontologies allows the interchange format to change on the fly, without effecting currently functioning systems.

The next area of presence for the Semantic Web is social network visualization or visualization as a whole. Native RDF (Resource Description Framework) viewers may replace html as the native format for viewing web content. The content itself will natively render social networks and other semantically tagged data. Current applications visualize RDF in a variety of ways using custom application viewers.

Moving to the next area, we use the Semantic Web for variable ontological storing, using column oriented storage. We use triple stores or Map/Reduce as a general purpose storage mechanism.

Finally, we use the Semantic Web for the adding of high value analytical data or reasoning. OWL (Web Ontology Language) provides mechanisms to represent higher order inferences. This type of work is the current state of research and enhancements of current specifications.

**Current**

In order to gain reasonable performance for semantically hosted data we use additional storage structures and redundant data. By using these additional storage structures, we achieve acceptable performing accesses. We store variable ontological data that is vocabulary centric. Moderate size graphs (less than 100k) return in a reasonable period of time (25ms). We store multiple terabytes of semantically enabled data and the results scale. Performance times are acceptable.

Intervise's Chief Technology Officer, as a Principal Investigator at the University of Texas at Dallas, solved the "Infinite Graph" problem. As a result, Semantic Web tools, such as Jena, may now handle virtualized graphs of any size. This solution was completed in late 2008 at was added to Jena in early 2009.

**What is not Solved**

Reasoning itself is not solvable as an enterprise solution. The lack of virtualization of graph models is just one shortcoming. Reasoning itself is not scalable and grows in a non-polynomial (NP) way. To solve this problem the problem must be restated as polynomial, which requires reasoning to work as a just- in-time service and leverage the additional structures available from OWL-2. This work leverages the virtualization of memory research, and is also being solved at the University of Texas at Dallas.

**Moving Forward**

In order for organizations to leverage the Semantic Web, it is important to have a well-suited problem and to avoid asking the Semantic Web to do more that it can at this point. Vocabulary centric solutions are ideal for today. Adding ontologies, at this point, is reasonable with the expectations regarding reasoning muted.