

[Share](#) [Report Abuse](#) [Next Blog»](#)

dr. data dictionary

[Ruminations on Declarative Systems, Metadata Publishing and the Semantic web](#)

Friday, July 02, 2010

Impressions of SemTech 2010

Two weeks ago I attended my fifth Semantic Technology conference in San Francisco. It was a great conference! This is my fifth time attending the conference and I plan to attend in the future! My biggest dilemma was which session to attend. At times there were as many as eight concurrent sessions going on.

There were a few big trends that I spotted.

The use of RDFa to annotate web pages using semantically precise elements was clearly a big trend. [Jay Myer](#) of [BestBuy](#) described how the BestBuy sales went up 30% when they added RDFa tags to their product pages. Although many search engines are not transparent about their use of RDFa tags in page rankings, Jay's results should make it clear that this strategy works.

Jay told a great story about how hard it was to find a fridge that met his specific criteria: black with a specific size etc. using a keyword-based search engine. The semantic web will change all of this!

The big factor here is [Martin Hepps GoodRelations](#) ontology for products and services. This can really bring the Semantic Web to the masses. Finally a way to code the hours of operation for your store so that you can use [Siri](#) to ask "What Sushi Restaurants are Open at 10pm Near Here"

There were also a lot of sessions on [LinkedData](#) and specifically on LinkedData in the government area. Jim Helder and his students at RPI are scraping every data set on data.gov and converting them all to RDF and storing them in a huge triple store.

Jim also provided one of the best sound-bites from the conference:

"Get AWAY from the Table"

This goes far beyond the [NOSQL](#) movement that we are seeing. It

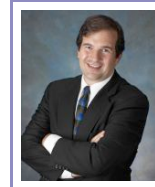
Blog Archive

- ▼ [2010](#) (1)
 - ▼ [July](#) (1)
 - [Impressions of SemTech 2010](#)
- ▶ [2008](#) (9)
- ▶ [2007](#) (18)
- ▶ [2006](#) (22)
- ▶ [2005](#) (8)

Links

[My Web Site](#)
[Wikipedia](#)

About Me



Dr. Data Dictionary

Technology Consultant living in Minneapolis

[View my complete profile](#)

gets to the core of the problems with innovation in many organizations.

Jim is a consultant to the US data.gov transparency process and a key advocate of open linked data standards. It is interesting to see a friendly rivalry between Jim and Tim Berners-Lee who is a consultant for the UK data transparency movement. Both are using RDF to convert data but each is using slightly different strategic approaches.

He was referring to the fact that many organizations over-use the relational model and they need to understand that there are many alternatives, especially when doing mashups of data sets from many sources.

There were also many presentations on natural language processing and finding the true "meaning" of words within unstructured and semi-structured data sets.

I always attend **Brian Sletton's** session on REST. Many people do not understand the relationship between REST, URI's and the semantic web stack. **Joe Wicentowski's work** at the US Department of State on the URL rewriting frameworks within eXist has really reinforced how this can be done easily in a single XQuery module.

I presented a 3.5 hour tutorial session on Entity Extraction. Since it was scheduled for the last session of the last day of the conference I thought the attendance would be very low. But the room was packed and most of the people stayed till the very end.

It was wonderful to finally meet Marie Wallace and DJ McCloskey from the **LanguageWare** team at IBM. Their support of the Apache UIMA standards will be a great step forward to the creation of interoperable language analysis pipelines. Marie and DJ introduced me to the Millennium restaurant just a few blocks from the hotel. My wife Ann and I went there over the weekend and we now own one of their cookbooks.

By the way Google is now also support recipes in their **rich snippets** in their search results so you will soon be able to find "only recipes that take under 30 minutes" in a Google search engine.

There were also many session on the need for good taxonomies and ontologies in the enterprise. The use of **SKOS** for controlled vocabularies was a very hot topic and there were dozens of presentation on how OWL is being used to capture and exchange business rules.

It was also really great to finally meet **Jeni Tennison** after reading her books and following her Tweets for a long time. She is deeply involved in the federal data transparency projects in the UK. Here use of RDF is breaking new best practices for the entire community.

It was also great to catch up with Mark Birbeck, one of the chief architects of the **Ubiquity XForms libraries**. I am looking forward to trying out some of the new tools based on his backplane JavaScript libraries.

The people from Facebook also gave a presentation on how they are adding the ability for people to add a few lines to each web site to allow users to add a “like” button on a web site using a variation of RDF. Many people were a little disappointed to hear that they will not be using namespaces in their interfaces. Facebook felt that most HTML coders could only handle a single namespace. Many people agreed with their findings and thought that until 90% of HTML coders knew what namespaces were that organizations like Facebook were stuck just adding new data to HTML meta elements.

One of the most interesting discussions I had was with the people from **Cray Computer**. Apparently a large unnamed government agency has given Cray Research a very large contract to build customized ASIC (FPGA) chips to do graph analysis. Their ThreadStorm **XMT** architecture has 128 register sets and allows the CPUs to get continuous feeds of graph queries without waiting for memory. Their claim is that the federal agency has told them they are getting a 100X improvement in complex graph queries. The challenge is that the API is currently a low-level C interface and they have not yet put a SPARQL compiler in front of an XML. So no good SPARQL benchmarks are yet available to compare the results of this hardware with a typical triple store. But even if it is fast, the price would be in the six digits for one of the XMP systems. So only very large organizations and government agencies might use this unless it was provided as a service.

One of the other things I found was how many people are using OWL and the OWL reasoners like **Pellet** as replacements for traditional rules engines from companies like FairIsaac. There are two reasons for this. First is that many rules engine companies only talk about their engine performance but not on the need for precise semantics in the data elements. With OWL and the methods behind the semantic web stack we attempt to put semantics higher in the requirements of a system a use pre-built and pretested ontologies.

The second big reason that people are use OWL is that the rules are created using open standards. That means you are not locked into any one vendors rule format. The new W3C standards for rule interchange (RIF) will also start to break down many of the barriers

to the creation and exchange of large industry vertical rule sets to that each organization does not have to start from scratch with a rule base. This will be big for industries like insurance where claims processing ontologies are just being created. Thanks to **Kendal Clark** for good information on this topic.

I also had a great time talking to many people in the publishing area. Seth Maislin gave a great presentation on Taxonomies. Seth and **Marlene Rockmore** joined me for a lunch at nearby Indian restaurant. Marlene is an expert on Taxonomy development. She was working on a book with O'Reilly on taxonomies that is on hold for now but we hope to see one in the future from her.

It was also nice to see how many federal agencies are now starting to use SPARQL in the intelligence community. I saw a good presentation by Dennis Wisnosky with the US DOD talk about how they plan to use NIEM and semantic technologies to cut their integration costs. Yeah NIEM! Most people don't understand that the NEIM is based on the RDF model, it just uses XML syntax to store the relationships. See my web-cast on the SemanticUniverse.com web site for more details. Hopefully his slides will be available in the future. Many people were taking photos of the slides since the DOD is not great at getting their slides out. Wonder why?

I also had lunch with several people from within the intelligence community that had great discussions about the pros and cost of pulling assertions out of NLP-analyzed annotated XML documents and the challenges relating to this. Keeping the links back to the original XML documents to is critical to verify the results of an assertion in context. Traceability, lineage, provenance and time-domain representations in RDF that don't cause a 10X growth in the number of triples is a difficult problem.

If there was some way to relate RDF assertion to a source document XML Node ID (the fourth column question) is a very difficult problem and one that needs close research with native XML and RDF systems. One of the things I learned about the **eXist-Lucene** integration done by Wolfgang Meier is that keeping the node-id as the document ID in Lucene allows non-programmers the ability to configure customized search rank rules past on the context of the keyword in a document. This is where highly customizable **structured search** rocks. I think this innovation needs to be added to RDF triples projects. Keeping the XML node-id of an assertion with an RDF triple context could be a great way to prevent RDF triple bloom for context.

Finally, I felt that David Wood and James Leigh's presentation on **Callimacus** project had the most potential to have a "big impact" on the adoption of Semantic Web tools. This new open source project

allows users to specify a simple HTML template with special XML tags that allows them to bring triples directly into a web application. This has the potential to allow far more non-SPARQL programmers to integrate RDF data directly into a web page much like XQuery does today. With some work it might be possible to auto-generate the XForms bind elements for form rules. This would be a huge win for the integration of rules into web forms without needing to write JavaScript.

You can also see my tweets on the conference here:
<http://twitter.com/dmccreary>

I hope to see more of you in San Francisco next year. Let me know if you are interested in co-presenting any papers!

Posted by Dr. Data Dictionary at **10:15 AM**

Labels: **owl**, **rdf**, **semantic-web**, **semtech**, **semtech2010**, **xml**

1 comments:



Abhishek said...

Very nice insight Dan! written very thoughtfully and concise.
Thank You!

10:51 AM

Post a Comment

links to this post

Create a Link

Home

Older Post

Subscribe to: **Post Comments (Atom)**