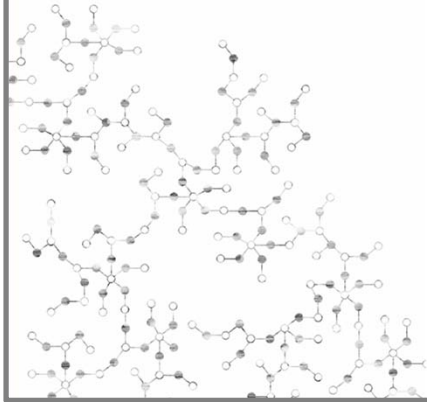


Introducing the Semantic Web

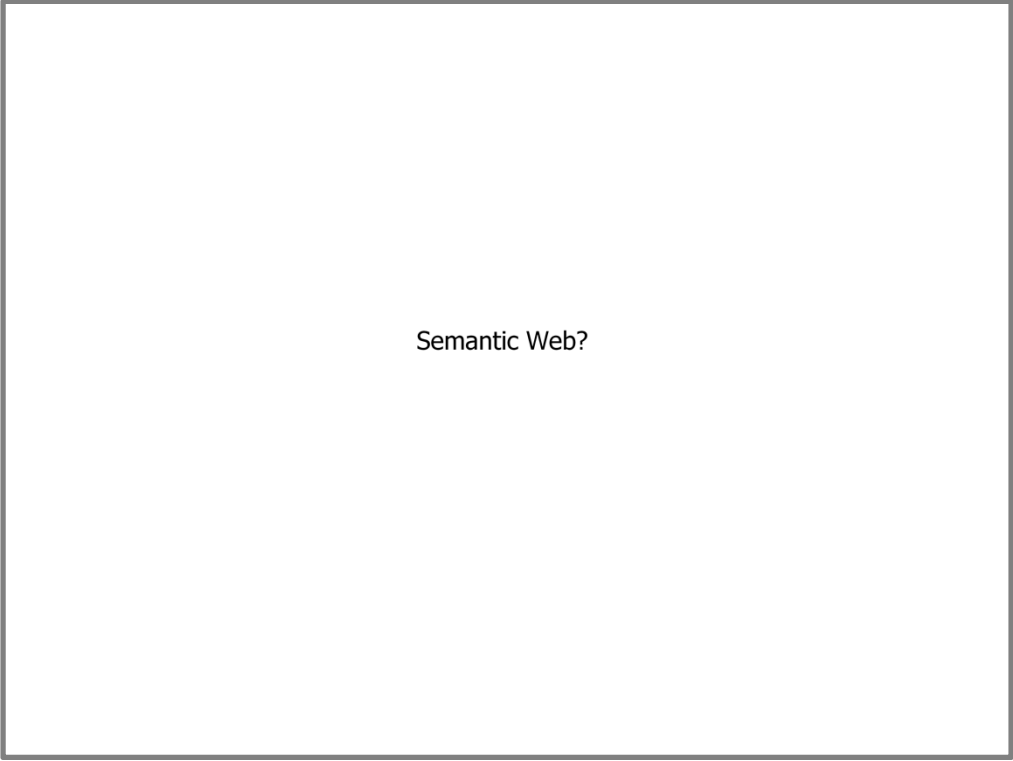
F. Tim Knight
Osgoode Hall Law School Library
York University

A Conversation on the Semantic Web and Legal Information

Annual Meeting Pre-Conference Workshop
Canadian Association of Law Libraries/
L'Association canadienne des bibliothèques de droit
May 24th, 2014 in Winnipeg, Manitoba

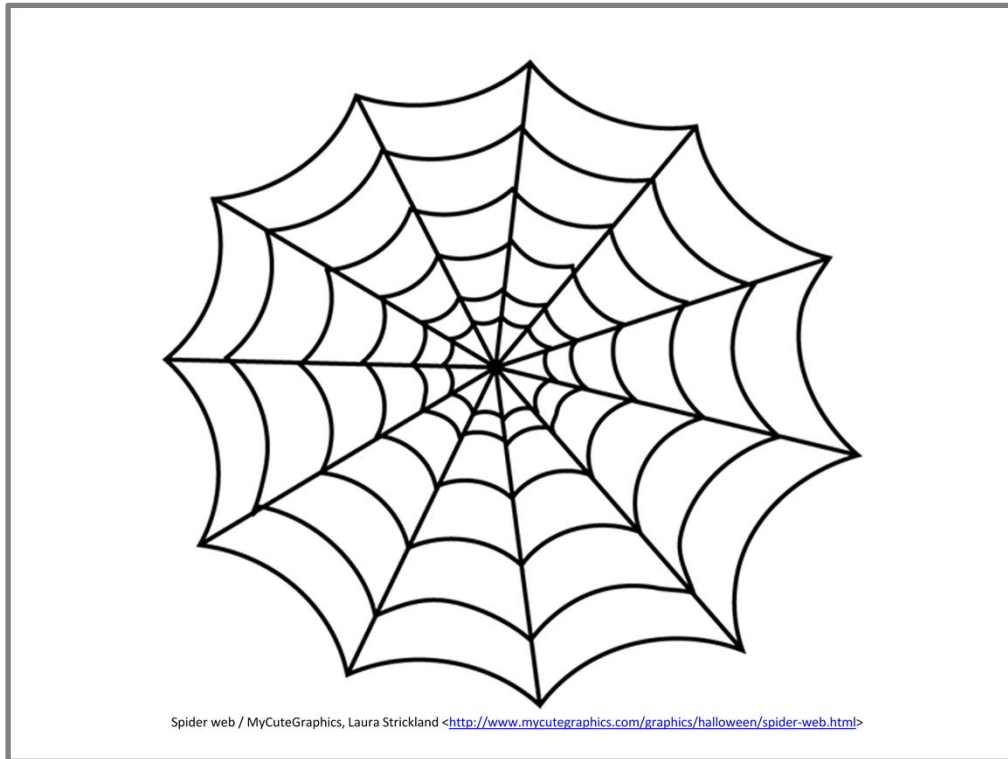


Let's take a high level look at the Semantic Web and see if we can flesh it out and give it some definition.

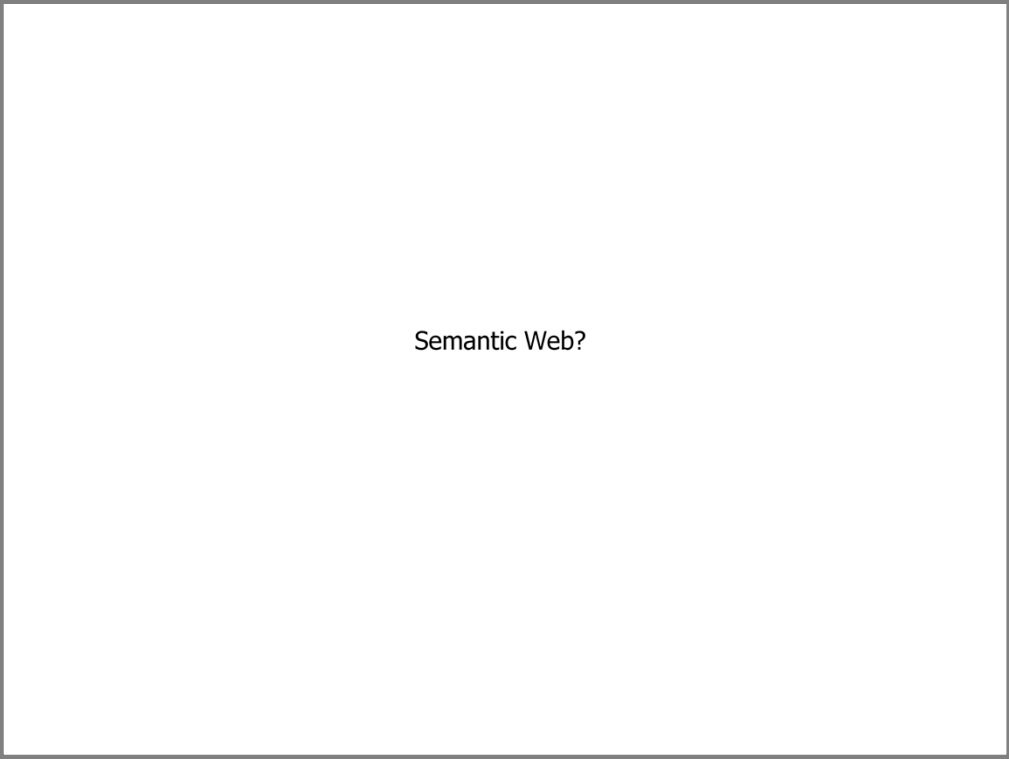


Semantic Web?

You've probably heard the term "Semantic Web." But you may not completely understand what it represents.



We know what a web is ... and in the context of the Internet, the Web is essentially a group of computers connected in a network, and then these networks are connected each other. The result is what's been called the "Web of documents."



Semantic Web?

... but what is a Semantic Web?



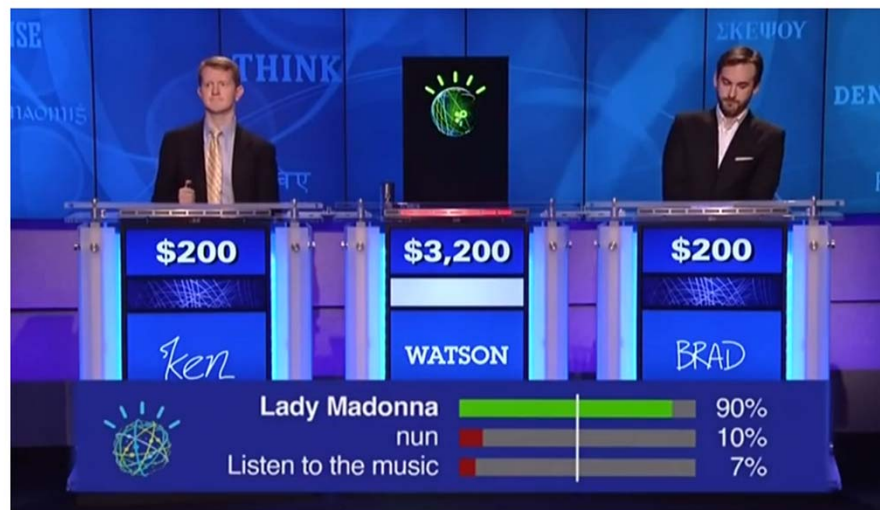
Tim Berners-Lee, founder of the World Wide Web, used the term Semantic Web in a Scientific American article he helped write in 2001. The word “semantic” can be defined as “of, pertaining to, or arising from the different meanings of words or other symbols.” So in this graphic on the cover of Scientific American, the message on the computer screen says, “I know what you mean ...” implying that the computer knows what we mean.

It turns out that this is something that is very difficult for a machine to do, that is, if you are not a human it’s hard to understand what we mean when we communicate and interact with it. And not just understand, but also learn from the interaction.



Nebraska Oddfish. 2013. I could be your light in the dark... <<http://www.flickr.com/photos/ditadata/9167721224/>>

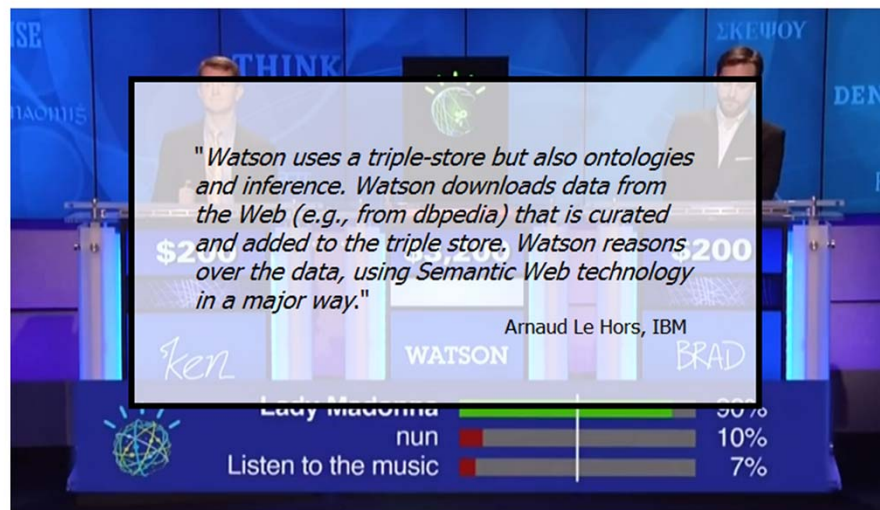
And really, we still don't fully understand how we as humans learn, or at least, what is the best way to present information to make learning easier. So figuring out how to break this all down for machines to learn is a difficult thing to do. As librarians we know that language can be ambiguous, and figuring out the correct context and perspective can be challenging. It's beyond the scope of our time here this morning to outline these problems in detail, but let's touch on something that we are all familiar with ...



Chase Rocker. 2011. Jeopardy - Watson vs. The Humans - Day 1[excerpt] <<http://www.youtube.com/watch?v=vjptrfCVDHI>>

Jeopardy! ... IBM's Watson appearing on Jeopardy in 2011. And not only that, it beat the two top contestants. This was a remarkable achievement for a machine.

In an interview with IBM's Arnaud Le Hors, talking about the development of the W3C's Linked Data Platform, he said this about Watson ...



Chase Rocker. 2011. Jeopardy - Watson vs. The Humans - Day 1[excerpt] <<http://www.youtube.com/watch?v=vJptrfCVDHI>>

"Watson uses a triple-store but also ontologies and inference. Watson downloads data from the Web (e.g., from dbpedia) that is curated and added to the triple store. Watson reasons over the data, using Semantic Web technology in a major way."

We'll talk about "triples" later this morning, but if you haven't heard the term "triple-store" before, that's a database of linked data triples.

So, keep Watson in mind as we continue with our conversation about the Semantic Web.

"The Semantic Web is ... an extension of the current one in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Tim Berners-Lee and others, "The Semantic Web," Scientific American, May 2001, p. 37

In the 2001 article on the semantic web, the sub-title states that the semantic web is, "a new form of Web content that is meaningful to computers." That's a good way to look at it.

And inside the article, Tim Berners-Lee and friends defined the semantic web as:

"... an extension of the current one in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

Later in 2007 Tim Berners-Lee wrote a short blog post he called the "Giant Global Graph." The Giant Global Graph is what Berners-Lee would like to rename the semantic web to get GGG, just like WWW for the World Wide Web. 😊

In this blog post he talked about the evolution of networks and the Web.

"It isn't the cables, it is the computers which are interesting."

"It isn't the computers, but the documents which are interesting."

"It's not the documents, it is the things they are about which are important."

Tim Berners-Lee, "Giant Global Graph"
<<http://dig.csail.mit.edu/breadcrumbs/node/215>>

The network was envisioned as a way to connect computers without having to physically attach one computer to another with a cable. This meant that computers could see each other on the network without using cables. So that was one of the first steps that helped shift thinking away from the cables and to start thinking about the computers themselves. This also meant that anything you created on your own computer could be delivered to any one of the other computer that were part of the network.

Berners-Lee says that what they found was, "there was re-use of the connections, in that, as the packets flowed, a cable which may have been laid for one purpose now got co-opted for all kinds of uses which the original users didn't dream of. And users ... found that they could connect to all kinds of computers which had been hooked up for various reasons, sometimes now forgotten. So the new abstraction gave us more power, and added value by enabling re-use."

Then, of course they connected the networks ...

"It isn't the cables, it is the computers which are interesting."

"It isn't the computers, but the documents which are interesting."

"It's not the documents, it is the things they are about which are important."

Tim Berners-Lee, "Giant Global Graph"
<<http://dig.csail.mit.edu/breadcrumbs/node/215>>

... which ultimately created the World Wide Web that we all know and love.

That was when folks started to realize that it's not the computers that we should be focussed on, but the documents. The network allowed you to access documents and resources without having to know exactly which computer it was stored in. And, he notes, it too, "allowed unexpected re-use. People would put a document on the web for one reason, but it would end up being found by people using it in completely different way."

The Net linked computers; and the Web linked documents.

"It isn't the cables, it is the computers which are interesting."

"It isn't the computers, but the documents which are interesting."

"It's not the documents, it is the things they are about which are important."

Tim Berners-Lee, "Giant Global Graph"
<<http://dig.csail.mit.edu/breadcrumbs/node/215>>

The final stage in the evolution they describe came with the realization that it's not really the documents that are important but the things that the documents are about.

But the trouble with that is, that that "thing" you are interested in, say an aspect of a piece of legislation, or a particular court case, may have parts distributed across the Web in a number of different documents or databases. There may be a number of documents containing information about the same or related things, but there is no easy or reliable way to discover them and bring them all together when they are needed.

The semantic web has the potential of creating this higher level of abstraction, floating above the document layer, that will provide the connections that will link them together.



I'd like to return now to the Scientific American article. The authors begin the article with a scenario that illustrates how the semantic web might be used to coordinate some health care arrangements.

I'm going to read through this passage and while I do that I'd like you to see if you can imagine a similar scenario that we might apply in a legally oriented scenario.

"Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments."

Tim Berners-Lee and others, "The Semantic Web," Scientific American, May 2001, p. 36

The scenario begins like this:

Lucy telephones her brother Pete from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions. Biweekly or something. I'm going to have my agent set up the appointments."

The agent she refers to here is her "semantic web agent."

Semantic Web Agent Scenario

- Retrieves information about her prescribed treatment from the doctor's agent
- Provides a list of several providers covered by medical insurance rated excellent or very good within 20 miles of home
- Calculates possible appointment times using Peter's and Lucy's scheduled availability
- Agent presents a plan

At the doctor's office, through her "handheld Web browser," Lucy tells her Semantic Web agent which she needs to get done. The agent promptly retrieves information about her Mom's **prescribed treatment** from the doctor's semantic web agent, looks up several lists of health care **providers**, and checks for the ones that are covered by her Mom's insurance, it also checks that the providers are within a **20-mile radius** of her **home** and have a **rating** of **excellent** or **very good** on the trusted rating services. It then tries to find a match between available **appointment times** (supplied by the agents of the individual providers through their Web sites) and Pete's and Lucy's personal schedules.

In a few minutes the agent presents them with a plan. Unfortunately, Pete didn't like it—the selected Hospital was all the way across town from their Mom's place, and he'd be driving back in the middle of rush hour.

Semantic Web Agent Scenario

- Pete sets agent to redo the search with improved parameters
- Lucy's agent trusts Pete's agent and assists providing access certificates and data already sorted through
- Calculates new appointment times with warnings
 - Pete needs to reschedule less important meetings
 - Provider not listed under physical therapists but verified by other means
 - Option to view details is provided

He gets his own semantic web agent to redo the search adding some stricter preferences about **location** and **time**. Lucy's agent, has **complete trust** in Pete's agent in the context of the present task, and automatically assists by supplying access certificates and shortcuts to the data it had already gathered and sorted through.

Almost instantly the new plan was presented: a much closer clinic and earlier times—but there were two warning notes. First, Pete would have to reschedule a couple of his **less important** appointments. He checks what they were—and rescheduling is not a problem. The other warning was something about the insurance company's list failing to include this provider under **physical therapists**: The "Service type and insurance plan status [was] securely verified by other means," the agent reassures him. And provides an option to view the details.

Lucy registered her assent at about the same moment Pete was muttering, "Spare me the details," and it was all set. (Of course, Pete couldn't resist the details and later that night had his agent explain how it had found that provider even though it wasn't on the proper list.)

"The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

Tim Berners-Lee and others, "The Semantic Web," *Scientific American*, May 2001, p. 37

Using this scenario as an example the authors conclude that, "The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users."

In the current Web environment the machine cannot easily ascertain meaning from what it finds. It might be able to verify stylistic elements or determine what might be a title, a keyword or a heading, if those things have been included by the document creator. But in a semantic web, consisting of structured data, a semantic web agent that discovers the clinic's Web page will know not just that the page has keywords, such as "treatment, medicine, physical, therapy," but also when the doctor's are available and then return appropriate appointment times.



Semantic Web agent in a legal situation?

So that's an interesting example of how things might work in a future imagined for us about 13 years ago. Let's use this example to start our first conversation. What I'd like to try and do now is quickly imagine how a semantic web agent might work on a legal problem ...

First of all can we think of some use cases? Where might this kind of agent be useful and for whom?