

“Sarkozy bites Obama child”

A commentary on the benefits and distractions of the Semantic Web

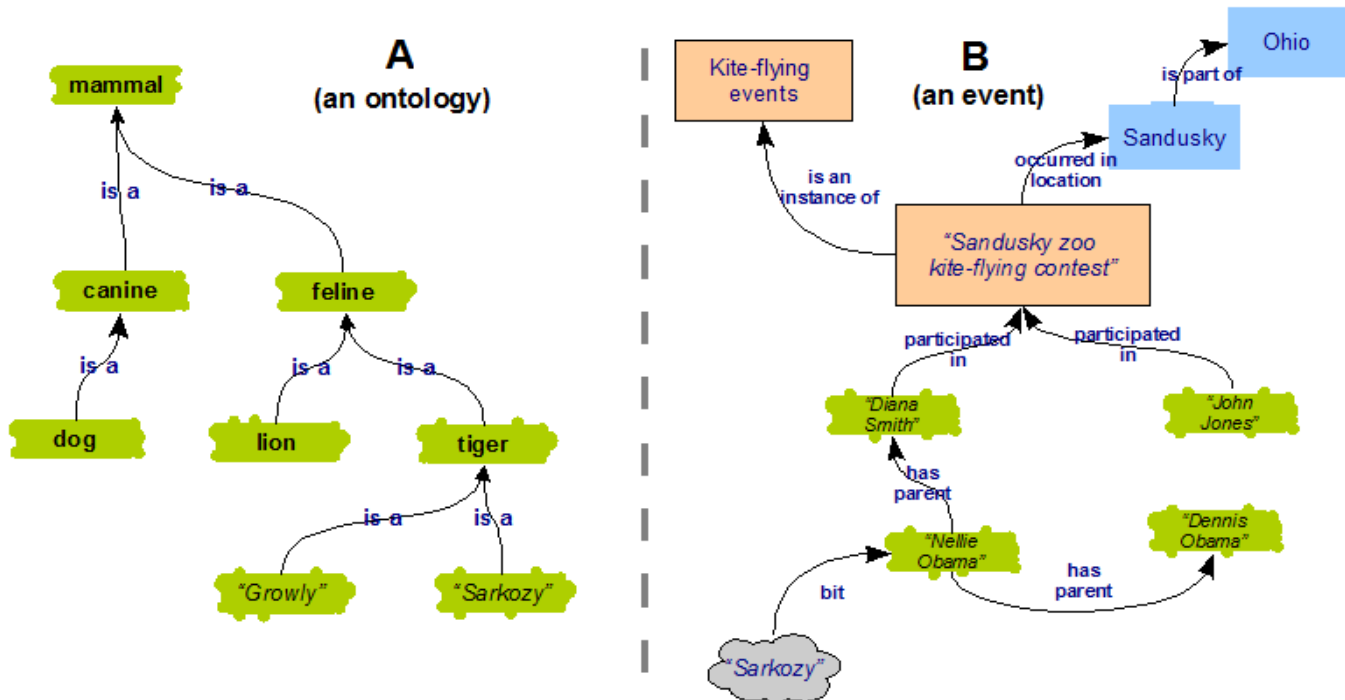
Philip C. Murray
Chief Knowledge Architect
The Semantic Advantage
phil.murray@semanticadvantage.com

For all its potential value, the [Semantic Web](#) could stall rapid economic progress by distracting us from the more critical goal of making work better and more productive. The Semantic Web focuses on making information more “meaningful” and more directly accessible by software applications. That sounds great. Heck, it **is** great. But making the meaning in vast amounts of information out there more accessible is only one challenge in the Information Age and, I would argue, it is not the most important challenge.

The Semantic Web — a very good idea

The Web gives us a network of information. The Semantic Web will give us what Stefan Decker of DERI Galway refers to as a “global network of knowledge.” At its heart, the Semantic Web is a very simple and good idea. The principles of knowledge representation have been around for decades, but if you add a standard way of describing the relationships among concepts to a standard method for giving those concepts a globally unique identity, voilà, you have the possibility of a gigantic snap-it-together knowledge network, a puzzle traversable by computers.

Visualize the puzzle subassemblies as graphs — with a computer ontology of mammals on the left and the **meaning** of “Sarkozy bites Obama child” (a description of an event) on the right — and you get the idea:



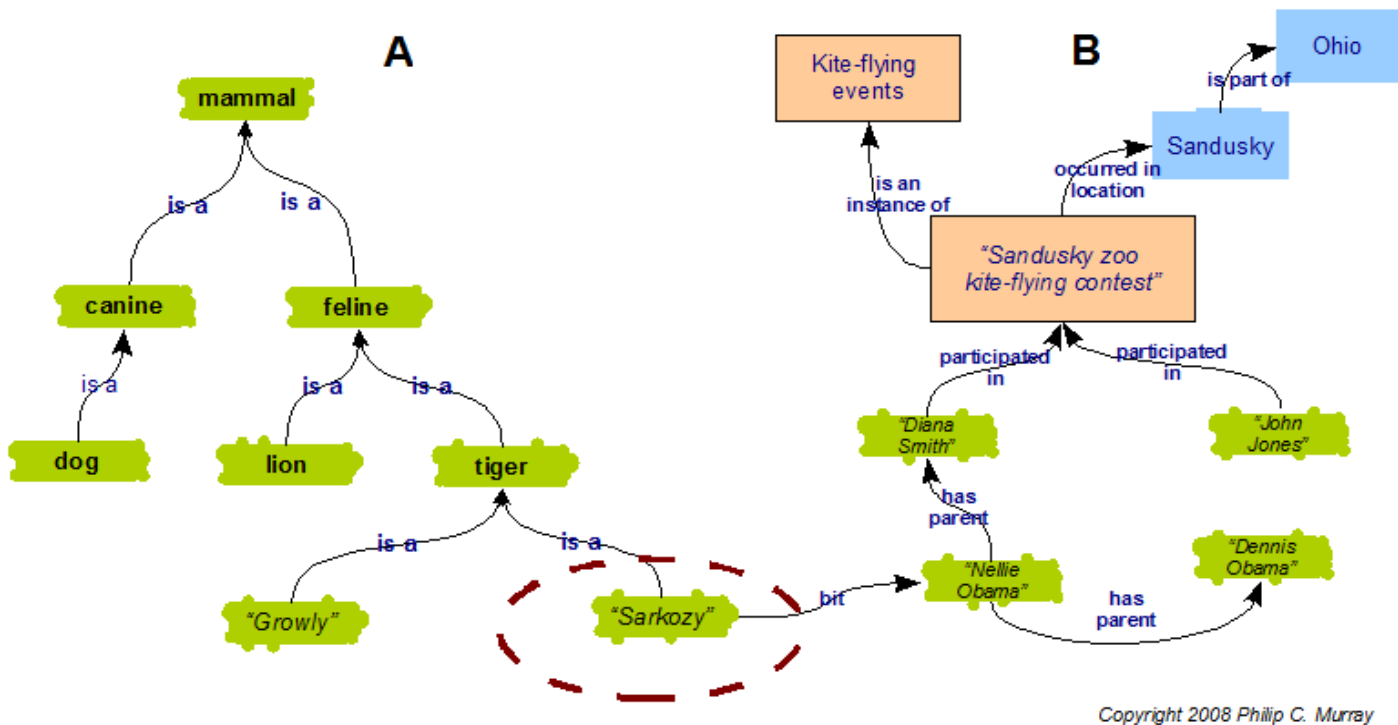
Copyright 2008 Philip C. Murray



You know at least two things from the separate graphs (A) and (B):

1. There is a tiger named “Sarkozy,” and
2. Something named “Sarkozy” has bitten a child of Dennis Obama and Diana Smith.

If you snap these two graphs together at the unique tiger named “Sarkozy” by asserting the two things named “Sarkozy” are the same (that is, by saying they have the same URI), you have connected the animal world with the world of kite flying ... and, in effect, disconnected President Sarkozy of France from a very heinous act.



What’s interesting is that a software application capable of deriving inferences from these connected graphs could use them to determine such “truths” as (1) the French head of state did **not** bite one of Barack Obama’s daughters (on a particular day in Sandusky, anyway) and (2) sometimes there are tigers in Ohio.

What’s equally interesting is that without personal experience of these events, after reading such a headline (1) you, the human reader, **can’t** know the truth behind the words of the headline and (2) you are highly likely, without additional information, to interpret the headline as a serious international faux pas ... and a good reason to return to calling French fries “Freedom Fries.”

It’s not just you. An application capable of interpreting natural language – but **not** supported by such explicit descriptions of meaning as those in these fictional graphs — might also, like a human, interpret that headline as an astonishing international incident. But if the application had access to the underlying meaning expressed in such graphs, it could make useful, valid inferences. It could determine truth in some limited way. So, for example, WOPR from the movie “WarGames” wouldn’t feel compelled to launch ICBMs against Paris.

Yes, you can quibble with my inferences and my graphs. (And if you’re creating ontologies, you should ... and almost certainly would.) But the point is clear: Sometimes unstructured information is deceptive, even when there is no intention to deceive and the information is technically accurate.

Knowing the meaning of a statement often requires something more than words, especially when the words have a very limited context. When we’re communicating directly with each other as humans, we can usually



negotiate that meaning. However, if we simply pass that headline to the next person without interpretation, she has to go through the same process. If she cares to make that effort.

But the Semantic Web is not enough

Like the World Wide Web before it, the simplicity of the Semantic Web is precisely what makes it so valuable. Both of Sir Berners-Lee's brainchilds distribute the processes of creation while enabling centralization of the, er, net results.

That's great. There will be lots of applications based on Semantic Web technologies and they will have very positive, practical benefits. There are hundreds already. At a recent Semantic Web Gathering in Cambridge, Massachusetts, I saw a brief demo of an application that helped stock traders navigate the constraints of the Sarbanes-Oxley Act by applying the underlying meaning of the act to what-if trading scenarios. Cool!

Those are the kinds of applications – often serving businesses much more directly than consumers – that are likely to drive the mainstream adoption of Semantic Web technologies, especially by large companies.

But that's just small change compared to what we might be able to do. Converting laws and contracts into computer-processable meaning would, all by itself, completely justify an investment of many billions into semantic applications. If you've looked at patents, you can also see the end-to-end benefits of enabling more meaningful descriptions of inventions.

Honoring the gods of information

Nonetheless, we are still kowtowing to **information**. Reacting to it. We're focusing on converting information into meaningful forms **after** we spent all that time recording meaning as information. We are increasingly engaged in a relentless, costly, exhausting, wasteful, and error-prone reverse-engineering of information into meaning. While the Semantic Web is right for such needs, there is a great risk – maybe even a likelihood – that we will continue to let the information problem distract us ... even as we congratulate ourselves on having outsmarted infoglut.

Here's where we can be led astray, even with the best of intentions:

- If we spend all our time trying to reverse-engineer information into meaning, we will be distracted from **examining and improving how people produce value** ... and how they benefit from that value. Work has changed – and continues to change – dramatically. Conventional, top-down definitions of what we do at work resemble Soviet 5- and 10-year economic plans. Is that the kind of control we're looking for?
- A corollary: **Economic and social progress now increasingly relies on creating coalitions** and managing the evolution of both meaning and cooperation from those multi-stakeholder cooperative efforts. Human communication and engagement don't simply intersect with meaning; they create meaning. In many cases, they create **the meaning that counts most**.
- **We may end up spending more time feeding the beast than being served by it.** Examples are no further away than your local doctor's office or a hospital ward, where medical personnel may spend more time making sure the records are right than putting their knowledge to work for patients.
- **The rush to codify the past as seen through information will be disastrous if the cost of doing so is diminishing our ability to ask for what we want.** There is a difference between trying to figure out what people want by examining and interpreting their communications – which assumes activities relative to a static, conventional understanding of the world – as opposed to



enabling people to ask for what they want even when no conventional names exist for their needs. **A huge difference.**

- If you focus on the surface manifestations of meaning – both unstructured information and large-scale social responses to that information – and on the best ways to aggregate those surface manifestations, you succumb to relying on small differences in meaning across large amounts of information as a source of value. This is analogous to stock trading derivatives. Look where that got us. **Everything is not miscellaneous. Everything is connected and meaningful.** Unless you're too lazy to figure out how.

Moving forward

We often do need to reverse engineer information into meaning, because past meaning has become history ... and we need to learn from history. The Semantic Web, knowledge engineering, and semantic technologies as a whole help this process in significant ways. For example, Text Analytics technology is often a great way to find new value in information resources that are too massive for people to digest. It's a start in the right direction, but framing the desired output in isolation from general work and productivity requirements is just another form of "siloining."



The Semantic Advantage
3 Sullivan Terrace
Barrington, RI 02806
401-247-7899

Blog: <http://semanticadvantage.wordpress.com/>
Web site: <http://www.SemanticAdvantage.com>

Turning Information into Value