Web 3.0 is only partly about semantics

By Vincent Maher

4 Aug 2008

There is nothing more frustrating than a gaggle of geeks sitting in your boardroom talking about simple concepts in an unnecessarily obfuscatory manner because their revenues are tied to your inability to understand what they're saying or the bills you're paying.

One of the prime targets for this confusion is the Semantic Web. They will tell you it's about artificial intelligence, accronyms such as RDF, object-oriented data structures and meta this and hypertext that. The bottom line is this: the Semantic Web is about bringing information to life.

This is achieved by providing context to the information you publish on the Web. Currently most of the information published on the Web is stored in pages of HTML, which is a language used to define how the information is displayed, not what the information actually is. This means that for others to make sense of it they have to read it, or you need the kind of infrastructure Google has to be able to separate the information into different types.

Practical example

Here's a practical example: You own a travel site that displays both hotel rooms and rates as well as flight information on the same Web page. Structurally there is no difference in the way these two types of information are delivered and there is no way for a search engine or flight-booking aggregator to visit your pages and know the difference. This is a problem because now, and even more so in the near future, you're going to want other sites that aggregate this type of information and send you traffic to be able to do this automatically without you even knowing about it.

The Semantic Web and its technologies allow you to embed tags around your information that are hidden to ordinary readers but visible to the bots, crawlers and other evil-sounding things that spend their digital lives going from page to page gathering information. These tags tell the bots that this type of information is flight information and that information is hotel room information, so they gather this information up and store it in a way that makes more sense for searching.

The potential outcomes of this new way of delivering information are staggering because it paves the way for applications that deliver information to the consumer in a much more intelligent way. A person wanting to travel to South Africa could suddenly be alerted about your new special offer, or a radical drop in your hotel room rates without ever visiting your web site. Someone interested in cooking could be alerted about an ingredient your shop sells while compiling a recipe for a dinner party. Suddenly and automatically, a blogger talking about a particular car has a link to your dealership because he has tagged the car semantically.

OpenCalais

Thomson Reuters recently started a project called OpenCalais, which we used extensively when building the new Mail & Guardian Online website. It's a complicated thing made very simple and it solves a lot of problems very quickly.

In most cases the decision to adopt a semantic approach to content is limited by two things. Firstly, you need to change the way you store and display information on your website. This can be solved by spending some money; it's not complicated technically. The second and more problematic aspect of this is what to do with your historical information. It is seldom practically feasible to manually tag each piece of information because this will take forever and you may never catch up.

Calais is the kind of service you will use to do this automatically. Again, you need a little money to have someone build the tool that sends each piece of archived information to Calais, gets the list of tags or semantic code and puts this into your database.

You can send about 40 000 pieces of information a day to Calais and it will send you back properly tagged semantic information free of charge. This means that most of South Africa's news sites could completely overhaul their archives in a couple of weeks. To do that would take a human a year and it would probably be less accurate.

Web 3.0 is more than just semantics

Despite the obvious value of all of this, semantic data is not going to be the visible aspect of the next generation of Web use that earns it the Web 3.0 title. The fundamental shift from Web 1.0 (1993 - 2004) to Web 2.0 (2004 to present) was the sudden and wide-ranging social use of information on the web brought about by broadband and always-on Internet access. The shift from Web 2.0 to Web 3.0 is going to be the increased automation of knowledge creation. Where Web 1.0 was a web of documents, Web 2.0 is a web of relationships around documents, and Web 3.0 will be a web of applications.

For businesses, this means a further relinquishing of control over where their data is displayed - let's face it, the data is the stick and the transaction is the carrot, so holding onto your data doesn't really make sense anymore anyway. At least half the people reading this just disagreed, probably because they are reading too much into what I just said. I am not suggesting you share or give away all your information; I am suggesting that you make sure the information your customers need is available to them without having to seek you out.

Another compelling reason to share some of the information you're holding onto is that you want to be positioning yourself as a knowledge leader in your field and therefore a trustworthy authority. Trust is easier earned by being useful than by advertising.

For more on Calais and the M&G Online:

- Matthewbuckland.com: The semantic web for publishers and bloggers
- Matthewbuckland.com: M&G ONLINE LAUNCH DIARY: Thou shalt semantically tag with Open Calais
- Vincentmaher.com: Reuters embrace the Semantic Web with Calais

ABOUT VINCENT MAHER

Vincent Maher is the chief innovation officer at Kagiso Media. Read more on his blog at vincentmaher.com, follow him on Twitter at @vincent_maher or connect on LinkedIn.

- [Trends 2015] e-Sports outstrips mainstream sports - 30 Jan 2015
- It's a short ride with your own wind in your sails - 6 Apr 2011
- Google turns whole web into social network - 8 Dec 2008
- How newspapers, mobile media can work together - but probably won't - 16 Oct 2008
- Web 3.0 is only partly about semantics - 4 Aug 2008

View my profile and articles...

For more, visit: https://www.bizcommunity.com