

# WEB 3.0

## Harnessing data intelligence



In 2004, Web 2.0 stood for a period of major innovation after the bursting of the Internet bubble in 2000. The goal was to return to the roots of the Internet: restore a space where everyone can participate. At the end of 2006, the term Web 3.0 appeared to describe the emergence of another important space that meets the requirements of knowledge organisation, groups and communities on

the web. In parallel, this term is used to describe the acceleration of the impact of web technologies on the computer industry.

The good news is that there is no contradiction between the intentions of Web 2.0 and those of Web 3.0. Whereas the former latched on to the experience of users, the latter wants to radically revise the basic technological platform of the Internet. Imagine being able to identify and link an endless number of items of information, data, concepts, as well as to be able to interrogate an immense database (Internet) by putting complex questions in natural language.

Web 3.0 has been in the pipeline for several years already but its development has been slowed somewhat by the excitement surrounding Web 2.0. In 2008, the latter has become a part of the furniture, so to speak: its social aspect, illustrated by the participation of Internet users in writing the web; its architectural aspect (the web as a platform) has taken shape due to technologies bringing the effectiveness of classical applications to a web browser, which is the entry point of all users wanting to contribute to the web; and its technical aspect due to the development of applications that are simple to use for most Internet users. Web 2.0 has established Ajax (Asynchronous JavaScript And XML), Silverlight, AIR, Ruby on Rails: tools allowing the building of Web 2.0



**Viewpoint:** Nova Spivack is the CEO of Radar Networks and the founder of the start-up Twine. He is also Web 3.0 guru.

Page 17

services with ergonomic properties superior to the traditional client-server architecture. By the same logic, RIAs (Rich Internet Applications) have supplanted RDAs (Rich Desktop Applications). Web 2.0, in the same way, heightened the marginalisation of closed and proprietary tools with the installation of open APIs and web services. Finally, in recent years there has been an explosive development of content-creation tools made available to all Internet users: mashups, blogs, wikis, RSS and the like, a spontaneous web indexing method, folksonomy, applied mainly via tags.

The enthusiasm surrounding these technologies translated into a wave of new start-ups. The best technologies have frequently been acquired along the way (such as acquisitions by Microsoft, Yahoo! and Google). Major successes of Web 2.0, social networks, have been bought (MySpace by News Corporation) or intensively courted (strategic agreement between Microsoft and Facebook). Web 2.0 has not generated fortunes for most web players, including the media, but it has contributed immensely to enriching services, content and interaction with their customers.

## Now is the time for 3.0

A consequence of Web 2.0 and the powerful call to people to contribute to the Internet ("Harnessing collective intelligence," is the slogan of Tim O'Reilly, inventor of the term Web 2.0) is that the web has become a data pit. "As the web gets vaster and more complex, and as consumers must work with a growing array of content and services, productivity is seriously being threatened – not only in search but also in every other area of our digital lives," says Nova Spivack, founder of Radar Networks and new Web 3.0 guru (his start-up, Twine, raised nearly US\$ 13 million). "Most of us who work intensively with knowledge and information already have a direct and intuitive experience of how information overload has grown, even in the last decade. Clearly something must be done about this or in another few years we will all be buried in our own information."

The most visible initiatives for Web 3.0 aim at the development of "intelligent web guides," mainly based on the semantic web, the spearhead of their activities. From the economic point of view, it is the technological development that offers the most obvious potential for most industries.

## Web 3.0 technologies

To speak solely in terms of the creation of an intelligent web would be a too-narrow definition of Web 3.0. Many see the next development of the web as an additional step towards the achievement of a "Web OS." In the next years, three technologies will reach a sufficient scale and maturity to advance the development of a web operating system: high-speed networks, wireless or not; network computing, a major group comprising components as diverse as SaaS (Software as a Service: applications and services accessible on the web), interoperability of web services or "cloud computing" (data storage and implementation of applications on the web, via a "cloud" of servers and computers); and the multiplication of open technologies (APIs, data formats, open source developments, Creative Commons and Open Data License).

In 2007, at the Seoul Digital Forum, Eric Schmidt, CEO of Google, described this scenario: "My prediction would be that Web 3.0 would ultimately be seen as applications that are pieced together and that share a number of characteristics: the applications are relatively small; the data is in the cloud; the applications can run on any device – PC or mobile phone; the applications are very fast and they're very customisable; and furthermore the applications are distributed essentially virally, literally by social networks, by e-mail. You won't go to the store and purchase them. ... That's a very different application model than we've ever seen in computing ... and likely to be very, very large. There's low barriers to entry. It solves a lot of problems, and it works everywhere."

The technological aspect of Web 3.0 concerns the media, mainly because of the revolution that it heralds in the IT area.

## In a nutshell

### What is Web 3.0?

A radical revision of the technological Internet platform and a new step towards a web operating system. Tools, such as the semantic web, allowing the identification and linking of an infinite number of data, documents, concepts and communities as well as the ability to put complex queries in natural language on the Internet.

### Threat or opportunity?

An opportunity to develop sophisticated search and navigation functions. An advantage for the media that manage more limited data volumes than general search engines. Openings for creating services (content and advertising). Specialists are banking on a real take-off towards 2010.

### How is it possible to turn it to benefit?

Support the emergence of standards, such as W3C, like that of HTML and XML.

Online: [www.ifra.com/flashlight](http://www.ifra.com/flashlight)

However, for the near future publishers will be more interested in the prospects (in terms of the creation of content and sophisticated businesses services) of an intelligent and automated indexing of web communities and resources.

## What interests publishers

The intelligent web has been associated with the semantic web since the publication in 1999 of Tim Berners-Lee's book "Weaving the Web." The objective of the semantic web is to use formal metadata, based on standard W3C protocols and languages to make web data usable and accessible by programs and applications. But do we not already have search engines for this? Berners-Lee recently answered this question in his blog: "Text search engines are, of course, good for searching the text in documents, but the semantic web isn't text documents, it is data. One thing to always remember is that the web of the future will have



**Barney Pell, CEO of Powerset.com (natural language search; created end of 2006, US\$ 12.5 million funding):**

*"What's happening now will have as big an impact on newspaper content as Web 2.0 did, but it's true there are two terms competing for attention here – 'Web 3.0' and the 'semantic web'*

*and I think when the dust settles they will come to mean the same. The good news is that it's great for business, for syndicating and making the most money off that, but also you don't have to do it all yourself; there are going to be manual tools for content owners but also automation – systems that can read and process content and extract out semantics."*



**Dr. Christian F. Hempelmann, chief scientific officer, hakia.com (semantic search engine, US\$ 21 million funding):**

*"Semantic search technologies do not rely on link referrals to rate web pages. Instead, technologies like hakia's OntoSem analyse the content for what it really means. One other thing that will have extreme importance in Web 3.0 is online advertising, especially for newspapers and media publishing companies. Semantic technology will be critical to this evolution. Understanding content is essential for online advertising, and can be delivered on a consistent basis only by semantic advertising systems."*

BOTH documents and data... The benefit of the semantic web is that data may be re-used in ways unexpected by the original publisher. So when a semantic web start-up either feeds data to others who reuse it in interesting ways, or itself uses data produced by others, then we start to see the value of each bit increased

through the network effect."

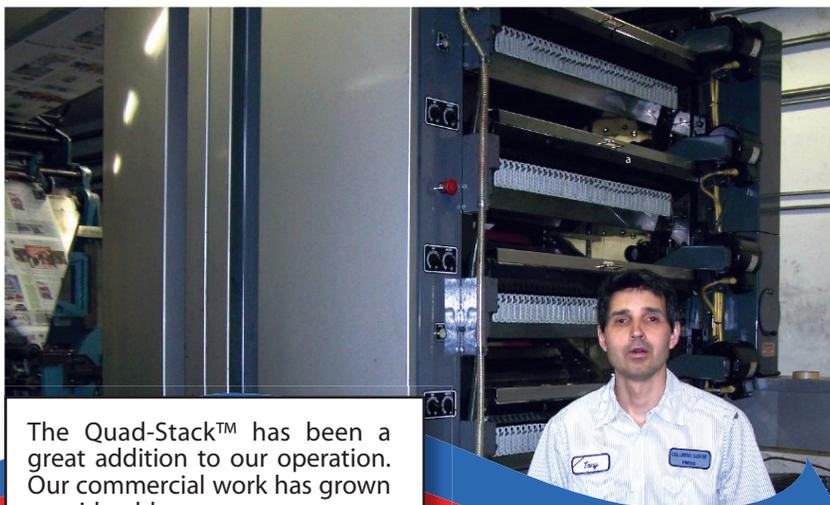
Semantic web operates with known specifications (HTTP, Uniform Resources Identifiers and XML) as well as more specialised standards, of which the best-known are: RDF (Resource Description Framework) as data description module and

RDF Schema for creating description vocabularies; OWL (Web Ontology Language) to permit the linking of documents relating to one and the same topic but spread throughout the web; SPARQL (the first query language for locating data based on semantic web technologies, published in January 2008). The tools exist but their development is complicated and the task is formidable on the scale of the entire web: "Building the semantic web runs up against two problems: building semantic databases is a work-intensive operation that depends on either a manual and community approach – which gives rise to problems of data reliability – or automatic extraction (giving rise to technological problems not yet completely resolved). These databases achieve considerable volumes, and the tools allowing the manipulation of such volumes are not yet in widespread use," says one of the co-authors of the first RDF in 2001 and XML specialist, Eric van der Vlist.

It is premature to talk about building a semantic web but it is already possible to set up semantic websites. The more limited applications, such as those for the media sector, quite easily escape the aforementioned difficulties: the volumes are more limited than in cases where the entire web is concerned and metadata is simpler to assemble reliably. The use of semantic web technologies should thus permit the media to develop search and navigation functions that cannot be developed, at least in the near future, by general search engines.

Valérie Arnould (arnould@ifra.com) and Steve Shippide

## GREAT ADDITION



The Quad-Stack™ has been a great addition to our operation. Our commercial work has grown considerably.

Waste has been reduced as well, due to registration being "on" from the point of start-up.

We hope to grow our business in the future by adding a Quad-Stack™ with U/V capability.

Tony Methvin  
Printing Manager  
Columbia Gorge Press  
Hood River, Oregon, USA

### QUAD-STACK™

**Low waste  
High resolution  
Four year warranty  
Complete presslines**

Web Press Corporation  
Web Leader International, Inc.  
+(253) 395-3343  
info@webpresscorp.com  
www.webpresscorp.com

