

B.5. Web semántica y vaporware del siglo XXI

Por **Lluís Codina**

Codina, Lluís. "Web semántica y vaporware del siglo XXI".
En: *Anuario ThinkEPI*, 2008, pp. 46-49.



Resumen: A propósito del concepto de vaporware se presenta una visión crítica del proyecto de la Web semántica señalando tanto los aspectos positivos como aquéllos que pueden ser únicamente un nuevo caso de vaporware, es decir, algo que desvía esfuerzos hacia un objetivo irreal. Se presentan algunas recomendaciones simples para evitar repetir de forma acrítica un discurso oficial sobre la Web semántica que, hoy por hoy, carece de evidencias.

Palabras clave: Vaporware, Web semántica, World Wide Web Consortium, metadatos, ontologías.

Title: The semantic web and vaporware in the 21st century

Abstract: Starting with the concept of vaporware, we present a critical point of view about the Web Semantic Project. Both positive aspects and others that may only represent a new case of vaporware (something that diverts effort toward meeting an unrealistic objective). A few simple recommendations are offered to avoid uncritical dissemination of official discourse about the Semantic Web that is as of yet unsubstantiated.

Keywords: Vaporware, Semantic Web, World Wide Web Consortium, Metadata, Ontologies.

Antecedentes en el siglo XX

VAPORWARE ES EL NOMBRE QUE RECIBEN los proyectos informáticos que se anuncian reiteradamente pero que no alcanzan una condición de existencia real (luego aclararé lo de real). Como es fácil ver, el término resulta de añadir a la palabra "vapor" el sufijo "ware" siguiendo una costumbre habitual (*hardware, software, etc.*). La verdad es que es un buen término, porque es difícil definir con más expresividad y economía de lenguaje a la vez lo que son, o en lo que se pueden convertir, ciertos proyectos.

Por supuesto, todo proyecto alguna vez es vaporware. Sin embargo, el término suele reservarse para aquellos casos que se estancan más allá de lo razonable y, en cualquier caso, más allá de cualquier previsión.

Por ejemplo, uno de los casos individuales más ilustres es el proyecto *Xanadu*, del gran **Theodore Nelson**. Es un proyecto de los años 60 que, según algunos, jamás se ha materializado en nada. Naturalmente, en todo hay opiniones. Porque según su autor es un

proyecto muy real. Está claro que en cuestiones de tecnología parece que el concepto de realidad es relativo. Puede que algunos seamos demasiado exigentes, pero otros creen que una cosa es real si aparece mencionado, por ejemplo, en un *Powerpoint*.

<http://www.xanadu.com>

<http://www.xanadu.com.aulted>

Existió otro caso de vaporware, de tipo colectivo, mucho mayor, se trata de la inteligencia artificial (IA) tal como se ha vendido desde los años 50 hasta los 80 y tantos. La IA es una disciplina muy digna y respetable que ha realizado grandes aportaciones no solamente a la informática, sino a otras áreas de conocimiento. Actualmente, lidera sectores de alta tecnología y sin sus contribuciones ni la industria moderna en general ni la informática serían lo mismo.

Vaporware es el nombre que reciben los proyectos informáticos que se anuncian reiteradamente pero que no alcanzan una condición de existencia real

En los cincuenta y sesenta se vaticinaba el logro de ordenadores pensantes para los setenta, en los setenta (¿se acuerdan de *2001*, aquella película de 1968?) para los ochenta o los noventa; en los noventa los chicos (y chicas) de la IA echaron la vista atrás, carraspearon un poco y reconocieron que tal vez se habían excedido en sus predicciones. Ahora ya no ponen fechas. Lo mejor de todo sería saber cuántos millones enterró la sociedad durante aquellos años en proyectos que prometían ordenadores inteligentes para ya.

¿El turno de la Web semántica?

Desde 1998, más o menos, si uno se deja convencer por el discurso oficial (lamentablemente amplificado por demasiados autores acrílicos) sobre el proyecto de la Web semántica para una internet en la que los ordenadores serán capaces de razonar y efectuar inferencias sobre los contenidos de las páginas web, uno puede creer que es una realidad tan tangible que si uno no ha reparado en ella es porque uno es un despistado o un inculto.

El problema del *vaporware* es que desvía esfuerzos para nada, hace invertir tiempo y dinero en cosas que a la larga se muestran inútiles y mantiene a miles de personas pendientes de algo que nunca llegará haciéndoles tomar de paso decisiones equivocadas; por ejemplo, no invertir en una tecnología actual y segura en espera de que el año que viene salga otra más avanzada. Lo bueno (esta visto que casi todo tiene algo bueno) es que algún caso, como el de *Xanadu*, han dejado al menos ideas útiles. Es cierto que **Nelson** nunca ha presentado nada real, pero su obra teórica fascinó y capturó la imaginación de muchos durante décadas y tuvo resultados indirectos tan inmensos como la Red, a la que, por cierto, él no aprecia mucho porque dice que es lo que "*Xanadu* evita ser" (ya ven, genio y figura...). La obra de **Nelson**, como creación intelectual, sigue teniendo un gran valor y seguramente seguirá teniendo capacidad para fascinar a nuevos creadores durante años.

Ahora bien, la cuestión es que algunos creemos que demasiados aspectos del proyecto de la Web semántica lanzado y liderado desde el *World Wide Web Consortium* (W3C),

Insistir en que las ontologías permitirán razonar a los ordenadores y realizar inferencias sobre los contenidos de la Web en su conjunto y en todos los ámbitos del conocimiento y de la realidad, es no querer ver precisamente la simple realidad

y que cuenta con aportaciones de centenares de investigadores del todo el mundo, se han adentrado en los caminos del peor *vaporware*, el que solamente sirve para crear *powerpoints* (y para obtener subvenciones, de paso).

Lo bueno y lo malo

Vayamos por partes. Casi nadie duda que el proyecto de la Web semántica también tiene aspectos muy positivos; aspectos con los que entiendo que quienes trabajamos en internet y de publicación electrónica debemos no sólo familiarizarnos sino incluso implicarnos activamente. La idea de separar contenido con xml (o con xhtml) y presentación con XSL (o con CSS) es potentísima, y la verdad es que por sí sola justificaría cualquier proyecto. El propio lenguaje xml es una creación intelectual de primer orden que ha generado una nueva era en la generación y gestión de documentos digitales, con implicaciones que abarcan no solamente la clase de documentación con la que trabajamos en centros de documentación y bibliotecas, sino en toda la ofimática.



Además, un sistema unificado para expresar metadatos, con una buena base lógico/matemática, como *RDF* puede ser útil para el intercambio de información e influenciar de manera positiva la próxima generación de sistemas de información. Las ontologías

como medio de expresar y formalizar conocimiento vienen siendo utilizadas desde hace al menos una década por terminólogos, lingüistas y expertos en semántica documental, aunque sea en dominios y aplicaciones muy concretos. Por eso, es simplemente fantástico que se haya desarrollado ahora un estándar como *OWL* que ayuda a la aparición de un sistema estándar de marcado y codificación de ontologías.

Veamos ahora la parte, en nuestra opinión, no tan buena. Los metadatos generados de forma humana tienen utilidad indudable en entornos controlados, por ejemplo, en una intranet, en una base de datos o en un repositorio, no digamos ya en un banco de imágenes, por ejemplo; pero para los motores de búsqueda, que están obligados a practicar una modalidad de recuperación de información con adversario (gracias al doctor **Ricardo Baeza-Yates** por el concepto), hacer caso a los metadatos que hay en cada página sería un suicidio. Bien, pues en el *Consortio (W3C)* nadie da señales de haber reparado en ello y el discurso oficial, siempre amplificado por autores un tanto acríticos, sigue siendo el mismo que a finales de los noventa, cuando aún no se había revelado el análisis de enlaces como una técnica mucho más fiable.

Además, insistir en que las ontologías permitirán razonar a los ordenadores y realizar inferencias sobre los contenidos, no de un dominio del conocimiento bien delimitado y para unas tareas específicas (como ya sucede), sino de la Web en su conjunto y en todos los ámbitos del conocimiento y de la realidad, es, hoy por hoy, no querer ver precisamente la simple realidad. No existe ninguna evidencia empírica ni base conocida alguna para que esto pueda funcionar, ni a corto ni a medio plazo.

Creo que casi todos los que trabajamos en temas de la Red tenemos un enorme respeto por el *W3C*, cuyas actividades, en muchos casos, constituyen monumentos intelectuales, como ya he intentado dejar de manifiesto. Con mis modestos medios, defiendo en general sus políticas sobre internet y estoy de acuerdo con el noventa por ciento de su actividad, pero, en mi humilde opinión, creer en las bases sobre las que se sustentan algunos aspectos de la Web semántica significa casi creer en la parapsicología.



Mi propuesta es la siguiente: en primer lugar, por supuesto recomiendo que desde la Documentación nos interese e impliquemos en el proyecto de la Web semántica. Tiene cosas muy valiosas que pueden cambiar para siempre la semántica documental.

Ahora bien, a la vez recomiendo que, de forma general, nos atrevamos a decir que el emperador está desnudo. Dos ejemplos concretos: debemos dejar de afirmar que gracias a los metadatos las búsquedas en la Red son más precisas. En primer lugar, no hay un sólo motor de búsqueda relevante (¿ven a lo que me refiero con lo de "real"?) que los utilice. En segundo lugar, si lo hicieran, las búsquedas se volverían mucho peores, no mejores, pero es que además la calidad del contenido de las páginas web se vería en peligro: si uno tiene suficiente con llenar de metadatos su página para obtener una buena posición en la página de resultados de los motores, ¿para qué molestarse en crear buenos contenidos?

El uso de metadatos en contextos determinados, como los señalados más arriba, los justifican de sobras, no es necesario ignorar la realidad para defenderlos.

Otro ejemplo: no dejemos creer a nuestras audiencias, repitiendo de manera acrítica el discurso oficial, que un fabuloso sistema de ontologías a escala, nada menos que de toda, pero toda, la Web permitirá razonar a los agentes de usuario (navegadores), al menos mientras quien sostenga ese discurso no presente alguna mínima evidencia.

Hoy por hoy, no existe la menor evidencia ni base conceptual que justifique tal cosa en un entorno tan abierto, descentralizado, multilingüe, intercultural y lleno de intere-

Actualmente no existe la menor evidencia ni base conceptual que justifique la existencia de un sistema de ontologías a escala en un entorno como es internet, donde cada uno creará las suyas propias

ses contradictorios como es internet, y donde cada uno creará sus propias ontologías literalmente *co-mo-le-dé-la-ga-na*; esto último en el supuesto que existan suficientes “ontólogos” para generar conocimiento que pueda leer una máquina sobre los varios millones de cosas que, según los expertos, sabe un niño de siete años, tales como que con una cuerda se puede estirar, pero no se puede empujar y sin embargo ignoran los ordenadores. Cosas con las que existe un consenso general en la IA: es imposible que una máquina “piense”, al menos de manera útil.

Ahora bien, ¿imagina alguien de verdad a los ordenadores razonando sobre aspectos como los derechos humanos, la Historia del Arte, la economía política o el estatus legal de Kosovo?, ¿imaginan búsquedas basadas en ontologías del estilo “buenas prácticas gubernamentales en derechos humanos”? ¿la supuesta ontología a escala de toda la Web sobre derechos humanos considerará la pena de muerte una buena práctica?, ¿y aplicar cadena perpetua? Desde luego, la que haga *Amnistía Internacional*, no. Pero, ¿no harán a su vez sus propias ontologías los defensores de la pena de muerte y de la cadena perpetua? Por ejemplo, con una disciplina sólida como la biología, ¿habrá una ontología unificada de la teoría de la evolución y de la teoría del diseño inteligente?, ¿no es más lógico creer que habrá en realidad un choque total de ontologías en una inmensa cantidad de cuestiones? Sobre las distintas lenguas, ¿conocerán los ordenadores las equivalencias de todas y cada una de las palabras de todas y cada una de las lenguas de todos y cada uno de los documentos de la Red y de los metadatos y de las ontologías?

Conclusiones

En esta grandiosa visión del W3C de una Web semántica con ordenadores razonando como agentes inteligentes resulta que cada punto de la cadena de razonamiento, desde suponer que los metadatos creados por los propios interesados son totalmente exactos y fiables hasta suponer que los ordenadores pueden razonar a la manera humana, pasando por la idea de que la lógica matemática puede representar todo conocimiento de manera fiable y además ¡unívoca!, carece de base.

Por tanto, a favor de la visión racional de un nuevo sistema de marcado, un nuevo lenguaje para metadatos y ontologías y el impulso a favor de una Red que incluye el proyecto de la Web Semántica, por supuesto que sí y a ser posible con nuestra implicación y apoyo. A favor de que lo anterior deba ir acompañado de un nuevo y gigantesco episodio de *vaporware* sólo para no contradecir el mensaje oficial, por supuesto que no. Con una vez fue suficiente y ya produjo una generación que se empeñó en decir que el traje nuevo del emperador era tan y tan bonito. ¿Por cierto, quién se acuerda de *Quaero*?

Fuentes

Página oficial del proyecto de la web semántica
<http://www.w3.org/2001/sw/>

Artículo de la *Wikipedia* sobre la web semántica
http://en.wikipedia.org/wiki/Semantic_Web

Página del proyecto sobre web semántica y sistemas de información documental (vinculado al autor)
<http://www.semanticweb.net/>

Shadbolt, Nigel; Berners-Lee, Tim; Hall, Wendy. “The semantic web revisited”. En: *IEEE Intelligent Systems*, 2006, v. 21, n. 3, pp. 96-101.
http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisited.pdf

Berners-Lee, Tim; Hendler, James; Lassila, Ora. “The semantic web”. En: *Scientific American*, 2001, mayo.
<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>