

# TELOS

## Cuadernos de Comunicación e Innovación

### Avanzando hacia la web semántica a través de la anotación de contenidos

Germán Toro del Valle de Telefónica I+D

Tecnologías



*¿Por qué el proyecto europeo INSEMTIVES contribuye a hacer realidad la web semántica? Obtenga la respuesta en este interesante artículo*

La promesa de una Web Semántica, esto es una Web más inteligente, más humana, ha estado entre nosotros desde que en el año 2001 el creador de la Web, el señor Tim Berners-Lee, publicó su archiconocido artículo “*The Semantic Web*” [1] en la revista *Scientific American*. Desde entonces, se han realizado grandes avances en este ámbito en la forma de especificaciones<sup>1</sup>, ontologías<sup>2</sup>, herramientas y plataformas<sup>3</sup>, etc. Con todo ello y salvo contadas excepciones, para el gran público la promesa de la Web Semántica no ha dejado de ser eso: una promesa. En este artículo presentamos el trabajo que estamos realizando en el proyecto INSEMTIVES (<http://www.insemtives.eu>) del 7º Programa Marco de la Unión Europea, cuyo objetivo no es otro que acercar y poner al servicio del gran público la enorme potencia y ventajas que ofrece la Web Semántica con vistas a convertir la eterna promesa en una realidad. **INTRODUCCIÓN** La Web, tal y como la conocemos hoy, ha experimentado una enorme evolución desde aquellos primeros documentos enlazados mediante hipertexto de principios de los años 90. Esta evolución ha ido siempre dirigida a satisfacer las necesidades de sus usuarios: necesidad de publicar y consumir contenidos (texto, imágenes, vídeos, etc.), necesidad de localizar los contenidos que nos interesan (buscadores), necesidad de mantener el contacto con otras personas (redes sociales), etc. En este sentido, todo apunta a que la necesidad de interactuar con la Web de forma más natural, más humana, guiará la evolución futura de la misma, tal y como en su momento predijo Nova Spivack<sup>4</sup> (ver Figura 1).

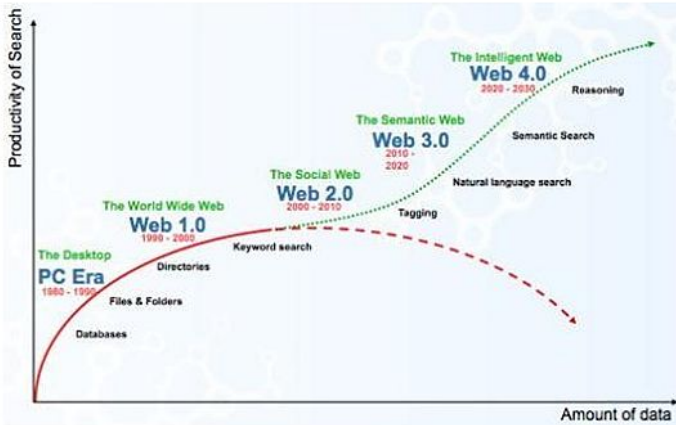


Figura 1.- Evolución de la Web (según Nova Spivack).

Esta nueva Web se caracterizará por la capacidad de gestionar no sólo datos e información, sino la semántica asociada a dichos datos, esto es, el significado de los mismos. Esto, que en principio pudiera parecer algo inmediato, presenta una enorme complejidad, ya que en definitiva se trata de conseguir que dos universos completamente independientes (como son el de las máquinas (ordenadores, servidores y equipos en general) y el de las personas) lleguen a entenderse. Se hace necesario definir nuevos modelos de información (ontologías), nuevos formatos así como nuevas herramientas para generarlos y consumirlos. Con todo ello, el enorme esfuerzo que esto supone se verá ampliamente superado por los beneficios que se obtendrán en forma, por ejemplo, de búsquedas más precisas, de una navegación más inteligente, de sistemas de recomendación avanzados, y muchas otras aplicaciones que probablemente hoy en día ni siquiera imaginamos. **LAS BASES DE LA FUTURA WEB SEMÁNTICA** La futura Web Semántica se basa en un concepto fundamental como es el de ontología<sup>5</sup>. A grandes rasgos, una ontología es una representación formal de conocimiento (datos más su semántica asociada) en forma de un conjunto de

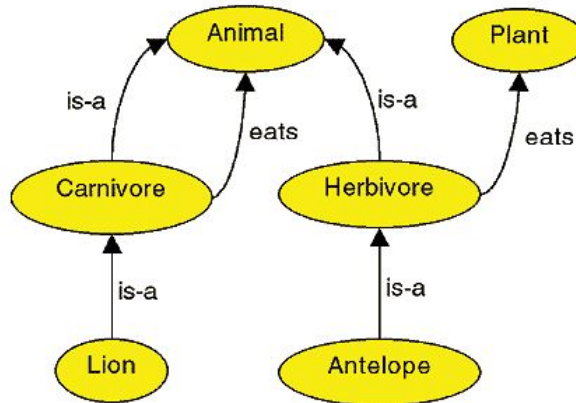


Figura 2.- Ejemplo de ontología.

conceptos y relaciones entre dichos conceptos (ver Figura 2).

El modelado de la información en forma de ontologías hace posible un consumo mucho más natural e inteligente de la misma. Por ejemplo, a partir de información modelada utilizando la ontología anterior (Figura 2) podríamos realizar búsquedas de páginas Web como las siguientes:

- Páginas Web donde se mencione un animal: obtendríamos como resultados páginas donde se mencione a un león, un antílope o cualquier otro animal modelado en nuestra ontología.
- Páginas Web donde se mencione un animal carnívoro: obtendríamos como resultados páginas donde se mencione a un león o cualquier otro animal carnívoro modelado en nuestra ontología.
- Páginas Web donde se mencione un animal herbívoro: obtendríamos como resultados páginas donde se mencione a un antílope o cualquier otro animal herbívoro modelado en nuestra ontología
- Etc.

En la Web actual, estas búsquedas no son posibles ya que la información expuesta no incluye la semántica de los datos (por ejemplo, un león es un carnívoro que es un tipo de animal), por lo que sólo podemos realizar búsquedas basadas en el contenido (i.e., palabras clave) y no en el significado de dicho contenido. A modo de ejemplo, si realizamos una búsqueda en cualquier buscador Web (<http://www.google.com> o <http://www.bing.com>, por ejemplo) con la palabra clave “león” obtendremos una enorme lista de resultados sobre la ciudad española y no sobre el animal que puede ser el objeto de nuestro interés. Con todo ello, el concepto o la idea de ontología no es suficiente para construir la futura Web Semántica. Junto con dicho concepto, la Web Semántica se está construyendo sobre un conjunto de especificaciones y estándares cuyo objeto no es otro que el de materializar el concepto de ontología de forma abierta y reutilizable. Dentro de estas especificaciones y estándares ocupan un lugar privilegiado las especificaciones publicadas hasta la fecha por el *World Wide Web Consortium*<sup>6</sup> (W3C) que hacen posible

el modelado semántico de la información de forma estándar. **¿UNA NUEVA WEB O UNA WEB RENOVADA?** La razón de que actualmente la Web Semántica no sea una realidad, a pesar de que la tecnología necesaria existe y está disponible, se debe al exceso de ambición a la hora de construirla desde cero. No han sido pocos los que han intentado desarrollar sus propias aplicaciones y servicios Web semánticos mediante el modelado semántico de los datos por ellos manejados, utilizando para ello las técnicas anteriormente mencionadas. Sin embargo, al margen del esfuerzo que supone re-modelar dichos datos de forma semántica desde cero, generalmente el pequeño volumen de datos con capacidades semánticas disponibles hace que las aplicaciones y servicios ofrecidos sean bastante limitados. En este sentido, es importante entender que la Web Semántica no es una nueva Web construida desde cero, si no la Web de siempre pero enriquecida con capacidades semánticas. Esto no puede ser de otra manera si atendemos al enorme volumen de datos e información disponibles en la Web actual. Aunque pueda parecer algo sorprendente, no fue hasta el año 2004 cuando empezó a considerarse la posibilidad de evolucionar la Web existente hacia una Web con capacidades semánticas en la forma de especificaciones como RDFa<sup>7</sup>. Es en esta línea, es decir, la de añadir una capa semántica sobre el contenido actualmente disponible en la Web, en la que se está trabajando actualmente en el proyecto INSEMTIVES (<http://www.insemtives.eu>) del 7º Programa Marco de la Unión Europea a través de técnicas de anotación semántica de la información. **ANOTACIÓN SEMÁNTICA DE CONTENIDOS** A grandes rasgos, una anotación semántica es un meta-dato que puede asociarse a cualquier contenido (i.e., textos completos o palabras concretas, imágenes o áreas de las mismas, vídeos o secciones de los mismos, etc.) con vistas a enlazarlo con determinadas entidades de una ontología. A modo de ejemplo, véase la Figura 3 donde se ha anotado semánticamente un texto. Como puede observarse, tras la anotación la cadena “XYZ” ha dejado de ser una simple cadena de caracteres para convertirse en la empresa XYZ establecida el 3 de Noviembre de 1978 y localizada en Londres (la capital del Reino Unido, no otra). Lo mismo ocurre con Bulgaria que ha sido anotada semánticamente como el país europeo. En este sentido es importante observar cómo el contenido original sigue estando intacto a pesar de haberse anotado semánticamente. El mundo de posibilidades que esta pequeña capa semántica sobre el contenido original nos ofrece es enorme siendo posible, por ejemplo:

- Realizar búsquedas semánticas del tipo “páginas donde se mencione una empresa europea fundada antes de 1980” o “páginas donde se mencione un país de la Europa del Este” (el texto anotado en la Figura 3 sería un resultado satisfactorio para ambas búsquedas en caso de utilizar la ontología adecuada sólo esbozada en la figura).
- Ofrecer al visitante contenidos semánticamente relacionados como, por ejemplo, información sobre el valor de las acciones de la empresa XYZ, o anuncios relacionados como, por ejemplo, anuncios de turismo en Londres, el Reino Unido o Bulgaria.
- Hacer posible sistemas de sindicación inteligente permitiendo a los usuarios suscribirse, por ejemplo, a cualquier noticia sobre “empresas europeas localizadas en el Reino Unido”, etc.

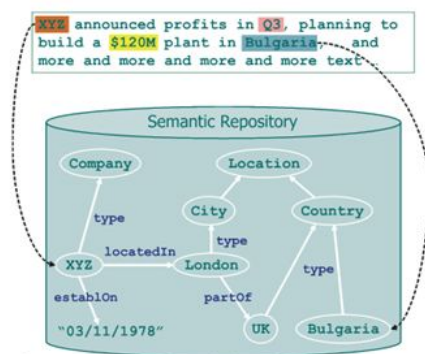


Figura 3.- Ejemplo de anotación semántica.

Como puede observarse, el modelado, o más concretamente, la anotación semántica de la información permite ofrecer servicios mucho más inteligentes y avanzados que los actuales, basados en los contenidos planos, sin semántica asociada. **EL PROYECTO INSEMTIVES<sup>8</sup>: INCENTIVES FOR SEMANTICS** El proyecto europeo INSEMTIVES está generando la tecnología necesaria (i.e., metodología, modelos, plataforma y herramientas) que permita añadir una capa semántica (en forma de anotaciones semánticas) sobre el contenido actualmente

disponible en la Web. Para ello, el proyecto combina técnicas de extracción automática de información semántica (i.e., procesamiento de lenguaje natural en textos, análisis de imágenes, etc.), las cuales no dejan de ser técnicas probabilísticas, con la que probablemente sea la fuente de información semántica más fidedigna de todas: las personas, ya que somos nosotros los que en última instancia dotamos de significado a la información. En este sentido, el proyecto INSEMTIVES está analizando distintas técnicas motivacionales y basadas en incentivos con vistas a conseguir que los usuarios de las páginas Web se vean motivados a realizar anotaciones semánticas. Entre las técnicas utilizadas se encuentran los denominados Juegos Con Un Propósito (JCUP, del inglés *Games With A Purpose*<sup>9</sup> (GWAP)) cuyo objeto no es otro que conseguir que los usuarios realicen anotaciones semánticas mediante juegos, técnicas basadas en la diversión como las que propugna la *Fun Theory*<sup>10</sup>, así como otros incentivos más tradicionales como incentivos basados en recompensas (i.e., premios), reputación o imitación [2]. Es importante advertir que el proyecto INSEMTIVES contempla la anotación semántica de cualquier tipo de contenido: texto (palabras, frases, párrafos, etc.), imágenes (incluyendo áreas concretas de una imagen), audio (incluyendo secciones concretas de la pista) y vídeos (incluyendo secciones concretas del vídeo). De hecho, en la actualidad, el proyecto está evaluando sus contribuciones en tres ámbitos muy distintos como son:

- Anotación de portales corporativos: permitiendo a sus usuarios anotar semánticamente la información allí expuesta con vistas a ofrecerles servicios avanzados de búsqueda, navegación y recomendación inteligentes. En las Figura 4 y 5 puede observarse un ejemplo de las herramientas que se están desarrollando en el contexto del proyecto INSEMTIVES para la anotación de páginas Web en portales corporativos.
- Anotación de servicios Web: con el objetivo de ofrecer capacidades inteligentes de búsqueda y localización de servicios (por ejemplo, localizar “un servicio que ofrezca información meteorológica de países de Europa del Este en unidades métricas”).
- Anotación de mundos virtuales: con el objetivo de ofrecer a sus creadores capacidades avanzadas para su construcción (por ejemplo, para obtener todos los tipos de animales domésticos disponibles en el mundo virtual con vistas a construir un entorno doméstico).

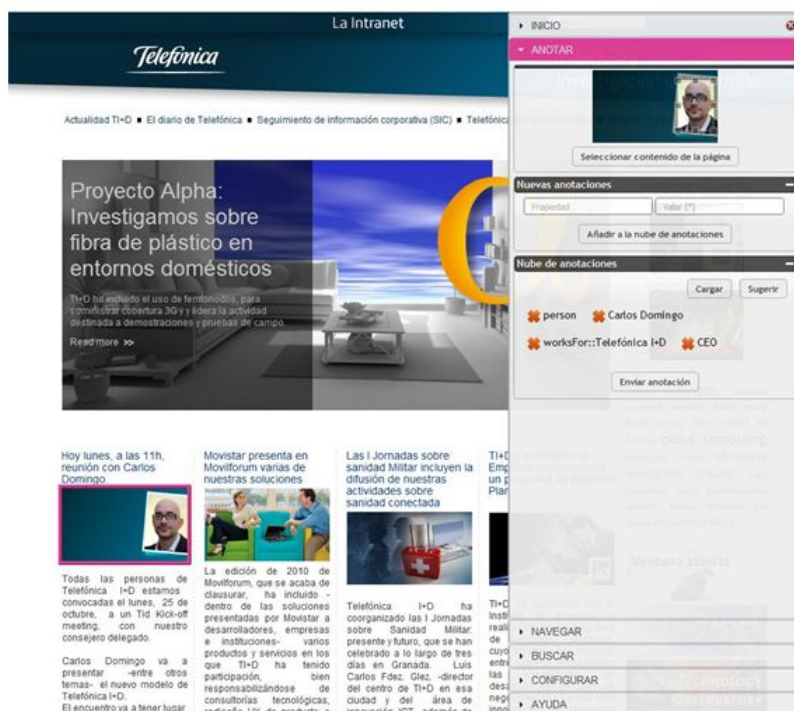


Figura 4.- Ejemplo de anotación utilizando la herramienta de anotación semántica.

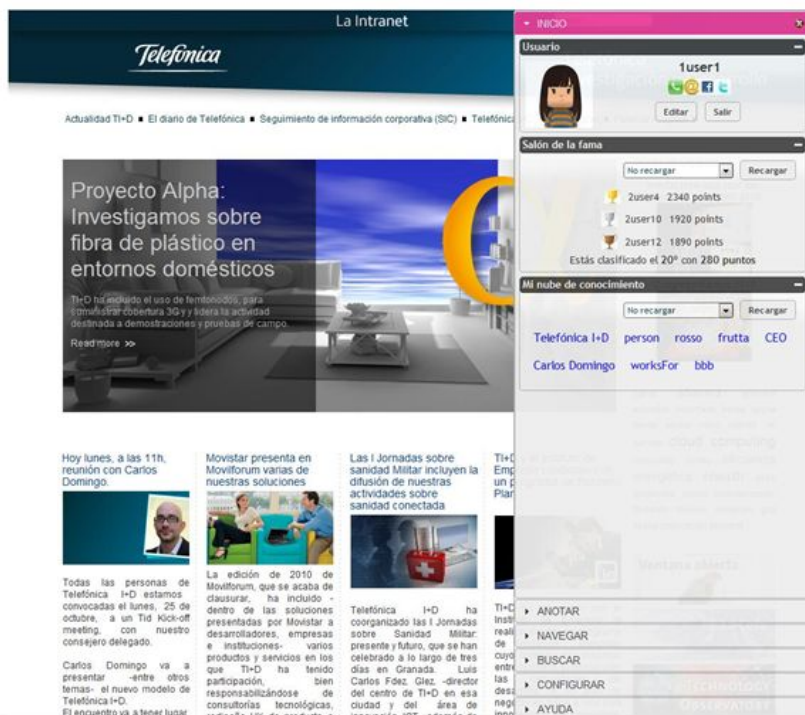


Figura 5.- Vista de la herramienta de anotación semántica tras la anotación.

Como decíamos anteriormente, el proyecto INSEMTIVES proporcionará la metodología, modelos, plataforma y herramientas de anotación semántica que nos permita avanzar hacia una Web cada vez más semántica. **ÁMBITOS DE APLICACIÓN** Es importante advertir que las tecnologías semánticas que permiten dotar de significado a la información son una familia de tecnologías horizontales que pueden utilizarse en los más diversos ámbitos. A modo de ejemplo, la *start-up* austriaca seekda! (miembro del consorcio del proyecto europeo INSEMTIVES) está aplicando esta tecnología en su buscador de servicios web<sup>11</sup> para dotar de una mayor inteligencia y efectividad a las búsquedas. Por ejemplo, en caso de que un desarrollador necesite un servicio de predicción meteorológica podrá introducir como término de búsqueda “weather forecast” obteniendo como resultado servicios en cuyo nombre o descripción aparezcan esas palabras (como, de hecho, ocurre actualmente) así como otros servicios que hayan sido anotados o bien con ese concepto o con conceptos relacionados como pueden ser, por ejemplo, “weather”, “weather conditions”, “atmospheric condition”, etc. Por otro lado, la anotación semántica de los servicios facilita enormemente las búsquedas multilingües ya que sólo es necesario mapear los conceptos en varios idiomas para que cualquiera de las búsquedas anteriores proporcione como resultados exitosos de la búsqueda servicios que hayan sido anotados con conceptos como “meteorología”, “predicción meteorológica”, “clima” o “*météo*” (en francés), “*tempo*” (en portugués), etc. Un segundo ejemplo de aplicación de la anotación semántica de contenidos lo ofrece la productora británica Pepper's Ghost Productions<sup>12</sup> (también miembro del consorcio del proyecto europeo INSEMTIVES) que está aplicando la anotación semántica a la construcción de mundos virtuales. Básicamente, la idea consiste en anotar semánticamente todos los elementos que pueden aparecer en un mundo virtual, que en definitiva no son distintos a los que existen en el mundo real: personas, animales, vegetación, etc. El objetivo no es otro que ofrecer a los desarrolladores de mundos virtuales herramientas más inteligentes durante la construcción de los mismos. Por ejemplo, si el desarrollador está construyendo una granja y desea añadir animales a la misma, el sistema puede sugerirle animales que pueden encontrarse en una granja como gallinas, cerdos, etc., eliminando otros que no aplicarían en este entorno particular como leones, tigres, etc. Por último, en el caso de Telefónica Investigación y Desarrollo estamos aplicando en unos casos, y considerando aplicar en otros, la tecnología de anotación semántica de contenidos en ámbitos muy diversos, entre los cuales es posible destacar los siguientes:

- Portales corporativos: consecuencia directa del caso de uso que lidera Telefónica Investigación y Desarrollo en el proyecto y de aplicación inmediata en todos los portales tanto de uso interno en la casa como desarrollados para terceros.
- Comunidades abiertas como Open movilForum : para permitir a sus usuarios la anotación semántica de los contenidos allí expuestos y típicamente contribuidos por la comunidad.

- Plataforma de ebooks de Telefónica: de próxima apertura, para permitir a los lectores realizar anotaciones semánticas sobre sus libros con vistas a ofrecerles servicios avanzados de recomendación y búsqueda (por ejemplo, encontrar “libros que transcurran en el Siglo de Oro español” o “que transcurran en una capital europea”).
- Imagenio/Mediabox: para permitir a sus usuarios realizar anotaciones semánticas sobre los contenidos ofrecidos con vistas, nuevamente, a ofrecerles servicios avanzados de recomendación y búsqueda (por ejemplo, encontrar “películas que se hayan filmado en Oceanía” o “protagonizadas por un actor irlandés”).
- Etc.

Como puede observarse, las posibilidades son infinitas. **CONCLUSIONES** La promesa de la Web Semántica ha estado entre nosotros desde principios del presente siglo, sin embargo, hasta la fecha no puede decirse que esta promesa se haya hecho realidad. Las causas de este retraso se deben a una estrategia equivocada a la hora de su materialización. De hecho, no debemos pensar en la Web Semántica como una nueva Web, sino más bien como la Web actual renovada o enriquecida con capacidades semánticas. En este sentido, las técnicas de anotación semántica surgen como una opción muy efectiva para añadir una capa semántica sobre el contenido ya disponible, que nos permita ofrecer a sus usuarios servicios más inteligentes a la hora de la explotación de dicha información (búsqueda, navegación y recomendación semántica, por ejemplo). El proyecto europeo INSEMTIVES (<http://www.insemtives.eu>), donde participa activamente Telefónica Investigación y Desarrollo, está estableciendo las bases que permitan avanzar hacia una Web cada vez más semántica, mediante la anotación semántica de sus contenidos. **REFERENCIAS** [1] Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web. Scientific American, 284(5):34-43, May 2001.

<http://www.scientificamerican.com/article.cfm?id=the-semantic-web> [2] Shivashankar Halan, Brent Rossen, Juan Cendan and Benjamin Lok. High Score! - Motivation Strategies for User Participation in Virtual Human Development. Lecture Notes in Computer Science, 2010, Volume 6356/2010, 482-488, DOI: 10.1007/978-3-642-15892-6\_52. <http://www.springerlink.com/content/q85217x144236169/>

**NOTAS AL PIE** <sup>1</sup> [http://www.w3.org/2001/sw/wiki/Main\\_Page](http://www.w3.org/2001/sw/wiki/Main_Page) <sup>2</sup> <http://semanticweb.org/wiki/Ontology> <sup>3</sup> <http://semanticweb.org/wiki/Tools> <sup>4</sup> Nova Spivack es el fundador de Twine (posteriormente adquirido por Evri (<http://www.evri.com/>)), uno de los primeros servicios Web en aplicar las ideas de la Web Semántica al consumo de información por el gran público. <sup>5</sup> Ontología proviene del griego , genitivo del participio del verbo ??, ser, estar, y ó, ciencia, estudio, teoría. <sup>6</sup> La W3C (<http://www.w3.org>) es el organismo encargado de publicar y mantener las especificaciones y estándares que han permitido construir la Web que hoy conocemos. <sup>7</sup> A grandes rasgos, RDFa (<http://www.w3.org/TR/xhtml-rdfa-primer/>) es una especificación de la W3C que extiende el lenguaje HTML (utilizado en todas las páginas Web) para añadir significado a la información expuesta en una página Web siguiendo el paradigma de la Web Semántica. <sup>8</sup> <http://www.insemtives.eu> <sup>9</sup> <http://www.gwap.com/gwap/> <sup>10</sup> <http://www.thefuntheory.com/> <sup>11</sup> <http://webservices.seekda.com/> <sup>12</sup> <http://www.peppersghost.com/> <sup>13</sup> <http://open.movilforum.com/>