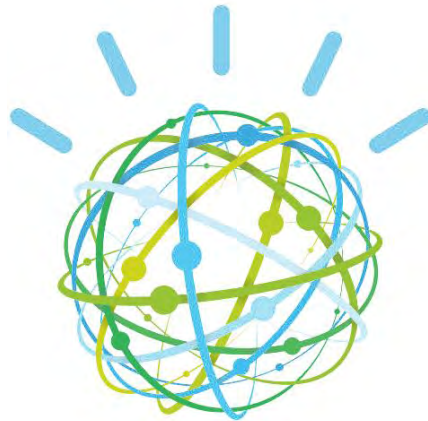


Universidad Nacional Autónoma de México
Facultad de Ciencias Políticas y Sociales



**EL INTERNET DEL FUTURO,
UNA TECNOLOGÍA QUE TE COMPRENDERÁ:
LA APLICACIÓN DE LA WEB SEMÁNTICA.**

Tesina para obtener el título de
Licenciado en Ciencias de la Comunicación
Especialidad: Periodismo
Presenta: **López de Santiago Enrique**
Asesora: Xochitl Andrea Sen Santos
Ciudad Universitaria, Ciudad de México, 2017





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice

INTRODUCCIÓN	1
CAPÍTULO 1. ¿QUÉ ES, DE DÓNDE SURGE Y CÓMO FUNCIONA LA WEB SEMÁNTICA?.....	5
EL ANHELO DE UNA WEB MÁS INTELIGENTE.....	5
ENTONCES ¿QUÉ ES LA WEB SEMÁNTICA?.....	12
PARECE CIENCIA FICCIÓN ¿EN VERDAD ESTÁ ALLÍ LA WEB SEMÁNTICA?	20
CAPÍTULO 2. LA WEB DE LOS CONCEPTOS ¿CÓMO FUNCIONA LA WEB SEMÁNTICA?	31
ENTRE TODOS ESTAMOS CREANDO UNA WEB MÁS INTELIGENTE.	31
PRIMERA HERRAMIENTA DE LA WEB SEMÁNTICA: RDF.....	37
SEGUNDA HERRAMIENTA DE LA WEB SEMÁNTICA: SPARQL.....	40
TERCERA HERRAMIENTA DE LA WEB SEMÁNTICA: OWL.....	46
RDF + SPARQL + OWL = WEB SEMÁNTICA.	50
CAPÍTULO 3. UNA RED LLENA DE DATOS... Y DE PERSONAS.	52
LA WEB SEMÁNTICA ESTÁ CREANDO NUEVAS MANERAS DE ORGANIZAR EL MUNDO REAL.	52
<i>Linked Data</i>	56
<i>Open Data</i>	58
<i>Big Data</i>	66
Web y Social Media.	67
Machine-to-machine (De máquina a máquina).	68
Big Transaction Data.....	69
Biometrics (Datos biométricos).	70
Human Generated (Generados por humanos).....	70
SUENA MUY BIEN, PERO ENTONCES ¿CÓMO PUEDO UTILIZAR TODO ESTO?	71
CAPÍTULO 4. REFLEXIONES Y NOCIONES PARA EL FUTURO DE LA WEB SEMÁNTICA.	78
BIBLIOGRAFÍA.	82
CIBERGRAFÍA.....	84

Introducción

El internet es una tecnología que millones de personas utilizamos prácticamente todos los días para nuestro entretenimiento y trabajo en una enorme gama de usos y formas, no cabe duda que el uso de la web llegó para quedarse en nuestro día a día y en consecuencia se convierte, como desde ya hace tiempo, en un campo de estudio vasto que concierne desde el análisis del uso cotidiano de sus usuarios hasta la investigación de las perspectivas que tienen las personas y organizaciones que se encargan de administrar y crear las reglas detrás de esta tecnología.

En este reportaje me he dado a la tarea de investigar sobre lo que parece ser la nueva era venidera del internet, una serie de protocolos y herramientas que harán de la web algo inteligente, predictivo, lleno de significado y que tiene la posibilidad de adaptarse a los usos y necesidades de los seres humanos, se trata de la web semántica; Esta tecnología que aún está en desarrollo ha evolucionado constantemente desde que comencé a dar seguimiento a la información en torno a ella desde el año 2009, en un inicio las intenciones para la web semántica pretendían que toda la web pudiera alinearse a los nuevos protocolos que establece esta tecnología, en otro momento se pretendió limitar su implementación a proyectos gubernamentales y privados, finalmente el consenso de las principales autoridades de los protocolos de internet llevó a comenzar su implementación en sitios gubernamentales y privados así como su liberación pública para particulares a modo de hacer de la web algo más estructurado y legible para todos los sistemas de cómputo como un protocolo estandarizado.

En la web actual todos los usuarios podemos buscar lo que necesitamos tan solo introduciendo algunas palabras claves en un buscador de nuestra preferencia (google, yahoo, bing, etc.) y como resultado tendremos una gran lista de sitios web y recursos relacionados con las palabras que buscamos, en cambio con la web semántica además podremos buscar por lo que queramos formulando preguntas o frases de cosas muy

específicas, interrogantes del tipo “¿cuántas personas habitaban la ciudad de Bogotá en 1997?”, “¿Qué animales migratorios están cruzando México durante este mes?”, “¿Qué convocatorias hay en este momento para becas sobre ciencia náutica en España?”; parece ciencia ficción pero en este momento todos los datos para contestar a esas preguntas ya están en internet, el problema es que los datos para resolver esas interrogantes están aislados, desconectados entre sí y es complicado que una persona consiga dar respuesta rápidamente a estas preguntas sin antes investigar todos los elementos aislados para dar al final una respuesta concreta, peor aún para un sistema de cómputo que no tiene idea de los elementos que hay que conectar para dar respuesta a estas preguntas, es allí donde entra la web semántica, una tecnología que puede dar sentido a los datos de modo que se puedan interconectar para dar un significado al conjunto de muchos datos que originalmente se encuentran separados.

Cuando una persona utiliza internet son las computadoras las que hacen de intermediario entre la red en la que estamos contactándonos y la información que requerimos, una vez que la información es entregada por la computadora ésta toma un rol pasivo hasta que le solicitemos información nuevamente ¿no sería útil si las computadoras pudieran analizar la información que nos han entregado y pudieran darnos información adicional basándose en los temas que hemos consultado?

La web semántica intenta conseguir búsquedas más inteligentes y menos ambiguas así como conectar enormes bases de datos navegables para su consulta en un solo clic, preguntas tan específicas como “¿Qué jugadores de futbol usan la camiseta número 11, han hecho anotación en algún estadio con capacidad de más de 40 mil asistentes y han nacido en un país de más de 10 millones de habitantes?” pueden ser contestadas en cuestión de milésimas de segundo sin necesidad de buscar toda esa información de manera aislada para luego ser cotejada manualmente, porque con los buscadores que usamos actualmente sabemos que podemos encontrar mucha información fidedigna en toda la web, pero el significado de esa información solo emerge al cruzarla y compararla

con otras bases de datos para resolver cuestiones específicas, para eso y para muchas más cosas sirve la web semántica.

El siguiente reportaje se ha construido a partir diversas fuentes de información que van desde textos académicos, artículos periodísticos, conferencias y entrevistas con personajes que utilizan y trabajan con esta tecnología me han permitido comprender y exponer el alcance de los nuevos protocolos de la web semántica que pretenden hacer del internet algo más inteligente, con más sentido y aún más interconectado para explotar el poder de la información que todos los seres humanos generamos día con día. Entrevistas con personajes como el Doctor David Gondek de los laboratorios de la empresa IBM me permitieron conocer las aplicaciones de punta en inteligencia artificial y web semántica con su sistema WATSON; el investigador en tecnologías digitales, educación y transmedia Hugo Pardo Kuklinski me compartió su visión sobre cómo los individuos percibimos la información y sobre como nosotros mismos nos convertimos en datos estructurados con la construcción de nuestra personalidad y curriculum digital. El periodista y ex editor de la revista WIRED, Ben Hamersley, me permitió ver las dificultades de implementar la web semántica al señalar las dinámicas sociales que surgen en web, la poca madurez que hemos alcanzado los usuarios con el internet y el inicio de la revolución cultural y digital que estamos logrando aún antes de utilizar la web semántica. Por otro lado conocí al Ingeniero Javier Solís González, gerente de INFOTEC y uno de los principales encargados de la implementación de la web semántica en portales del gobierno mexicano y de empresas de talla internacional que operan en nuestro país, también platicué con el divulgador y periodista David Ochoa que es uno de los divulgadores de tecnología y *podcasters* con más experiencia en México y quien me compartió también su visión del uso de esta tecnología y las áreas de negocio que existen en torno a la web semántica.

Por otro lado, este reportaje no puede omitir las explicaciones más técnicas que hay detrás de web semántica, no para ser un manual de uso ni mucho menos para ser una guía práctica para aquellos que desean introducirse de lleno al estudio de esta nueva

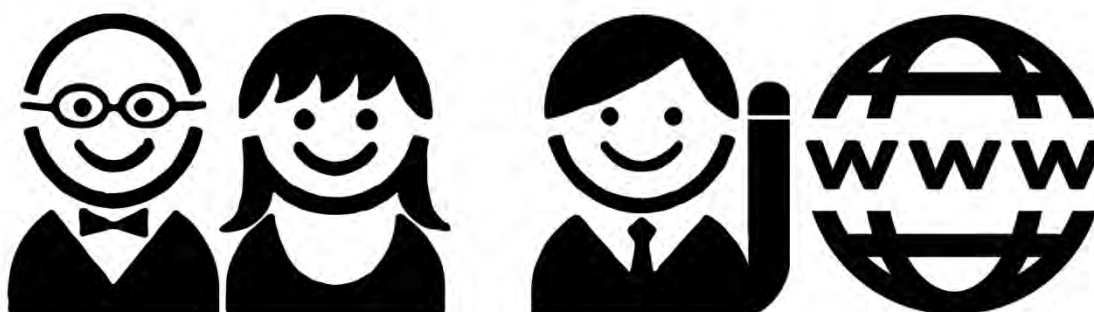
tecnología, más bien la intención de este texto es exponer el cómo las ciencias informáticas y de sistemas están utilizando nuestro lenguaje y su estructura para crear un sistema de significación que pueda servir de puente entre las bases de datos informáticas y la interacción humana; recordemos lo siguiente, las máquinas no son capaces de entender las palabras ni los significados de ellas, para un sistema de cómputo las palabras son simplemente una combinación de caracteres que en dado caso pueden conectar con otras palabras o con documentos, pero nada más. La web semántica pretende crear un sistema donde todas las palabras y los objetos cobren un sentido a partir de su relación con otros objetos y otras palabras, suena complicado y en realidad lo es, por ese motivo he decidido ocupar una sección de este reportaje a diseccionar los lenguajes de programación detrás de la web semántica para así exponer con claridad la forma en la que se está utilizando el lenguaje para un sistema informático que pretende dotar a la web con significados.

El conjunto de estas entrevistas e investigación documental me permite poner en manos de los lectores un texto que puede dar una perspectiva sencilla, pero no por ello simplista, de las innovaciones del internet y de las tecnologías informáticas que vienen para los próximos años así como dar una pequeña muestra de las capacidades que poco a poco estarán en manos de los usuarios que diariamente transitamos el ciberespacio.

Capítulo 1. ¿Qué es, de dónde surge y cómo funciona la web semántica?

El anhelo de una web más inteligente.

Cada día se conectan a la web casi 3 mil millones¹ de personas en todo el mundo, mientras están conectadas todas ellas intercambian datos entre sí, consultan información de una inimaginable variedad de tópicos, comparten sus fotos y videos, conversan, comparten contenidos propios y ajenos, crean todo tipo de archivos, se relacionan con sus amigos, familiares y hasta con desconocidos; estudian, trabajan y se divierten. Esta variedad de actividades deja claro que la web se ha convertido en algo más que una herramienta tecnológica y ha trascendido a ser parte del día a día en la vida de las personas con acceso a ella.



Para 2014 se calcula que en promedio 1 de cada 3 personas tiene acceso a internet en el mundo

Se calcula que cada día se transmiten casi 2 mil millones de gigabytes² de datos en la web derivados de la acción humana y se estima que esta cantidad continuará elevándose año con año por factores como el crecimiento de los usuarios con posibilidades de conectarse a internet, la aparición de nuevos servicios web, el aumento en el uso de teléfonos inteligentes (*smartphones*) así como el aumento en la velocidad de las conexiones a nivel global.

¹“Internet users in real time - Internet Live Stats”. Se calculaban más de 2,930,265,0008 de usuarios. Consultado el 8 julio 2014. <http://www.internetlivestats.com/watch/internet-users/>

²“1 Second - Internet Live Stats”. Se calculaba un tráfico de 1,884,515 TB (1,929,744,000 GB). Consultado el 8 julio 2014. <http://www.internetlivestats.com/one-second/#traffic-band>

En una escala más cercana se calcula que en promedio cada persona utiliza poco más de 650MB diarios tan solo al utilizar correos electrónicos, subir, bajar y consultar imágenes, audio y video, redes sociales, blogs, videoblogs, participaciones en foros o comunidades, juegos online, texto acumulado dentro de los servicios de mensajería instantánea y la lista puede seguir a una cantidad casi incontable de formas de consumo de la web.

Con todas estas cifras es más fácil imaginar la colosal cantidad de información que los usuarios sumamos cada año a la web y aun así es difícil de visualizar la suma que corresponde a toda la información que conforma toda la web en su conjunto.



Si quisiéramos guardar toda la información que se genera en un solo día en internet necesitaríamos un poco más de 30 millones de ipods de 64GB para almacenarla.

Entre esos millones de gigabytes que componen toda la web está casi cualquier información que deseemos consultar y es probable encontrar en ese proceso vínculos a otros temas cercanos o relevantes a nuestra búsqueda; sin embargo, a pesar de la cantidad de información y los buscadores web disponibles, sigue siendo tecnológicamente complicado tener un acceso preciso y confiable entre toda esa maraña de información, es

por eso que es importante exponer algunas creencias o prejuicios que existen sobre la web para comprender esta tecnología con mayor precisión:

1. Es mentira que los motores de búsqueda de la web (como google, bing o yahoo!) entregan la información más cercana, clara y exacta a nuestra búsqueda;
2. Las sugerencias que el buscador devuelve no siempre son las más acertadas;
3. Los buscadores web no pueden discernir y juzgar entre la calidad de un sitio web y otro del mismo tema para entregar el mejor resultado;
4. Los buscadores web no otorgan los mismos resultados a todas las personas aunque su búsqueda sea la misma;
5. No toda la información que contiene la web puede buscarse mediante google;
6. El contenido de internet no está ordenado, clasificado ni se puede acceder a él con facilidad.

La idea de que el internet sea altamente eficiente y ordenado existe prácticamente desde su concepción, se remonta a los años 80's donde la web aún era una tecnología en desarrollo en los pasillos del CERN³ donde fue creado, desde entonces se entendía que el internet, como una red de redes, tendría dificultades para transmitir información entre muchos equipos de cómputo⁴ y por eso se tendrían que crear reglas para poder compartir la información en una red donde bien podrían conectarse computadoras con diferentes piezas de hardware, configuraciones, marcas, sistemas operativos o software.

³ La Organización Europea para la Investigación Nuclear, comúnmente conocida por las siglas CERN, es el mayor laboratorio de investigación en física de partículas del mundo.

⁴ Berners-Lee, Tim, y Mark Fischetti. Tejiendo la red: el inventor del world wide web nos descubre su origen. Siglo XXI de España, 2000. p. 17.

El problema de organizar y encontrar información en la web puede imaginarse como entrar en un enorme súper mercado donde buscamos una sabrosa mermelada de fresa; dentro de este súper mercado seguramente un dependiente (un buscador web, como google, bing, yahoo!, etc.) podrá llevarnos al pasillo donde están las mermeladas y más específicamente a donde están las mermeladas de fresa, pero en este súper mercado ocurre que hay cientos o hasta miles de diferentes mermeladas de fresa, así que el dependiente de la tienda amablemente nos ofrecerá las mermeladas que la gente ha consumido con mayor frecuencia, no es que él sepa si son mejores que el resto de las que ofrece, solo les da prioridad porque son las que la mayor parte de la gente consume.



Algo parecido ocurre en la web, desde el año 2009, Google, el buscador más utilizado en el mundo, implementó uno de sus proyectos más ambiciosos bajo el nombre de “proyecto cafeína (Google Caffeine)”⁵; este consistió en dar relevancia a los resultados de las búsquedas a partir de las palabras que usamos y de la interacción y prestigio que

⁵ Barnett, Emma. “Google reveals caffeine: a new faster search engine”, el 11 de agosto de 2009. Consultado el 10 de junio de 2014 <http://www.telegraph.co.uk/technology/google/6009176/Google-reveals-caffeine-a-new-faster-search-engine.html>

genera cada página que el buscador encuentra de acuerdo a las palabras que los usuarios solicitamos. En realidad Google, o el dependiente de nuestro ejemplo, no sabe qué contiene exactamente lo que está ofreciendo, más bien que se basa en las coincidencias de las palabras de la búsqueda y las ofrece al usuario basándose en qué tanto otros usuarios han consumido las páginas que ofrece.

Si regresamos a nuestro pasillo de mermeladas ocurre que cada día se suman nuevas marcas de mermeladas de fresa, cada una con sus particularidades, claro, algunas serán copias de otras más populares, algunas más serán ligeramente diferentes y otras a pesar de ser ricas y nutritivas podrían estar hasta atrás de las repisas por su impopularidad ¿cómo encontrar e identificar la diferencia entre todas las mermeladas si nuestro dependiente siempre da prioridad a las mermeladas más consumidas y populares basándose en el consumo de los demás usuarios? Es aquí donde entra la web semántica, o algo así como un dependiente que además de conocer dónde se encuentran las mermeladas tiene noción de qué contiene cada una y puede responder a búsquedas más específicas ¿qué tal si necesitamos la mermelada con menos calorías o aquella con menor cantidad de conservadores? O mejor aún, qué tal si buscamos las mermeladas con mayor cantidad de colorante que sean fabricadas en Colombia, que sean propiedad de empresas con menos de 10 años en el mercado y que vengan en empaque de plástico. Bien, es una búsqueda muy específica que nuestro dependiente semántico podría responder.

Dejemos de lado el pasillo de mermeladas y llevemos este ejemplo a la web, imaginemos entonces que buscamos una respuesta sobre algo en extremo específico, preguntemos por ejemplo: ¿Qué jugadores de futbol usan la camiseta número 11, han hecho anotación en algún estadio con capacidad de más de 40 mil asistentes y han nacido en un país de más de 10 millones de habitantes? Seguro si comenzamos a hurgar en las bases de datos de la FIFA, los diarios deportivos más importantes y en foros especializados de aficionados al futbol podremos encontrar por un lado a todos los jugadores que han usado la camiseta número 11; muy probablemente tardemos menos tiempo en encontrar

la información de todos los estadios del mundo con capacidad de 40 mil o más asistentes y en consecuencia podríamos comenzar a separar a los jugadores que han hecho anotaciones en ellos, así por último sólo nos restaría revisar las nacionalidades de nuestro selecto grupo de goleadores para al final contrastarlos con el número de habitantes de su país de nacimiento, así, después de una laboriosa pero exitosa búsqueda, tendremos una lista de jugadores que cumplen con las condiciones para responder a nuestra pregunta⁶. Y claro, no es bueno quedarse con una duda como esa bajo la idea de una supuesta búsqueda en la red, así que aquí está la respuesta con los 10 mejores jugadores que corresponden a nuestra pregunta:

⁶ Conferencia: *Web semántica, la nueva generación de portales y aplicaciones*. Jueves 1 agosto 2013, 16:00 – 17:00. Ponente: Javier Solís González. Gerente de desarrollo de nuevos productos y servicios en el Fondo de Información y Documentación para la Industria (INFOTEC) de CONACYT.

	Jugador	Club	Nacionalidad	Capacidad del estadio	Habitantes del país de nacimiento
1	Cicinho	Real Madrid	Brasil	80354	187560000
2	Gonzalo Fierro	Colo-Colo	Chile	62000	16432674
3	Lukas Podolsky	FC Bayern Munich	Polonia	69901	38536869
4	Mark Gonzales	Liverpool FC	Sudafrica	45362	47432000
5	Michael Thurk	Eintracht Frankfurt	Alemania	52000	82438000
6	Ramón Morales	Chivas de Guadalajara	México	72480	107784179
7	Robin Van Persie	Arsenal FC	Holanda	60432	16336346
8	Stefano Mauri	S.S. Lazio	Italia	82656	58751711
9	Stein Huysegems	Feyenoord Rotterdam	Belgica	51177	10419000
10	Tarek El-Said	Al-Ahly	Egipto	74100	78887007

En realidad para encontrar esta información no se tuvo que buscar en muchas fuentes de información de forma manual para después cotejar los datos y conseguir los resultados, más bien se dio por hecho que la información que buscábamos ya estaba en la web, solo que no se encontraba en la forma que deseábamos, por decirlo de alguna manera la información estaba dispersa; Por un lado tenemos las listas de jugadores, en otro lado los datos de los estadios y en otro el número de habitantes de las naciones del mundo, sólo había que juntar toda esa información aislada para que un mecanismo de la web pudiera darle un sentido encaminado a responder a nuestro dilema futbolístico, es igual que nuestro tendero del pasillo de mermeladas de fresa que nos puede dar mermeladas muy específicas si se las sabemos pedir, de esto se trata la web semántica, aunque claro, con millones de pasillos de todo lo que nos podamos imaginar.

Entonces ¿Qué es la web semántica?

Si en la web actual podemos buscar lo que necesitamos tan solo introduciendo algunas palabras claves, entonces en la web semántica podremos buscar por lo que queramos formulando preguntas o frases.

La web semántica se trata de una web más inteligente y práctica para los usuarios, es un mecanismo de la web que permite dar significado a la información que allí se encuentra y que pretende simplificar tareas de búsqueda de información que pueden resultar tediosas y frustrantes. Con la web semántica se trata de utilizar las enormes cantidades de datos que se generan diariamente y darles significado para obtener respuestas rápidas y sencillas a preguntas que podrían demorarnos mucho tiempo buscando manualmente, se trataría entonces de tener una web que otorgue resultados a las búsquedas mejor definidos, más exactos y específicos⁷.

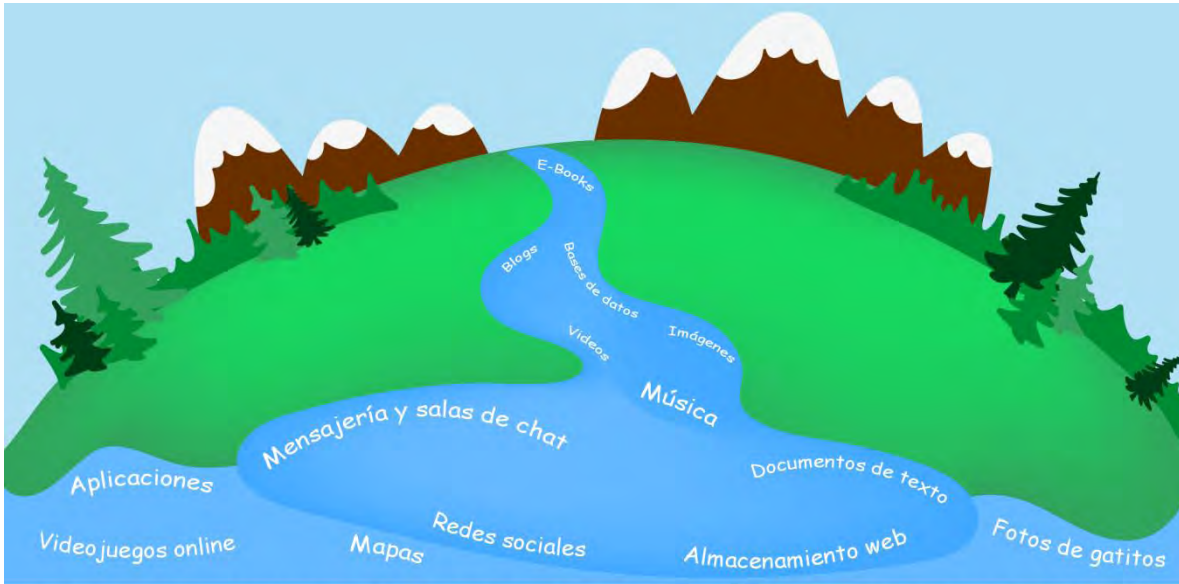
⁷ Berners-Lee, Tim, y Mark Fischetti. *Tejiendo la red: el inventor del world wide web nos descubre su origen*. Siglo XXI de España, 2000. p. 145.

La web en la actualidad es utilizada todos los días por millones de personas, en ella se tiene la ventaja y problema de una sobresaturación de datos que muchas veces se acumulan en la web sin un orden o significado específico, por lo que es difícil que otros usuarios puedan acceder a ellos; con la web semántica se pretende dotar a toda la información de la web con identificadores que le permitan organizarla para hacerla más accesible.

El desarrollo de internet ha generado que las personas utilicemos día a día esta herramienta para millones de actividades, desde la forma en que nos comunicamos, hacemos negocios o realizamos nuestro trabajo, de esta manera generamos y utilizamos recursos sin importar nuestra geografía o idioma, este uso cada día más generalizado ha creado uno de los problemas más grandes que tiene la web, se trata de la sobrecarga de información; El problema de la sobrecarga de información consiste en tener cantidades colosales de información generándose sin parar a cada segundo, esto causa lo que se conoce como un problema de heterogeneidad de fuentes, lo que quiere decir que a la web llega mucha información generada por diferentes tipos de computadoras y programas lo que causa un problema de interoperabilidad, lo que quiere decir que es casi imposible que todas las computadoras puedan hacer uso de toda información existente en la web debido a que mucha de ella es diferente tanto en su contenido como en su formato.

Los problemas de la web actual los podemos resumir brevemente:

Sobrecarga de información: Quiere decir que siempre se está generando nueva información, es mucha y cada vez más de nuevos tipos, formas, tamaños y formatos, de esta manera es cada vez más difícil controlar toda la información que se genera en la web eficientemente.



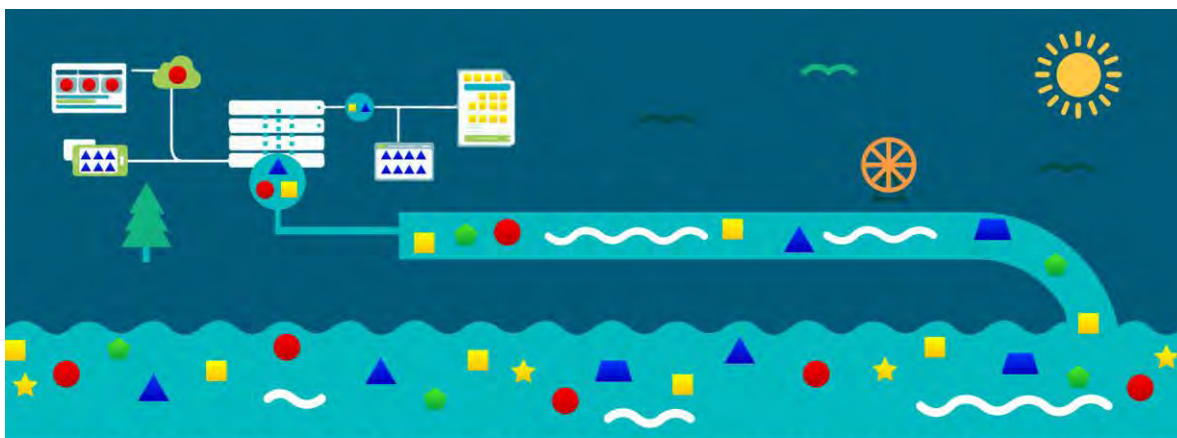
La sobrecarga de información es como un río muy caudaloso donde siempre fluye nueva información y que aumenta su caudal con nueva información de diferentes tipos

Heterogeneidad de fuentes: Quiere decir que toda la información que llega por la sobrecarga de información, además de ser información distinta entre sí es generada por computadoras o dispositivos muy distintos entre ellos.



El problema de la heterogeneidad es como si juntáramos mucha agua de diferentes fuentes, cada una con sustancias distintas que no necesariamente se combinan entre sí.

Problema de Interoperabilidad: Quiere decir que hay tanta información generada por tantos equipos diferentes que es difícil que la web pueda darle lectura, organización y sentido a toda la información que se genera a cada segundo en el internet.



Interoperabilidad es como si tu computadoras solo pudiera leer información de formas redondas, cuadradas y triangulares, mientras que en todo el caudal hay información de muchas formas distintas que son incompatibles con tu equipo

Con ayuda de la web semántica se pretende resolver los problemas de interoperabilidad y heterogeneidad de fuentes al hacer que la web logre procesar toda esa información y contenido, más adelante platicaremos sobre los métodos y protocolos con lo que se pretende que la web pueda interpretar la información, combinarla, realizar deducciones lógicas y lograr así resolver búsquedas automáticamente a partir de toda la información que obtiene de muchas fuentes de información de manera simultánea.

En su libro Tejiendo la red, Tim Berners-Lee, quien es considerado como el creador del internet, cuenta sobre el origen de la web y el objetivo que persigue la web semántica: “... los aparatos de búsqueda han demostrado ser muy útiles para combinar largos índices rápidamente y para encontrar oscuros documentos. Pero han demostrado ser notablemente inútiles, (...) porque no tienen modo de evaluar la calidad de un documento.

“... El problema es que los aparatos de búsqueda suelen limitarse a ver la existencia de palabras en documentos; algo que es una pista, pero que nos dice muy poco de lo que realmente dice o trata el documento.”⁸

De todo esto trata la web semántica, una tecnología que los usuarios podemos entender como una mera actualización a las herramientas de búsqueda que utilizamos habitualmente; sin embargo, la tecnología de la web semántica representa una posibilidad enorme llena de dificultades para los desarrolladores y creadores de la web de todo el mundo.

Los problemas que acompañan a la web semántica van más allá de un apartado técnico que ingenieros y técnicos deban resolver, también se trata de un problema sobre el lenguaje y la forma en la que los seres humanos nos comunicamos, hay que recordar que incluso las personas que hablamos un mismo idioma podemos llegar a tener problemas para comunicarnos ya sea a causa de palabras que surgen como parte de una moda, por causa de un uso regional que hace confusas algunas palabras o frases entre personas de distintas regiones, todo esto sin sumar la propia acción natural de mutabilidad del lenguaje, es decir, el simple hecho de que con el tiempo las palabras comienzan a cambiar tanto su forma de uso como la frecuencia con que se utilizan; Hablar de la web semántica no es solo el hecho de organizar cantidades colosales de información y tratar de dales significado, también se trata de una tecnología que trabaja con el uso del lenguaje y la manera en que las personas lo utilizamos para comprender e interactuar con el mundo de la información.⁹

En entrevista con el ingeniero Javier Solís González, gerente de desarrollo de nuevos productos y servicios en el Fondo de Información y Documentación para la

⁸ Berners-Lee, Tim, y Mark Fischetti. *Tejiendo la red: el inventor del world wide web nos descubre su origen*. Siglo XXI de España, 2000. p. 166.

⁹ Bush, Vannevar. “As We May Think”. *The Atlantic*, julio de 1945.
<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

Industria (INFOTEC) de CONACYT¹⁰, comentó que para comprender qué es la web semántica es importante tener en claro cómo ha evolucionado la web en sus eras o fases más importantes.

Primero está la web 1.0, en ella las personas se conectaban a la web y solo podían consultar la información que otros ya habían puesto allí, en general en la web 1.0 no se podía interactuar ni compartir a menos que se tuviera un nivel avanzado de conocimiento en el uso de computadoras, para el grueso de los usuarios simplemente se podía consultar información; Durante esta etapa simplemente había usuarios que consumían la información y administradores que subían cosas a la web para su consulta.



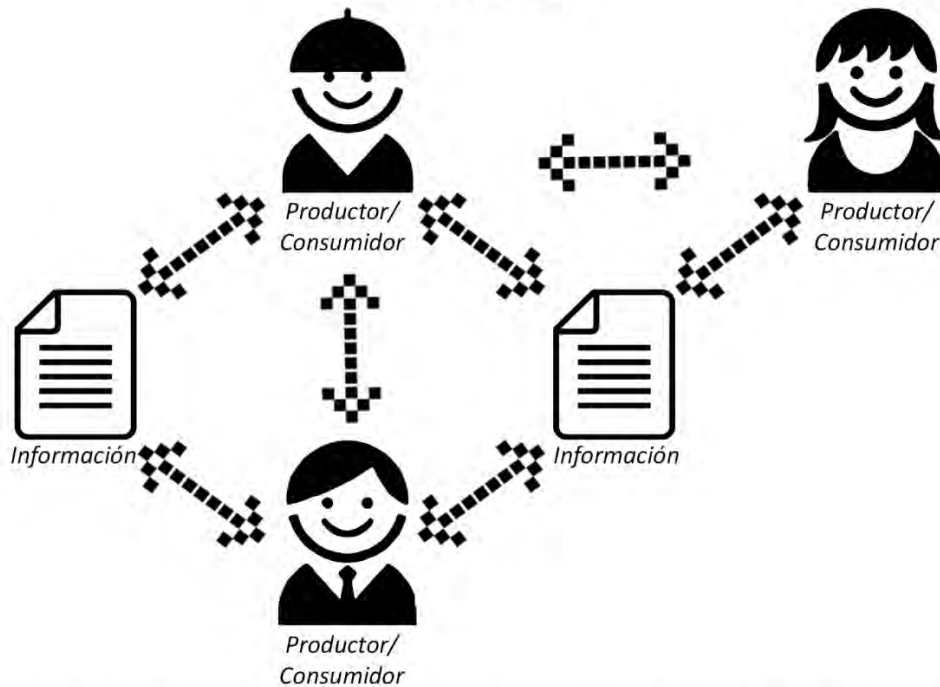
En la web 1.0 la información circulaba de los productores de la información hacia los consumidores. Los consumidores no podían interactuar de vuelta con el contenido o su productor.

En la web 2.0 las cosas cambiaron, las personas podían ser participantes de la web, ya no eran necesarios muchos conocimientos técnicos para poder abrir un blog, un foro, una página sencilla o un wiki, es aquí donde los usuarios comienzan a tener la oportunidad de ser más participativos, de crear, conversar y formar comunidades.

La web 2.0 la podemos definir como la era donde todos los usuarios son capaces de modificar la web.

¹⁰ Entrevista realizada en el centro de convenciones Expo Zapopan durante el evento Movistar Campus Party México 2014 #CPMX5, 30 de julio 2014, Zapopan, Jalisco, México.

LA WEB 2.0

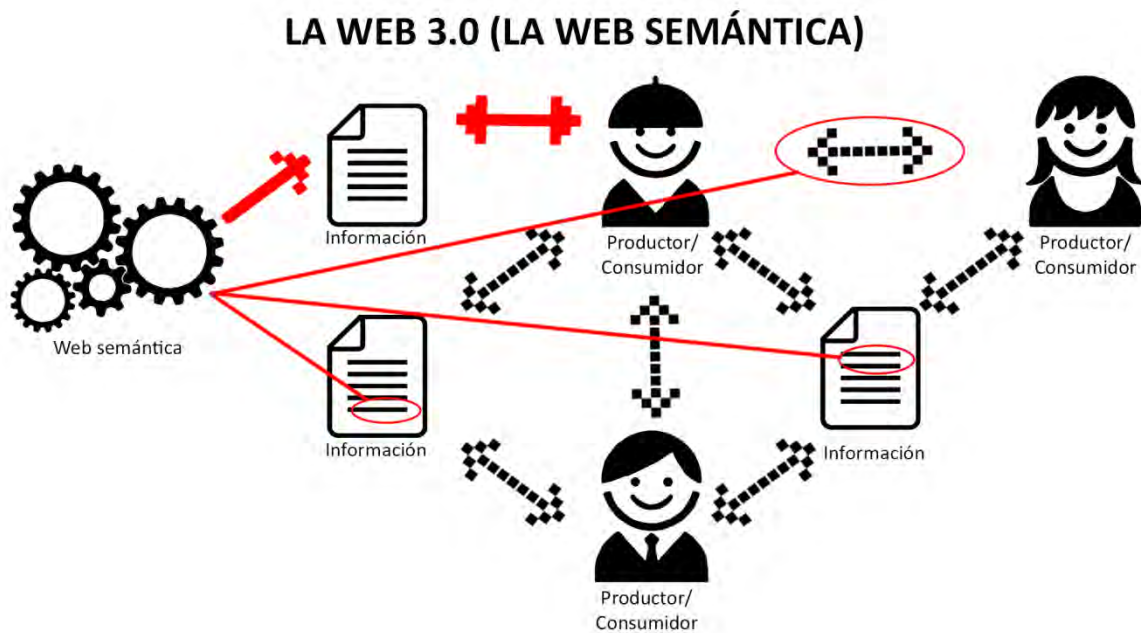


En la web 2.0 los usuarios pueden ser productores y consumidores de información. Todos los usuarios pueden manipular la información, compartirla, modificarla y crearla.

La idea con la web 3.0 es que ahora las máquinas ayuden a los usuarios a buscar, crear y modificar información; En esta etapa de la web, las herramientas informáticas ayudarán a los usuarios en todo momento y serán generadoras de información, los usuarios se comunicarán con otros usuarios, las máquinas con los usuarios y las máquinas también se comunicarán con otras máquinas para generar información, no como antes en la web 2.0 o web 1.0 donde únicamente los humanos podían generarla y consumirla. Así en la web 3.0 la web semántica aparece como parte de los mecanismos para lograr una experiencia más inteligente, útil y versátil para los usuarios.

Al seguir con la conversación, Javier Solís hizo énfasis en la diferencia que debería de existir entre la web 2.0 y la 3.0; En la primera, la web está ordenada por datos y trozos de información que solo nos dan palabras, esas palabras tienen sentido para las personas que las buscamos pero en realidad no tienen mucho sentido para las máquinas, mientras

que en la web 3.0 la intención es que la web no esté llena sólo de palabras, sino de conceptos bien definidos que sean comprensibles por igual para las personas y para las máquinas de modo que éstas puedan relacionar la información, compararla y de alguna manera comprenderla bajo un contexto bien definido.



En la web 3.0 las máquinas podrán generar información, modificarla y ayudar a distribuirla a otros usuarios o computadoras. La web semántica trata de potenciar y mejorar la experiencia de los usuarios de la web.

Con todo lo anterior es fácil pensar que la web semántica es una especie de inteligencia artificial o algún tipo de magia que le permite a las máquinas comprender los conceptos y palabras de los humanos, pero en realidad tan sólo se trata de una tecnología que trata de resolver problemas bien definidos a partir de resolver operaciones bien estructuradas al utilizar datos de la web que se encuentren bien delimitados.

Todo esto describe a la web semántica, que es una tecnología que no solo se resume a la búsqueda de información sino que también se resume en aplicaciones con la capacidad de mejorar otras herramientas y servicios existentes; Asuntos pertinentes al

gobierno, el comercio, la enseñanza, la cultura, el turismo, la investigación científica y hasta el entretenimiento pueden beneficiarse de la gran estructura de la web unificada semánticamente.

Parece ciencia ficción ¿en verdad está allí la web semántica?

Las ventajas de una web semántica también pueden aplicar al entorno social, si pensamos en la web como un punto de reunión público entonces la web no solo es un espacio de intercambio de información, también es un espacio de intercambio entre personas y organizaciones, como una especie de enorme plaza pública digital. En ese entorno, de darse el caso de una web semántica masificada, habría posibilidades de encontrar y posicionar personas de acuerdo a sus capacidades y conocimientos de manera más abierta y pública, además gracias a una web semántica esas relaciones humanas podrían organizarse de manera estructurada y podrían también ser potenciadas al hacerles llegar datos e información relevante a partir del espectro social al que pertenecen.¹¹

La existencia de la web como un espacio social y de la web semántica como una herramienta capaz de procesar información acerca de la propia interacción humana en la web hace pensar que la humanidad está entrando en un estado donde nunca antes, al menos de eso habla Ben Hammersley.

Ben Hammersley es uno de los periodistas pioneros sobre el tema de internet para el diario británico *The Times*, ha sido reportero para *The Guardian* y también es reconocido por su labor como director general en la revista *Wired* de la editorial Condé Nast, hasta la fecha es reconocido por ser uno de los más influyentes exponentes en el periodismo tecnológico. Al platicar con él durante su visita a Campus Party en 2010 en la ciudad de México¹² hizo énfasis en que para entender las tecnologías de la web 3.0 y las

¹¹ Tecnológico, INFOTEC Centro Público de Investigación y Desarrollo. Consultado el 18 de agosto de 2014. http://infotec.com.mx/en_de/infotec/articulo_web_semantica

¹² Entrevista realizada en el centro de convenciones Expo Santa Fe durante el evento Movistar Campus Party México 2010 #CPMX2, 11 de agosto 2010, Ciudad de México, México.

posibilidades de la web semántica era necesario ser conscientes sobre la era del postdigitalismo y la transformación global que ha generado la masificación de internet “La humanidad nunca había estado tan conectada, estamos en una nueva era, como si hubiéramos descubierto el fuego o la rueda; las ventajas de la nueva era son grandes, pero no todos las comprenden, algunos por la falta de acceso, otros porque no entienden la dinámica de un mundo sin fronteras donde todos podemos participar”.

Al conversar con Ben sobre las posibilidades de la web semántica como algo tangible en el internet de uso común, él dejó claro que aunque la web semántica se utiliza en nichos muy avanzados dedicados al desarrollo de la web difícilmente podría implementarse a toda la web a corto plazo. “Las computadoras nos han ayudado a crear una gran red de información, pero su incapacidad para organizar y comprenderla no solo se debe a incapacidades técnicas para solucionar el problema, también se debe a que no hemos terminado de comprender cómo funcionan los flujos de información de la web, por tanto aún es difícil crear un sistema semántico que le pueda dar sentido a toda la información que ya existe en la web, sin contar la información que está surgiendo en tiempo real a cada segundo y mucho menos podemos prever del todo el futuro cercano del internet para poner en marcha un mecanismo que logre interpretar todas las posibilidades que hay allá afuera”.

Ante esa respuesta era inevitable preguntar si entonces la web semántica era un esfuerzo sin sentido que pretendía abarcar más de lo que era posible, ante esto Ben comentó que la web semántica no era una fantasía ni mucho menos algo imposible, al menos no en el largo plazo, por el momento para él era más importante analizar los tres cambios fundamentales que ha traído el *post digitalismo* a la humanidad para a partir de ello conocer la mejor ruta de acción para poder implementar mecanismos semánticos a mayor parte de la web: “Lo primero que debemos entender es que la humanidad está pasando por cambios fundamentales gracias a tres puntos que hasta hace un par de

décadas no existían, primero tenemos la libertad de comunicación, luego el *feedback*¹³ y al final el rompimiento de las distancias”.

Sobre el primer punto, acerca de la libertad de comunicación, Hamersley ejemplificó que hoy en día la gente puede abrir una cuenta en cualquier servicio web e interactuar con otras personas, se puede escribir un blog con lo que sea que pienses, puedes crear videos donde puedes hacer prácticamente cualquier cosa o expresarte mediante un show en vivo en audio o video con una facilidad impresionante, recalca que algo que hoy nos parece normal, hace siquiera 20 años, aún parecía algo que solo los más experimentados, poderosos o ricos podrían hacer, eso hoy ya no es así y esa facilidad para comunicar cosas es un cambio radical en comparación con cualquier otra era de la humanidad.



En la era postdigital hay nuevos poderes de todo y para todo. Gracias al internet los usuarios tienen nuevas habilidades que antes solo eran posibles para los más poderosos y ricos.

¹³ La capacidad de un emisor para recoger reacciones de los receptores y modificar su mensaje, de acuerdo con lo recogido.

reunión además sería demoroso, pero resulta que ahora ya no tenemos ese problema, las distancias físicas ya no importan para poder comunicarte con otras personas, eso también es nuevo.”

En cuanto a la web semántica Ben Hamersley señaló las posibilidades de su uso inmediato en el ramo de las *Post digital Politics* (políticas post digitales): “hay grandes problemas que podemos resolver con nuevas herramientas, con la web y la web semántica cambian las maneras de hacer las cosas”, el mejor ejemplo es que los gobiernos tiene cantidades enormes de información bien organizada y delimitada que podría administrarse mediante tecnologías semánticas, si los gobiernos se ponen en marcha es posible generar información que ayude a encontrar soluciones a problemas que hoy nos son muy complicados.



Con la web cambia la manera de hacer las cosas, gracias a la velocidad, al rompimiento de las distancias y a la cooperación masiva se pueden dar soluciones novedosas y atacar nuevos problemas.

Hammersley compartió su idea de que en el post digitalismo vivimos dentro de una revolución donde estamos acompañados una generación que cambia los paradigmas de todo lo que se ha hecho antes: “En la era post digital somos como revolucionarios, pero sin rifles ni sangre corriendo por todos lados, lo importante es que el efecto es el mismo, estamos generando y empujando a un cambio en las maneras de hacer y pensar.”

Para Hamersley apenas estamos aprendiendo del potencial de tecnologías como la web semántica, además falta mucho por hacer para implementarla como una tecnología generalizada en toda la web, “si eso fuera una realidad ya tendríamos que ser conscientes del precio de estar conectados, lo que nos llevaría primero a tener que entender los problemas de la web para a la vez participar en resolver los problemas del mundo, estar conectados y romper fronteras significa también que somos parte de los problemas del mundo, debemos apoyar a resolverlos, es nuestra obligación. Sí, eso es duro, pero es la verdad”.

Tras conversar con Ben Hammersley quedó claro que para comprender este asunto de la web semántica había que ahondar más en asuntos del *post digitalismo* y comprender aún más el contexto en el que se planea desarrollar una tecnología que intenta comunicar y administrar la información de máquinas a usuarios y de máquinas a máquinas en un entorno donde la existencia de la web como espacio social servirá de puente para la realización de la web semántica. Para esta tarea fue oportuna la investigación del Doctor Hugo Pardo Kuklinski, quien es CEO y fundador de CampusMovil.net, miembro del Laboratorio de Mitjans Interactius (LMI) de la Universidad de Barcelona, es también profesor del Departamento de Comunicación Digital de la Universidad de Vic, además de ser el productor de digitalismo.com¹⁴ además de ser autor de diferentes libros sobre la era digital, post digitalismo y economía en la era digital.

En una breve pero muy concisa video llamada¹⁵, el doctor precisó que los esfuerzos para alcanzar la web semántica son constantes y cada vez más expertos se interesan en colaborar con la W3C¹⁶ para lograr un internet donde podamos dar un mejor uso a la colosal cantidad de información que se genera cada día, al mismo tiempo se suman esfuerzos para dar un uso práctico a esta tecnología para crear buscadores que sean

¹⁴ “Digitalismo.com”. <http://digitalismo.com/>

¹⁵ Entrevista realizada vía Skype a Barcelona, España, 17 de febrero 2014, Ciudad de México, México.

¹⁶ W3C es el consorcio dedicado a establecer los protocolos y estándares de la web y es uno de los organismos con mayor autoridad respecto a lineamientos y herramientas para generar una web ordenada, estandarizada y de libre acceso.

intuitivos para el grueso de los usuarios al mismo tiempo que grupos de investigación se suman para obtener una mejor comprensión de lo que ocurre en la web mediante herramientas semánticas.

El Doctor Kuklinski comentó que él y su equipo han encontrado algunas características en los usuarios que pueden dar pistas al momento de desear implementar tecnología semántica a la información que se genera en internet en tiempo real “no solo se trata de catalogar y darle sentido a la información, también existe la variable sobre el tiempo, es decir, sobre el momento en que es relevante, difícilmente podemos decir que lo que se consume hoy se consumirá mañana o en un mes, tampoco sabemos cómo medir la eficacia de los recursos en línea para poder separar el ruido de lo relevante”.

El Doctor señaló a la investigación de Mayer-Shönberger¹⁷, quien afirma que la tecnología invirtió nuestra relación con la memoria: “En la historia, las tecnologías favorecieron el olvido, pero la digitalización nos agobia con el volumen de información a nuestro alcance y ha eliminado la amnesia, por lo que se recomienda a las personas y a las plataformas web que aprendan a olvidar para ir eliminando la información irrelevante”.

Otro de los puntos de la investigación del Doctor Kuklinski trata sobre la creación de curriculums personalizados en la web, esto quiere decir que los usuarios cada vez tienden más a utilizar los instrumentos digitales al alcance para crear archivos multimedia, escribir ensayos, compartir conocimientos relevantes, entre otras cosas con el fin de aportar algo de conocimiento social y ser partícipes de las grandes conversaciones de la sociedad mediante la web. De esta manera surge la creación de un perfil único y personal que puede dar prueba del mosaico de identidades que crean los usuarios a partir de su interacción con la web, “en este sentido tenemos herramientas que permiten a los usuarios crear perfiles que muestran sus títulos o credenciales profesionales, sin embargo

¹⁷ Kuklinski, Hugo Pardo. *Geekonomía: un radar para producir en el postdigitalismo*. Edicions Universitat Barcelona, 2010. pp. 129.

lo ideal sería que con la web semántica pudiéramos concentrar currículums más específicos basados en la participación de los usuarios en la web, en sus iniciativas en causas sociales y en su identidad única como ciudadanos, a final de cuentas hoy día muchos de los acercamientos a otras personas son primero en la web antes que en cualquier otro lado por lo que para los usuarios será necesario construir una identidad basada en la web y en la administración personal de los datos personales para permitir expresar la identidad y capacidades individuales. Construir eficientemente esa identidad puede llevar años, pero es una tarea necesaria”.

Como complemento se agregó la importancia de la creación de historias *transmediáticas*, en ese sentido debe entenderse que cada vez es más recurrente la creación de contenido creado en formatos distintos con la intención no de recrear o competir, sino de complementar y facilitar la comprensión de temas amplios y/o complejos de todo tipo y temática: “La web semántica tendría entonces que poder analizar audios, videos, textos, imágenes y sitios web de manera tan eficaz que debería poder reconocer el contenido, su significado, su autor y la relevancia que cada una de esas piezas tiene con respecto a sus contrapartes y al resto de la información del mismo tema que circule en la web”. Así aunque todo esto tiene un valor agregado muy poderoso aún representa dificultades importantes al momento de cómo debemos plantear la información y los mecanismos semánticos para que tanto máquinas como humanos podamos utilizar la información eficazmente.

Los usuarios de la web también están desarrollando fenómenos que antes, aunque existían, no eran tan frecuentes, tal es el caso de la cultura del *remix* que ha resultado en una aceleración en los procesos de aprendizaje mediante el auto aprendizaje guiado y cognitivo que se mezcla con los conceptos y prácticas en ambientes formales como el trabajo o los centros de estudios como las universidades, “ese trabajo de disgregación y mezcla ha demostrado proporcionar mayores conocimientos a los usuarios especialmente

cuando estos adoptan la filosofía *pirata*¹⁸, pero sin hacer trampas y sin confundir la intertextualidad con el *copia y pega*¹⁹, de esta manera, afirma Kuklinski, se fomenta la creatividad de tal manera que la cultura del remix y la filosofía pirata potencian el aprendizaje colateral y las capacidades de innovación.



Filosofía pirata: Si la información se encuentra en la web entonces tiene el potencial de tomarse, recrearse y convertirse en algo nuevo sin ninguna restricción.

Con estas ideas en mente queda claro que hoy día estas prácticas no solo van de la mano de los usuarios experimentados, sino que cada vez se hacen más comunes en la mayoría de los usuarios ¿quién no ha bajado de internet alguna imagen para después usarla en alguna tarea o simplemente para ilustrar un texto o darla a otra persona a modo de regalo? Bueno, esas interacciones con la web tienen un impacto que la web semántica pretendería analizar para interpretar su significado o el nuevo significado que se le está dando, es por eso que es importante entender las características de los usuarios que hacen uso de la web pues dejan entrever que la información no sólo sube a la web como algo estático, sino que tiene el potencial de recrearse, transformarse y hasta resignificarse.

¹⁸ Hace referencia a la libertad para tomar recursos libremente de la web con el fin de compartirlos y crear compilaciones o repositorios de temas específicos para después crear o recrear nuevas obras derivadas sin fines de lucro.

¹⁹ Stallman, Richard. *Los piratas son los padres. Historias en los albores de la era digital*. Exgae, Barcelona, 2008. http://exgae.ney/docs/Los_piratas_son_los_padres.pdf

Así el doctor Kuklinski compartió su idea sobre no temer al conocimiento volátil, las contradicciones ni a la construcción de ideas dentro de la web, “estamos en los tiempos de la barbarie digital, la información fluye y los usuarios pueden hacerse de conocimientos increíbles con solo estar frente a la computadora si se lo proponen, con un conocimiento autodidacta y una constante generación de nuevos contenidos tanto los usuarios de la web como los responsables de establecer los mecanismos de una web semántica deben trabajar con los mismos problemas, en primer lugar ambos tienen que enfrentarse constantemente a tratar de resolver problemas mal definidos, ambos grupos tienen que enfrentarse también a tolerar la ambigüedad de la información y a trabajar en medio de la incertidumbre”.

En ese sentido y para finalizar, el doctor citó la investigación de Bain (2007)²⁰ donde afirma la necesidad de ser clara y explícitamente conscientes de las lagunas de información disponibles. “Es imperativo reconocer que en algunos problemas se tiene que llegar a extraer una conclusión o se toma una decisión en ausencia de la información completa siendo capaz de tolerar la falta de certeza”.

Ante las reflexiones y opiniones que compartieron Ben Hamersly y el Doctor Kuklinski se hace entrever a la web como una herramienta que se está redefiniendo constantemente entre los usuarios como un mediador entre las necesidades de consumo de información y la información en sí misma; es difícil entonces no escuchar el eco lejano de Marshall McLuhan, uno de los teóricos clásicos de la comunicación, quien afirmaba que los medios crean nuevos patrones sociales que reestructuran las percepciones, más allá de los contenidos que emita el medio este tiene su propio efecto en la percepción humana y la reconfigura²¹ por lo que es fácil pensar en la web como un medio que ha penetrado en la sociedad y que efectivamente ha ayudado a reconfigurar la percepción de la realidad y de las posibilidades que tenemos de manera inmediata como usuarios. Así al mismo

²⁰ Bain, Ken. 2007. *Lo que hacen los mejores profesores universitarios*. Universitat de Valencia. Valencia.

²¹ McLuhan, Marshall. 1964. *Understanding media: The extensions of man*. McGraw Hill. Nueva York

tiempo la web semántica se presenta como una tecnología o una actualización para la web con el potencial de generar nuevas dinámicas en cuanto a la mediación de la información ¿será entonces que McLuhan tenga razón y que gracias a la mediación de un medio, como ahora lo es la web, se generen las condiciones para causar un efecto en la percepción humana y la reconfiguren? Eso sólo el tiempo nos lo dirá.

Capítulo 2. La web de los conceptos ¿Cómo funciona la web semántica?

Entre todos estamos creando una web más inteligente.

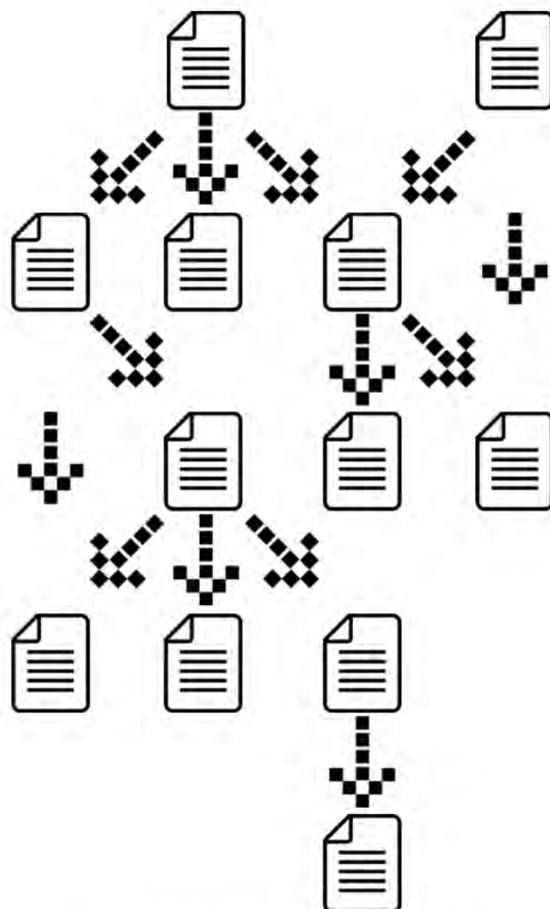
Cuando una persona utiliza internet son las computadoras las que hacen de intermediario entre la red en la que estamos contactándonos y la información que requerimos, una vez que la información es entregada por la computadora esta toma un rol pasivo hasta que le solicitemos información nuevamente ¿no sería útil si las computadoras pudieran analizar la información que nos han entregado y pudieran darnos información adicional basándose en lo que hemos consultado?

Para que eso pudiera suceder la información de la web tendría que estar en un formato que cualquier máquina pudiera leer de manera natural, eso o al menos tener alguna forma de convertir esa información a un formato comprensible universalmente por las máquinas, de esa manera podríamos tener una red donde toda la información podría ser procesada directa o indirectamente por cualquier computadora.

El ejemplo es sencillo al recordar las veces en que un usuario al buscar por alguna palabra en su buscador web preferido, se encuentra con una lista de miles de resultados con sitios web que contienen la palabra que solicitó y que al verificar los resultados se da cuenta que muchos de los resultados que se recibieron son irrelevantes y no tienen mucho sentido con lo que se estaba buscando realmente, esto es porque “los servicios de búsqueda web a pesar de contar con cada vez más elementos para hacer predicciones y delimitar la información siguen limitándose a ver la existencia de palabras en documentos; algo que es una pista, pero que dice muy poco de lo que realmente dice o trata la página o el documento.”²²

²² Berners-Lee, Tim, y Mark Fischetti. *Tejiendo la red: el inventor del world wide web nos descubre su origen*. Siglo XXI de España, 2000. Pp. 163.

Si sumamos que además la web se ha convertido en un espacio social en crecimiento entonces tenemos que además de páginas web y documentos hay usuarios cuyos perfiles o avatares delimitan su relación con otros usuarios, foros y fuentes de información; la web entonces encontraría a los usuarios como fuentes particulares que consumen y generan información bajo una estructura personalizada y con interrelaciones con el resto de la web, dicho de otra manera, la web podría escribir y consultar un mapa del perfil de los usuarios en la web para poder ofrecerles información más precisa basándose en la información y comportamiento previo del usuario.



En la web actual la información hace referencias de un documento a otro de forma lineal, de esta manera los mecanismos de la web no tienen idea de qué trata la información de donde vienen ni idea de qué trata la información a donde van.

Estos análisis no pueden ser completamente automatizados actualmente, en parte porque no hay una herramienta que pueda englobar toda la información de un usuario en cuenta entra a la red y por otro lado porque muy poca información de la web se encuentra en una forma que pueda ser analizada por una máquina para darle un significado.

Podemos decir que la web está diseñada para que los humanos podamos consultar información fácilmente del modo en que está expuesta por lo que podemos interpretarla y utilizarla sin mayor problema, pero éste no es el caso para las máquinas que la tienen difícil para saber qué es lo que están mostrando o qué relación tiene esa información con cualquier otra existente en la web.

Hay que poner atención en que servicios como Google, Facebook, Stumble Upon o Twitter que han hecho esfuerzos por crear dentro de sus propios servidores tecnología semántica que ayude a organizar la información que se genera dentro de sus redes tanto para mejorar la experiencia de los usuarios como para obtener información sobre el comportamiento y las tendencias creadas por la acción de los mismos; sin embargo, cuando hablamos de web semántica se debe pensar en mecanismos que puedan ser aplicables a la totalidad de la web y no solo a redes específicas como es el caso de los servicios mencionados.

Para crear una web semántica que pueda ser útil para la llegada de la web 3.0 el Ingeniero Javier Solís, gerente de desarrollo de nuevos productos y servicios en el Fondo de Información y Documentación para la Industria (INFOTEC) de CONACYT, comentó que el principal paso para lograr una web semántica funcional es que hay que cambiar la forma en que se organiza la web y pasar de una web de datos y documentos a una web de conceptos y relaciones entre conceptos; “la web semántica ya no se trata de una web llena de hipervínculos que llevan de una web a otra, sino de una web de conceptos bien definidos y de relaciones bien definidas entre los conceptos ¿para qué queremos hacer

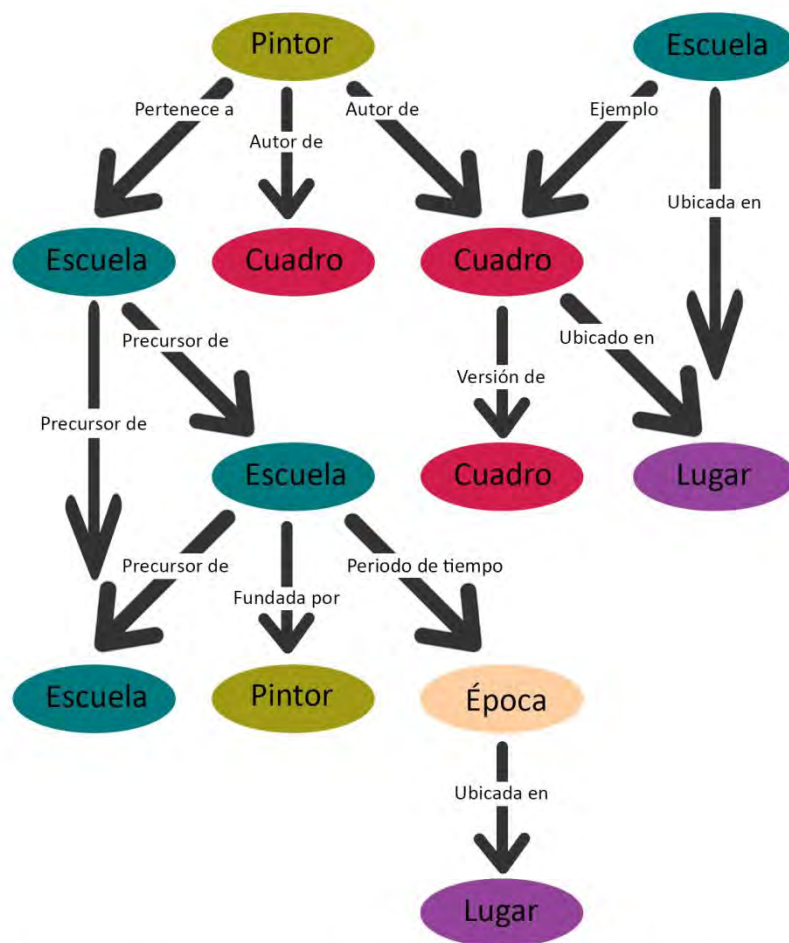
esto? Para lograr que las máquinas logren hacer inferencias (o comparaciones) entre la información y puedan realizar búsquedas más precisas”.

En una web semántica llena de conceptos podríamos buscar, por ejemplo, “cuáles equipos de beisbol jugaron el último domingo del mes pasado a una temperatura ambiente de 22° centígrados y se generaría una página donde podríamos saber, tal vez, que tanto el equipo de los Pericos de Puebla en México como los Cardenales de San Luis en Estados Unidos jugaron su partido del último domingo de mes a una temperatura ambiente de 22° grados.” Esta misma búsqueda con las herramientas de la web 2.0 que usamos actualmente enviaría una lista gigantesca de posibles respuestas en las que manualmente se tendría que buscar la información para luego compararla y después de un rato de observación y algo de lógica se obtendría la respuesta.

Buscar información en la web de esa manera no se trata de una inteligencia artificial ni de ningún tipo de magia, se trata de una serie de estándares que se tienen que seguir al momento de crear un sitio web o que se tienen que agregar a los sitios ya existentes para que puedan pertenecer a una infraestructura global en la que es posible compartir y reutilizar datos y documentos entre diferentes aplicaciones y usuarios, para obtener esa adecuada definición de los datos las herramientas que componen la web semántica se valen de 3 mecanismos ya estandarizados: Los lenguajes de programación RDF, SPARQL y OWL.

Sin embargo no hay que preocuparse demasiado, dado que esta lectura no está enfocada a crear expertos en lenguajes de programación ni pretende ser un manual al respecto; no se ahondará mucho en los tres lenguajes; sin embargo, es importante mencionarlos y sobrevolar rápidamente entre ellos para comprender su función dentro de la web para hacer más amplio el conocimiento sobre cómo funcionan los mecanismos con los que se construye la web semántica, la web 3.0 y los mecanismos de la web que están ayudando al desarrollo de un internet más potente, estructurado y útil para los usuarios.

Primero que nada hay que dejar en claro lo que es un lenguaje de programación, en general se trata de un método delimitado, conveniente y sencillo para describir estructuras de información y las secuencias de acciones que son necesarias para que una computadora ejecute una tarea en concreto. Los lenguajes de programación utilizan el alfabeto tal y como nosotros hacemos para escribir y comunicarnos, solo que en este caso sirven para comunicarnos con las computadoras o lograr que las computadoras puedan interactuar bajo reglas e instrucciones limitadas que proporciona el lenguaje de programación que se esté utilizando.²³



En la web semántica o web de los conceptos la información está estructurada de tal manera que los mecanismos de la web pueden analizar de qué trata la información y de qué manera se relaciona con otra.

²³ Aguilar, Luis Joyanes. *Metodología de la programación*. Mc-Graw Hill, México, 1991

Al igual que los lenguajes humanos como el español, el inglés o el ruso los lenguajes de programación poseen una estructura propia que permite combinar palabras y símbolos de forma que se puedan expresar frases aceptables y con significado dentro de su estructura, sin embargo los lenguajes de programación no son tan amplios y flexibles como los lenguajes naturales humanos donde podemos construir frases que aunque no sigan al pie de la letra las leyes de la gramática pueden ser comprendidas por otras personas, en el caso de los lenguajes de programación la gramática y la sintaxis tienen que ser respetados al pie de la letra o de lo contrario las instrucciones brindadas a la computadora carecerán de significado o producirán resultados indeseables.

Podemos resumir que un lenguaje de programación es un conjunto de reglas, símbolos y palabras especiales que permiten construir un programa de computadora para que ésta cumpla funciones específicas. Existen muchos lenguajes de programación, cada uno con sus características y capacidades muy diferenciadas para labores específicas, además cada lenguaje de programación consta de sus propias instrucciones y enunciados que al combinarse permiten construir programas de computadora o aplicaciones.²⁴



La web está formada por muchos lenguajes de programación, todos ellos distintos y para diferentes propósitos, esto es una gran ventaja y a la vez un gran problema

²⁴ Terréense W. Pratt, Marvin V, Zelkowitz. *Lenguajes de programación: diseño e implementación*. Prentice-Hall Hispanoamericana, S.A. Tercera edición, México 1998.

Primera herramienta de la web semántica: RDF.

El primero de los lenguajes que utiliza la web semántica para lograr su funcionamiento es el Marco de Descripción de Recursos (del inglés Resource Description Framework) mejor conocido simplemente como RDF, la función de este lenguaje es ser un formato de intercambio de archivos que permite definir las propiedades de la información a partir de metadatos, los metadatos los podemos definir simplemente como pequeños trozos de información que describen las propiedades de otra información, para lograrlo RDF utiliza un sistema de tres elementos o metadatos que describen a una información en “sujeto - propiedad - objeto”. Con estas tres cualidades se crea un modelo de datos para los recursos y las relaciones que se puedan establecer entre ellos; de esta manera, RDF se encarga de proporcionar información descriptiva y simple sobre los recursos que se encuentran en la web de una manera estandarizada y bajo un orden que pueda ser analizado y consultado por cualquier administrador o mecanismo de la web que siga las reglas del protocolo RDF.



```
<monumento a la revolución><hecho de><piedra>  
<monumento a la revolución><hecho de><acero>  
<monumento a la revolución><usado para><mausoleo>  
<monumento a la revolución><tiene><explanada>  
<monumento a la revolución><es un><museo>  
<monumento a la revolución><pertenece a><gobierno del DF>  
<monumento a la revolución><tiene><lámparas de LED>  
<monumento a la revolución><tiene><mástiles>  
<monumento a la revolución><tiene><fuentes>  
<monumento a la revolución><tiene><material historiográfico>  
<monumento a la revolución><es un><espacio público>  
<monumento a la revolución><es un><espacio turístico>  
...
```

Para describir al Monumento a la Revolución un RDF tendría que expresarse mediante propiedades de “sujeto – propiedad – objeto”

Las tripletas de metadatos son sentencias o descripciones que ayudan a definir un concepto, a partir de estos grupos de datos el RDF tendría entonces la capacidad de almacenar las propiedades de un concepto, que en nuestro ejemplo serían las

propiedades que delimitan lo que es el Monumento de la Revolución, sus partes y las características de sus partes para con el objeto que se describe.

Esta organización por tripletas comienza a hacerse más compleja conforme más partes tiene el concepto que se desea describir, a la vez que es normal que las partes de un objeto tengan que describirse a sí mismas como una subcategoría del objeto descrito, tal es el caso del museo que se encuentra en el Monumento a la Revolución que dentro de un RDF tendría que describirse bajo las mismas reglas “sujeto – propiedad – objeto” llegando a un punto tal que el Monumento a la Revolución se convierta en un objeto para describir una de las partes de sí mismo por lo que la descripción de tripletas terminaría algo equivalente a:



```
<monumento a la revolución><es un><museo>  
<museo><trata de><historia>  
<museo><trata de><mexico>  
<museo><trata de><guerras de intervención>  
<museo><está echo de><cristal>  
<museo><está echo de><pedra>  
<museo><está echo de><acero>  
<museo><pertenece a><Gobierno del Distrito Federal>  
<museo><se encuentra en><monumento a la revolución>  
<museo><se encuentra en><Delegación Cuauhtemoc>  
<museo><trata de><arte>  
<museo><tiene><cafetería>  
...
```

Cada objeto contenido en un RDF también despliega su propio paquete de tripletas de “Sujeto - Propiedad - Objeto” con esto se pueden hacer aún más especificaciones sobre lo que contiene y trata la información

Definir un concepto con base en tripletas puede resultar un proceso largo y con miles de tripletas como resultado; este proceso, aunque a simple vista puede parecer algo lento y engorroso es muy eficiente al momento de querer delimitar un concepto para una máquina; además, la intención de este sistema es que en la web existan descripciones ya definidas sobre una gran cantidad de objetos, de esta manera al momento de que a la web se le solicite información sobre, por ejemplo, museos, la web podrá consultar el RDF

asignado al t3pico y as3 definir qu3 es lo que compone un museo y la clase de informaci3n que pertenece a ese concepto.

En resumen podemos decir que gracias a RDF se puede describir un concepto con base en un identificador de tripletas y ese identificador tiene reglas estandarizadas para poder definir las cosas que quiere expresar, de esta manera ese concepto puede ser 3til para describir un objeto, como por ejemplo un museo, pero al mismo tiempo otros objetos de la misma categor3a con diferentes propiedades particulares, as3 el identificador <museo> puede ser 3til para iniciar la descripci3n sobre el Museo de las Intervenciones en M3xico y al mismo tiempo puede describir al Museo Smithsoniano en Estados Unidos bas3ndose en el concepto b3sico de lo que es un museo para luego establecer las particularidades de cada uno.

La idea al final de RDF es que cada objeto, persona o concepto tenga un identificador en la web que lo defina claramente y de manera estandarizada, de tal manera se pretende poder extraer informaci3n de cualquier elemento que exista en la web haciendo comparaciones con grandes cantidades de descripciones de conceptos para de all3 poder extraer informaci3n.²⁵

En la pr3ctica RFD ha tenido limitantes que le permiten ser eficaz solo con un n3mero limitado de tripletas descriptivas para trabajar por lo que bases de datos que contienen m3s de 50 millones de tripletas comienzan a ser conflictivas y arrojar resultados imprecisos, al implementar RDF en otro sistema de bases de datos dedicados 3nicamente a este formato de descripciones como lo es el sistema Big Data se han alcanzado resultados exitosos con hasta 320 millones de tripletas de forma efectiva.

²⁵ John Hebel, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez. *Semantic Web Programming*. Wiley Publishing. Primera edici3n, 2009

Por otro lado el sistema llamado “open graph” ha logrado procesar bases de datos con 50 billones de tripletas descriptivas utilizando el procesamiento de súper computadoras dedicadas a ese proceso.²⁶

A pesar de que 50 billones de descripciones son un gran número éste se vuelve relativo cuando se piensa en que la tarea de RDF es describir todos los objetos, sujetos y conceptos existentes en toda la web, por lo que aunque la web semántica y el RDF están en marcha en entornos cerrados para labores académicas, comerciales, gubernamentales y científicas sigue estando fuera de un entorno abierto debido a que se requeriría una gigantesca capacidad de procesamiento computacional para poder llevar a cabo esos procesos de manera eficiente con bases de datos bastante mayores a 50 billones de tripletas.

Segunda herramienta de la web semántica: SPARQL.

La segunda herramienta para lograr aplicaciones semánticas es SPARQL que es un acrónimo de inglés “Protocol And RDF Query Language”. Este lenguaje sirve para navegar en la información que se encuentra en formato RDF usando técnicas y algoritmos para consultar la información con el fin de buscar los datos que el usuario requiera.

Como ya hemos comentado RDF es el formato en el cual se almacenan las descripciones que señalan las propiedades de un concepto, para consultarlo se puede hacer uso de distintas herramientas; sin embargo, SPARQL llega como una recomendación de la W3C (El Consorcio WWW, en inglés: World Wide Web Consortium, es un consorcio internacional que genera recomendaciones y estándares que aseguran el crecimiento de la World Wide Web a largo plazo y que fue fundado por Tim Berners-Lee) para estandarizar la manera en que se hacen consultas a RDF y así poder tener reglas más

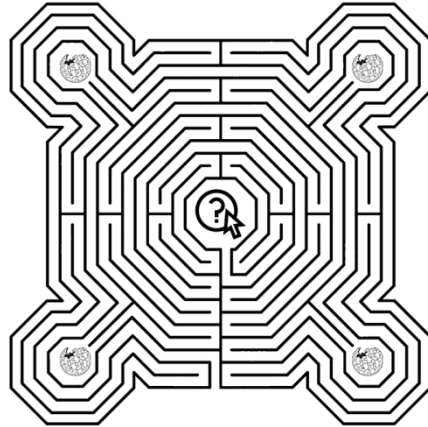
²⁶ Conferencia: Web semántica, de la teoría a la práctica, martes 29 julio 2014, 19:00 – 20:00. Ponente: Javier Solís González. Gerente de desarrollo de nuevos productos y servicios en el Fondo de Información y Documentación para la Industria (INFOTEC) de CONACYT, Guadalajara, Jalisco, México.

uniformes tanto en la forma en que se plantea la información en RDF como en la manera que debe ser analizada mediante SPARQL, de esta manera se pretende tener informes semánticos estandarizados y útiles en cualquier entorno.

Una forma de ejemplificar el funcionamiento de SPARQL es a partir del proyecto DBpedia.org, esta iniciativa construida a partir de un modelo Crowd-Sourced, es decir a partir de la colaboración abierta de la comunidad, está dedicada a extraer información estructurada de muchas entradas distintas de la Wikipedia para arrojar nueva información de consulta hacia la web.

Con la DBpedia es posible utilizar cadenas de consultas sofisticadas desde la wikipedia y crear paquetes de datos con información de un tipo o tema en particular para extraer información específica a partir de un gran volumen de información.

El ejemplo del capítulo 1 donde nos preguntábamos sobre cuáles jugadores de futbol usaban la camiseta número 11, habían hecho anotación en algún estadio con capacidad de más de 40 mil asistentes y han nacido en un país de más de 10 millones de habitantes se resolvió utilizando la DBpedia, de esta manera se pudo cruzar información sobre los jugadores profesionales de futbol, sus números de camiseta y sus nacionalidades a la vez que se comparaba la información de los estadios con capacidad mayor a 40 mil asientos y los goles anotados en ellos. Toda esa información estaba únicamente en la wikipedia de forma aislada, pero utilizando la DBpedia y su buscador basado en SPARQL se pudo extraer esa información y transformarla en respuesta que se solicitó.



Miles de entradas de Wikipedia tienen información aislada que al juntarla y compararla con otras entradas puede arrojar respuestas a preguntas complejas. DBpedia hace posibles realizar búsquedas con Wikipedia que no son posibles con un buscador web común.

Las herramientas de DBpedia son públicas y están abiertas a toda persona que desee realizar consultas así como para quienes quieran participar en el desarrollo de las herramientas para mejorar la DBpedia así como para quienes quieran apoyan en la extracción de información de la Wikipedia para generar nueva información que ayude a hacer más sofisticada tanto la DBpedia como la propia wikipedia así como a la creación de nuevos mecanismos para la navegación y la conectividad entre información; la intención de este esfuerzo es hacer más fácil la obtención de datos útiles para asistir a diferentes fines prácticos así como académicos entre los que está la inteligencia artificial, las tecnologías web, las ingenierías de software, el comercio digital o la administración pública por solo mencionar algunos de los usos posibles para la información extraída con DBpedia. Un dato importante de SPARQL es que se trata de un lenguaje específicamente diseñado para interactuar con bases de datos en RDF por lo que su implementación es muy efectiva para su uso entre computadoras, paradójicamente este lenguaje especializado en ser amistoso con la comunicación entre máquinas no está diseñado para ser interpretado con facilidad por los seres humanos por lo que es necesaria un interfaz que logre mediar entre la información solicitada de RDF mediante SPARQL y una forma de expresar esa información para ser consultada por el usuario de una forma que tenga

sentido para ser interpretada por un humano. Para resolver este problema se utilizan buscadores que hacen de intermediarios como el Virtuoso SPARQL Query Editor.²⁷

No se necesita ningún conocimiento de RDF o de SPARQL para realizar búsquedas en la DBpedia, aunque sí es requerido conocer las reglas para generar búsquedas semánticas a partir de los lineamientos de los buscadores.²⁸ Al visitar la wikipedia es muy probable poder encontrar en la parte superior derecha del artículo un recuadro con información específica sobre el tema, por ejemplo si se entra en el artículo dedicado a México²⁹ en el recuadro se encontrarán datos muy concretos sobre el país como su himno nacional, escudo, bandera, latitud, longitud, altura, número de habitantes, moneda, densidad de población, entre otros datos que definen lo que es México en cuanto a geopolítica se refiere, si se consulta otro país, por ejemplo Japón³⁰, encontraremos el mismo recuadro igualmente con datos específicos del país, aunque claro, puede que la información que se proporciona en el artículo de México difiera con el de Japón debido a algunas diferencias propias de cada país y a que la información que se vierte en la wikipedia es proporcionada por los propios usuarios y no se sigue un lineamiento obligatorio para los datos que se deben proporcionar.

El recuadro de información de la Wikipedia del que hablamos es importante pues justo de él se parte para extraer información con el uso de la DBpedia, SPARQL y las

²⁷ “Virtuoso SPARQL Query Editor”. Consultado el 12 de agosto de 2015. <http://dbpedia.org/sparql>

²⁸ “Google Code Archive - Long-term storage for Google Code Project Hosting.” Consultado el 12 de agosto de 2015. <https://code.google.com/archive/p/tdwg-rdf/wikis/Beginners6SPARQL.wiki>

²⁹ “Mexico”. *Wikipedia, the Free Encyclopedia*, el 13 de agosto de 2015. <https://en.wikipedia.org/w/index.php?title=Mexico&oldid=735470769>

³⁰ “Japan”. *Wikipedia, the Free Encyclopedia*, el 13 de agosto de 2015. <https://en.wikipedia.org/w/index.php?title=Japan&oldid=735771550>

United Mexican States Estados Unidos Mexicanos	
	
Flag	Coat of arms
Anthem: <i>Himno Nacional Mexicano</i> (English: "Mexican National Anthem") 	
	
Capital and largest city	Mexico City <small>19°26'N 99°08'W</small>
Official languages	None at federal level
Recognised regional languages	Spanish · 68 native language groups, ^[1]
National language	Spanish ^[2]
Demonym	Mexican
Government	Federal presidential constitutional republic ^[2]
- President	Enrique Peña Nieto (PRI) 
- Secretary of the Interior	Miguel Ángel Osorio Chong (PRI) 
- President of the Senate	Miguel Barbosa Huerta (PRD) 
- President of the Chamber of Deputies	Silvano Aureoles Conejo (PRD) 
- Supreme Court President	Juan Silva Meza
Legislature	Congress
- Upper house	Senate
- Lower house	Chamber of Deputies
Area	
- Total	1,972,550 km ² (14th) 761,606 sq mi
- Water (%)	2.5
Population	
- 2013 estimate	118,395,054 ^[4] (11th)
- Density	57/km ² (142nd) 142/sq mi
Currency	Peso (MXN)
Time zone	See Time in Mexico (UTC-8 to -6) - Summer (DST) varies (UTC-7 to -5)
Drives on the	right
Calling code	+52
ISO 3166 code	MX
Internet TLD	.mx
<small>¹ Article 4.° of the General Law of Linguistic Rights of the Indigenous Peoples.^[3]</small> <small>² Spanish is the <i>de facto</i> official language of the Mexican federal government.</small>	

Japan 日本国 Nippon-koku Nihon-koku	
	
Flag	Imperial Seal
Anthem: "Kimigayo" 君が代 	
	
Capital	Tokyo <small>35°41'N 139°40'E</small>
Official languages	None ^[1]
Recognised regional languages	Ainu itak Ryukyuan languages Eastern Japanese Western Japanese several other Japanese dialects
National language	Japanese
Ethnic groups (2011 ^[2])	98.5% Japanese 0.5% Korean 0.4% Chinese 0.6% other
Demonym	Japanese
Government	Unitary parliamentary Constitutional monarchy
- Emperor	Akihito
- Prime Minister	Shinzō Abe
- Deputy Prime Minister	Tarō Asō
Legislature	National Diet
- Upper house	House of Councillors
- Lower house	House of Representatives
Area	
- Total	377,944 km ² ^[4] (62nd) 145,925 sq mi
- Water (%)	0.8
Population	
- 2014 estimate	126,434,964 ^[5] (10th)
- 2010 census	128,056,026 ^[6]
- Density	337.1/km ² (36th) 873.1/sq mi
Currency	Yen (¥) / En 円 (JPY)
Time zone	JST (UTC+9) - Summer (DST) not observed (UTC+9)
Date format	yyyy-mm-dd yyyy年m月d日 Era yy年m月d日 (AD-1988)
Drives on the	left
Calling code	+81
ISO 3166 code	JP
Internet TLD	.jp

Los datos de estos recuadros son similares entre sí y se usan para compartir información entre ellos y entre todos los de la misma categoría, no solo eso, tienen también su equivalencia en diferentes idiomas para poder organizarlos y consultarlos dentro de una sola base de información de un mismo tópico.

tripleas de las que hablamos en el apartado de RDF y claro, no solo de este recuadro se extraen datos de paíes sino de casi cualquier t3pico que pueda ser relacionado con otros de su misma categoría, por ejemplo los libros, películas, video juegos, personalidades, empresas, organizaciones, t3picos especializados en temas científcos, etc3tera.

Para 2014 la versi3n inglesa de DBpedia tiene un paquete de datos que describe m3s de 4 millones de “conceptos” y m3s de 580 millones de “hechos o datos” sobre esos conceptos. Adem3s se han creado las versiones localizadas a 125 idiomas que todas juntas describen m3s de 38 millones de “conceptos” de los cuales m3s de 23 millones ya est3n relacionadas con su contraparte en ingl3s.

Así la DBpedia completa contiene m3s de 38 millones de paquetes de datos en 125 idiomas, m3s de 25 millones de links a im3genes y casi 30 millones de links a sitios ajenos a la wikipedia; hay tambi3n 81 millones de links a categorías de wikipedia. Todo esto junto hace que la DBpedia conste de 3 mil millones de piezas de informaci3n en tripleas de RDF de las cuales 583 millones han sido extraídas de la edici3n inglesa de wikipedia, 2 mil millones y medio han sido extraídas de las ediciones en otros idiomas y cerca de 50 millones son links a paquetes de datos externos.³¹

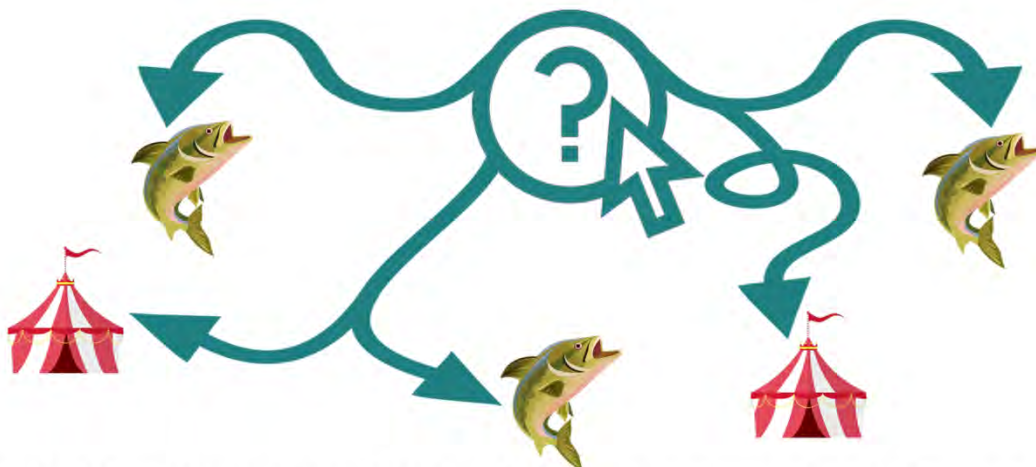
Con todo esto la DBpedia se conforma como uno de los m3s grandes esfuerzos por utilizar herramientas sem3nticas de formas estandarizada bas3ndose en RDF y SPARQL al extraer varios tipos de informaci3n estructurada (o semi-estructurada) de 125 ediciones de Wikipedia en sus diferentes idiomas y combina esta informaci3n en una enorme base de conocimiento.

³¹ “DBpedia”. Consultado el 13 de agosto de 2015. <http://wiki.dbpedia.org/Datasets>

Tercera herramienta de la web semántica: OWL.

La última de las herramientas para la creación del proyecto de web semántica es la “Web ontology lenguaje” en español llamada “Lenguaje de Ontologías Web” y conocida como OWL; este lenguaje de programación está diseñado para su uso en aplicaciones que requieren procesar la información que se contiene dentro de los documentos. Esta habilidad de OWL es peculiar pues se dedica a analizar la información contenida en los documentos para darle un sentido y así hacerla legible para otras computadoras, esta habilidad peculiar de OWL destaca pues lo normal es que los lenguajes traten de exponer la información de forma que los humanos podamos darle lectura, mientras que en este caso se trata de un lenguaje que analiza la información para hacerla legible y con sentido para las otras máquinas.

El lenguaje OWL puede ser usado para representar de manera explícita los significados de términos que hay en la información dentro de los documentos y organizarla así mediante vocabularios que la definan y delimiten para a su vez crear relaciones entre esos términos.



La confusión es algo común para los buscadores web, distinguir entre palabras iguales que significan diferentes cosas es un conflicto común, en este caso si se busca por “carpa” podemos encontrar al pez carpa y a la vez la carpa de un circo sin que la web distinga la diferencia entre una y otra.

La representación de términos y sus interrelaciones se les denomina ontología.³² Una ontología se puede entender como un conjunto de especificaciones y delimitaciones de conceptos que pueden ser utilizados por las máquinas, gracias a esto la web semántica se puede valer de las ontologías generadas con OWL para realizar inferencias lógicas y generar mapas de relaciones sobre jerarquías conceptuales distintas para poder utilizar descripciones tanto para el ámbito para el que fueron creadas como para su reutilización fuera del tópico para que fueron previstos.³³

Utilizar las ontologías para crear mapas del significado de diferentes términos para las máquinas nos ayudará en un futuro a solucionar el problema de la ambigüedad semántica existente en los documentos de la web.

La habilidad de OWL para expresar ontologías interpretables para otras computadoras lo hace una herramienta poderosa que al combinarse con RDF y SPARQL capacita extraer información muy precisa a partir de consultar cantidades enormes de documentos para crear nuevas ontologías que den significado a la información analizada en muy poco tiempo.

En otras palabras podemos decir que las partes que componen a las herramientas de la web semántica funcionan de la siguiente manera:

En primer lugar aparece el RDF como el lenguaje que almacena las propiedades de la información que describen lo que se contiene en los documentos que se van a analizar en una consulta semántica; a continuación el SPARQL es el mecanismo que señala las reglas con las que se va a analizar la información planteada en RDF para señalar cuáles son elementos que está solicitando la búsqueda para su análisis; finalmente OWL se encarga

³² “OWL Web Ontology Language Overview”. Consultado el 15 de agosto de 2015.
<https://www.w3.org/TR/owl-features/>

³³ Revista Digital Universitaria, 10 de julio de 2005, Vol.6, No.7 ISSN: 1607 – 6079, Publicación mensual.
<http://www.revista.unam.mx/vol.6/num7/art72/art72-3.htm#a> visto 1 de octubre de 2014

de crear las interrelaciones que existen entre la información que está arrojando RDF y SPARQL, de esta manera el lenguaje OWL funciona como un esquema que interrelaciona la información para a partir de ello crear inferencias, comparaciones, distinciones, restricciones, relaciones y otras operaciones más para arrojar resultados a las búsquedas que solicitan información específica de información que se encuentra dispersa en diferentes documentos o lugares de la web.

Debido a que OWL es el lenguaje encargado de crear las relaciones entre los conceptos es necesario dividirlo en 3 distintas versiones o especies de OWL, cada uno diferenciado por su capacidad expresiva.

En primer lugar se encuentra OWL Lite, ésta es la versión más básica de OWL y su función es definir los elementos básicos para crear de manera simple ontologías fáciles de procesar; OWL lite está diseñado para dar el mejor rendimiento al realizar clasificaciones jerárquicas y restricciones simples al grado de, por ejemplo, hacer restricciones de cardinalidad con valores únicamente de 0 o 1, lo que quiere decir que al realizar una consulta a RDF tan solo podrá asignar diferencias entre la información de 0, es decir “no” o 1, lo que quiere decir “sí”; ese sub-lenguaje trata de procesar la información de formas tan simples como jerarquías que determinen el mayor o el menor valor sobre si algo corresponde o no corresponde a una variable o si una información pertenece a un grupo de conceptos o no, como ya se ha dicho OWL lite trata de crear ontologías de la manera más simple posible.

En segundo lugar está OWL DL que incluye todas las funcionalidades de los lenguajes OWL, pero imponiendo restricciones basándose en las propiedades que RDF puede usar; este sub-lenguaje está diseñado para tener una máxima expresividad a la vez que conserva su compatibilidad computacional para seguir siendo información procesable por herramientas semánticas. Aunque OWL DL incluye todas las herramientas de OWL

éste solo puede usarlas bajo algunas restricciones determinadas por una lógica de descripción que no rompa la estructura de la información que se está solicitando.

Por último está la versión más completa llamada OWL Full, como su nombre lo indica esta versión de RDF incluye todas las funcionalidades del lenguaje y admite propiedades típicas de RDF como la reificación, es decir que permite sentencias lógicas que hacen referencia a otras sentencias lógicas.

OWL Full está creado para quienes desean una máxima expresividad y libertad en el uso de RDF aunque esto tenga que significar el sacrificio de garantías computacionales, por lo que es probable que con OWL Full se logren búsquedas muy potentes en ambientes de información cerrados y estructurados mientras que se pierde compatibilidad con sistemas ajenos o no controlados por el mismo usuario.

Cada uno de estos sub-lenguajes es una extensión de su predecesor más simple y por lo tanto también es una extensión de lo que se puede expresar y validar correctamente. Este proceso solo es válido de forma ascendente y no a la inversa, por lo que el uso de OWL lite puede generar ontologías validas para OWL DL y OWL Full, pero no a la inversa, de modo que una ontología generada en OWL Full no es válida en OWL DL o en OWL Lite, dicho de otra manera, los grupos de relaciones ontológicas se mantienen, pero no así las relaciones inversas:

- Cada ontología legal de OWL Lite es una ontología legal de OWL DL
- Cada ontología legal de OWL DL es una ontología legal de OWL Full
- Cada conclusión válida de OWL Lite es una conclusión válida de OWL DL
- Cada conclusión válida de OWL DL es una conclusión válida de OWL Full

Dependiendo de las necesidades del desarrollador de herramientas semánticas éste deberá utilizar la versión de OWL que más convenga, ya que entre mayor sea la capacidad expresiva de una versión, mayores serán también las dificultades que se tendrán para operar de una forma eficiente.³⁴

RDF + SPARQL + OWL = Web semántica.

Con el uso de los lenguajes de programación de la web combinados de manera que se complementan se han creado los estándares para la creación de la web de los conceptos, esta web semántica poco a poco ha delimitado sus reglas para poder administrar cantidades inimaginables de información así como bases de datos de las que se puede extraer información en sí misma.

En suma, una vez que se ha estructurado la información de manera que pueda ser procesada por las máquinas es necesario contextualizarla dentro de un dominio concreto, es decir, toda la información debe ser colocada dentro un depósito en internet, un servidor o un repositorio específico en una computadora para así procesarla e inferir conocimiento a partir de ella. De esta manera a partir de la suma de los conceptos que hemos depositado de manera estructurada se pueden crear las relaciones que establecen entre sí nuevos conceptos y axiomas generados sobre los primeros conceptos que depositamos; así con el uso de RDF, SPARQL y OWL se pueden generar ontologías que sirven para que las máquinas obtengan un esbozo más claro de la realidad.

³⁴ Edrerra-Viedma, José M. Morales-del-Castillo. *Referencias a la web semántica desde la perspectiva de la documentación*. Revista Investigación Bibliotecología, vol 21, num 43, Julio/diciembre, 2007, México. Visto el 10 de octubre de 2014. <http://www.revistas.unam.mx/index.php/ibi/article/view/4138/3672>



Con esto las computadoras logran hacer consultas a software inteligente que trabaja con conceptos dentro de un contexto, este avance supone que en vez de utilizar los buscadores como meros recuperadores de información se conviertan en herramientas capaces de obtener respuestas precisas a preguntas concretas. Es como si en vez de solo poder extraer información ahora se pudiera extraer conocimiento de las computadoras.

Capítulo 3. Una red llena de datos... y de personas.

La web semántica está creando nuevas maneras de organizar el mundo real.

- *Computadora, cual es la cantidad de enfermedades a las que puedo estar expuesto en la ciudad de México.*
- *Hay registro de más de 24,000 enfermedades posibles relacionadas con ese territorio.*
- *Computadora, delimita a cuántas de ellas estaría expuesto dentro de la Facultad de Química de la UNAM.*
- *Hay más de 1,500 enfermedades posibles bajo condiciones regulares.*
- *Qué productos químicos que hay en la Facultad podrían causarme alguna de esas enfermedades.*
- *Las sustancias son: Amianto, Cloruro de vinilo, Monóxido de carbono, Plomo, Estireno, Tolueno, Tricloroetileno...*

Esta breve conversación bien podría ser parte de un show de ciencia ficción o de alguna novela donde hubiera computadoras con una inteligencia artificial destacable; sin embargo, a pesar de que en este momento no podemos realizar preguntas así de precisas a nuestras computadoras de uso personal, eso no quiere decir que no sea posible dotar a un sistema de cómputo con la información necesaria para que pueda responder a las preguntas de nuestra conversación imaginaria. Un ejemplo de ello es Watson, un sistema de cómputo creado por la empresa IBM³⁵ que puede entender el lenguaje humano y que es capaz de responder a preguntas de manera precisa, incluso con la suficiente precisión y rapidez para competir contra los mejores jugadores de *Jeopardy!*³⁶, Watson es una propuesta para obtener un motor de búsqueda de información basado en un sistema de

³⁵ International Business Machines (IBM), es una reconocida empresa multinacional estadounidense de tecnología y consultoría con sede en Armonk, Nueva York. IBM fabrica y comercializa hardware y software para computadoras, y ofrece servicios de infraestructura, alojamiento de Internet, y consultoría en una amplia gama de áreas relacionadas con la informática, desde computadoras centrales hasta nanotecnología.

³⁶ IBM. *IBM and the Jeopardy Challenge*, 2010. Consultado el 8 de diciembre 2015.
<https://www.youtube.com/watch?v=FC3IryWr4c8>

búsqueda de respuestas e interrelaciones en vez de los sistemas de búsquedas de información basados en coincidencia de palabras y metadatos.

Cuando hablamos de Watson hablamos de un sistema informático en un entorno cerrado y controlado, lo que quiere decir que la información, los mecanismos y los protocolos que sigue Watson son controlados por el equipo de desarrollo de IBM, de esta manera el sistema adquiere una gran estabilidad y es relativamente sencillo encontrar los errores que se puedan generar al momento de procesar la información.

Para hablar más sobre los sistemas de búsqueda y de cómo las computadoras pueden programarse para analizar el lenguaje humano se entrevisté al Doctor David Gondek, él ha estado desde el principio del proyecto Watson y es líder del equipo de estrategias de respuesta del sistema, su labor es desarrollar el sistema de aprendizaje de algoritmos para catalogar y estimar la mayor confiabilidad de las respuestas que IBM Watson puede ofrecer.

Al hablar sobre los mecanismos que utiliza un sistema como Watson y los sistemas con que se hacen búsquedas en la web hay que entender algunas cosas sobre el problema que representa el sólo hecho de crear un buscador de información que pretenda resolver una búsqueda de cualquier tópico posible; uno de esos puntos es que los motores de búsqueda, sean cuales sean, siempre darán una respuesta sin importar la pregunta que se le esté realizando, “para una computadora no existe una respuesta como “no sé” o “no sé qué es lo que estás preguntando”, “no entiendo a qué te refieres”. Los buscadores en realidad parafrasean, buscan palabras y términos que les sean coincidentes o similares, buscan sinónimos o metadatos, pero en realidad no entienden qué es lo que estamos preguntando, las computadoras no razonan, son simplemente un mecanismo”, comenta el Doctor David Gondek.

Y es que una paradoja fundamental de los buscadores es que son muy diferentes a las personas, “a una persona puedes darle mucha información y lo más probable es que con el tiempo entre más información acumule más probabilidad habrá de que pueda responder de forma acertada a una pregunta, con las computadoras no es igual, con un motor de búsqueda ocurre que entre más información agreguemos a su base datos más tonta se volverá en cuanto a las respuestas que puede ofrecer”, y este problema especialmente se acentúa cuando el usuario no se es muy preciso con las preguntas que plantea a una máquina.



Una persona puede aprender y relacionar información entre más fuentes tenga a su alcance, con los sistemas de cómputo actuales pasa lo contrario, entre mayor sea el volumen de información mayores dificultades tendrá para extraer información correcta.

Para sobreponerse a ese problema IBM ha generado un sistema llamado *Automatic Open-Domain Question Answering*, al cual se le nutre con grandes volúmenes de preguntas y respuestas predefinidas, al mismo tiempo este sistema se complementa con otra herramienta de IBM llamada *Decision Maker*, la cual trata de analizar la forma en que están formuladas un gran volumen de preguntas para así producir respuestas congruentes al tipo de cuestionamiento que se le realiza.

“Partimos de la base de que las computadoras hacen comparativos y que no tienen idea de lo que están buscando, es por eso que tampoco tienen una idea aproximada de lo que necesitan encontrar, al realizar su tarea son totalmente carentes del significado de lo

que están buscando y de lo que encuentran cuando aciertan en su labor, es por eso que es difícil encontrar información exacta cuando tenemos un gran volumen de información.

Al momento de que los usuarios usamos cualquiera de los motores de búsqueda tenemos que ser muy precisos con la manera en que se plantea la búsqueda, de otra manera esta herramienta no será realmente útil para encontrar la información que buscamos a pesar de que la respuesta que necesitamos sí se encuentre dentro del alcance del motor de búsqueda”, continúa el Doctor David Gondek

Con Watson, la empresa IBM no pretende hacer bases de datos de cada posible pregunta, sino que busca asignar un valor a las posibles respuestas para así poder calcular cual es la opción más fiable para contestar la pregunta que se le planteó, se busca basar las respuestas en fuentes especializadas y fiables de manera que se puedan analizar resultados que previamente han generado éxito en búsquedas y preguntas anteriores.

Tomando todo lo anterior en mente podemos entender cómo puede crearse un sistema que pueda analizar el lenguaje humano para poder responder correctamente a preguntas de manera abierta e incluso de forma ambigua, pero si todo este trabajo lo puede realizar un sistema controlado y cerrado como Watson ¿es posible que los buscadores web que usamos los usuarios todos días puedan tener la misma capacidad?

La respuesta es sí... pero no de manera inmediata.

“La web es un ente vivo, siempre se está agregando contenido nuevo, no todo el contenido de la web está creado bajo los mismos parámetros e incluso fácilmente se puede subir a la web información falsa o difamatoria, no hay manera de crear un sistema de fiabilidad para todo lo que se sube a la web e incluso si lo hubiera la información que colocan los usuarios siempre estaría regida por un criterio totalmente subjetivo tanto de quien sube la información como del momento en que se sube y de quienes lo consumen y

comparten; en la red se crean modas, se deforma y transforma el uso de palabras y expresiones, lo que en el mundo fuera de la web tiene una connotación en la web puede tener otra o muchas más connotaciones; la web es tan caótica que no hay mecanismo lo suficientemente capaz para analizar lo que hay y se genera en su totalidad; pero qué tal si en la web dejamos de tratar de analizar y dar un significado a todo lo que hay allí, en vez de eso sería mejor solo enfocarse a información específica, fiable, verificable y delimitada en universos temáticos de fuentes y proveedores que sigan protocolos y metodologías claras.”

Para eso tarea sí hay las herramientas y los mecanismos para controlar de dónde surge, a dónde va, cómo se consume y qué significa la información, se trata de protocolos que ya están en uso detrás de muchos sitios web, es decir Linked Data, Open Data y Big Data³⁷.

Linked Data.

Uno de los protocolos para manejar grandes volúmenes de información en la web y con el uso de tecnología semántica es *Linked Data*, este protocolo es una propuesta para publicar datos estructurados de tal forma que éstos puedan ser conectados entre sí, de esta manera todos esos datos vinculados pueden ser consultados de forma simultánea con la certeza de que pertenecen a un mismo tópico y de esa manera ser expuestos y consultados en la web de forma más útil para los usuarios.

³⁷ Business, Author: Klint Finley Klint Finley. “Open Source Is Going Even More Open—Because It Has To”. *WIRED*. Consultado el 28 de julio de 2015. <http://www.wired.com/2015/07/open-source-going-even-openbecause/>



Linked Data permite construir la web semántica, con ella internet se convierte en una gran base de datos interconectados, procesados y distribuidos en la web, de esta manera los datos se vinculan y se exploran entre sí de manera que se puede extraer información de muchas fuentes para generar respuestas completas, acertadas y con significado.

La intención de usar *linked data* es crear grandes bases de datos que puedan ser expuestas en la web y que a su vez diferentes bases de datos puedan conectarse para crear una enorme base de conocimiento bien estructurada que nutra a los servicios de la web. Para el creador de la web, Tim Berners Lee, *Linked Data* debería pasar de una web de datos a una web conformada por una gran unión de bases de datos interconectados y distribuidos en toda la web³⁸.

Los datos enlazados de Linked Data como un mecanismo que es parte de la web semántica se basa en la aplicación de los principios que conocimos en el capítulo anterior sobre RDF, SPARQL y OWL; la idea es crear una forma unificada de escribir las bases de datos de la web de manera tal que todas las bases de datos puedan vincularse entre sí de manera congruente, de esta manera se reutilizaría la información de todas esas bases de datos de una forma que pueda dar los resultados esperados e incluso algunos inesperados

³⁸ Berners-Lee, Tim. *Tim Berners-Lee acerca de la futura Web*. Consultado el 29 de enero de 2015. http://www.ted.com/talks/tim_berniers_lee_on_the_next_web?language=es

al cruzar grandes volúmenes de información, esto daría un nuevo valor agregado para los usuarios de la web.

Dicho de otra manera, linked data hace capaz a los mecanismos de búsqueda de información de la web de encontrar información relacionada entre sí aunque ésta se encuentre en sitios de la web distintos entre sí, con esta habilidad los usuarios pueden ser capaces de realizar búsquedas complejas que requieran comparar información de diferentes tipos en una sola respuesta, tal como lo haría la computadora ficticia del principio de este capítulo.

La tarea de la web semántica y del uso de *Linked Data* aún es difícil, para el 2012 un estudio dirigido por Internet Data Corporation (IDC por sus siglas en inglés), y patrocinado por la empresa en tecnología EMC, devela que únicamente el 3% de toda la información de internet estaba etiquetada para su análisis y de esa lista solo el 0.5% podía estar disponible para trabajar con ella en grandes volúmenes³⁹, y aunque esta cifra es muy baja en comparación con el total de la información con posibilidades de ser analizada representa una parte importante si consideramos que ese 3% está conformada por información gubernamental y académica que sigue los protocolos de *Linked Data* y de una filosofía de la web conocida como *Open Data*.

Open Data.

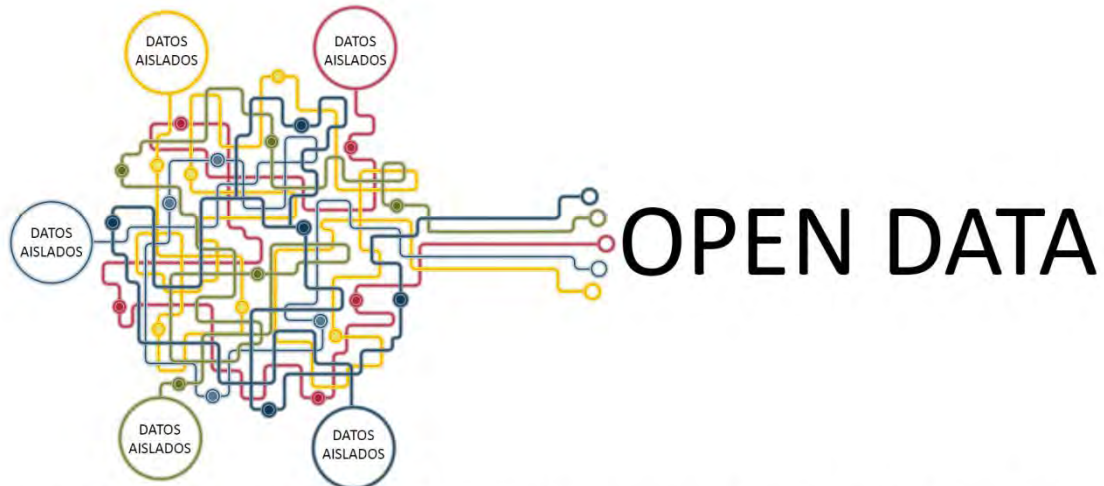
Para tener una gran red de información se requieren muchos datos de muchas fuentes distintas, además es necesario conseguir todas las bases de datos para unirlos en una gran red de redes; esto requiere un gran esfuerzo de todos los que generan y poseen datos útiles para los demás, no solo se trata de los usuarios de la web como individuos sino de un asunto que involucra instituciones, organizaciones e incluso gobiernos enteros.

³⁹ Burn-Murdoch, John. “Study: Less than 1% of the World’s Data Is Analysed, over 80% Is Unprotected”. *The Guardian*, el 19 de diciembre de 2012, sec. News. Consultado el 17 de junio 2014.
<https://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume>

Es claro que la utilidad de las herramientas semánticas y de *Linked Data* abren un enorme potencial para extraer conocimiento útil de grandes volúmenes de información, por ejemplo, qué pasaría si vinculáramos la información generada por todos los hospitales de la ciudad de México con las bases de datos de población de toda la ciudad a la vez que se vinculan las bases de datos de los fabricantes de medicamentos de todo el país. Tal vez lo que podría ocurrir al cruzar toda esa información sería que se encontrara que una de las enfermedades que más afecta a las personas jóvenes de la ciudad de México y que causa un número importante de fallecimientos es tratada con medicamentos que son adquiridos del extranjero y ocurre que un fabricante de medicamentos de México puede manufacturar ese medicamento que es de muy alta demanda solo en la ciudad de México y que de hecho puede proveerlo de manera más económica que lo que cuesta por su vía de importación; tendríamos entonces que los hospitales de la ciudad de México podrían ahorrar algo de dinero en ese medicamento, se podría entonces tratar con mayor efectividad esa enfermedad, disminuirían el número de personas jóvenes enfermas en la ciudad de México y el fabricante de medicinas tendría la oportunidad de hacer un gran negocio.

Si bien el dicho popular dice que la información es poder, la web semántica y *Linked Data* nos dicen que la información que se comparte es aún más poderosa. De eso se trata *Open Data*, una idea que surge de los círculos de creación de software sobre el razonamiento de compartir las instrucciones con que está hecho un programa de computadora, de esa manera cualquier persona es capaz de usar, modificar o ampliar ese programa para después regresarlo a la comunidad como algo revisado y mejorado por más personas que sólo el autor⁴⁰.

⁴⁰ DiBona, Chris, Sam Ockman, Mark Stone. *Open Sources: Voices from the Open Source Revolution*. San Francisco: O'Reilly, 1999.



Open Data trata de crear una filosofía basada en compartir bases de datos de manera abierta, pública y escrita con formatos estandarizados para que cualquier persona pueda hacer uso ellas.

Acerca de Open data, el ingeniero Javier Solís González, Gerente de desarrollo de nuevos productos y servicios en el Fondo de Información y Documentación para la Industria (INFOTEC) de CONACYT, habla sobre esta filosofía como algo que ha permeado en los círculos científicos de todo el mundo, “incluso en México lleva algunos años que el propio Gobierno Federal ha abierto sus bases de datos para compartirlos y enlazarlos con otras dependencias gubernamentales. Incluso aquí, donde se habla de bases de datos que se comparten del gobierno y para el gobierno se encuentran dificultades, especialmente para aquellas bases de datos viejas o que necesitan ser capturadas de vuelta en formatos estandarizados para poder compartirse y vincularse de manera eficaz”.

Para evaluar qué tan útil es la información para su vinculación con *Linked Data* se ha creado un protocolo internacional basado en un sistema de estrellas, donde 1 estrella señala lo que es poco sensible a ser vinculado y 5 estrellas a aquella información que pueda ser utilizada sin problemas en bases de datos semánticas.

★ Los datos con una estrella son aquellos que son públicos y que pueden estar en cualquier formato, no importa si son difíciles de manipular o consultar, son aquellos donde hay algo de información como puede ser una hoja escaneada, un pdf cerrado o la foto de un documento.

★★ Son archivos públicos que contienen información estructurada en archivos que permiten modificación, pero que no necesariamente pueden ser vinculados, por ejemplo un archivo de Excel o de cualquier otro programa que permita la consulta y modificación del contenido.

★★★ Son archivos públicos que contienen información modificable y estructurada y que están en un formato libre de licencias, por ejemplo un archivo CVS en vez de uno de Excel, esto se debe a que para consultar un documento de Excel es necesario contar con un programa que requiere de una licencia, mientras que CVS puede ser leído libremente prácticamente por cualquier equipo sin restricción alguna.

★★★★ Es información pública que está contenida en la web bajo una URI (identificador de recursos uniforme, del inglés “Uniform Resource Identifier”)⁴¹ y que puede señalar qué es lo que contiene y las propiedades de lo que contiene de manera que se pueda vincular con *Linked Data*, no necesariamente se utiliza el formato RDF, aunque es lo recomendable.

★★★★★ Es información pública que está contenida en la web bajo una URI y que señala qué es lo que contiene y las propiedades de lo que contiene, además

⁴¹ No confundir con URL (localizador de recursos uniforme, del inglés Uniform Resource Locator) La diferencia es que una URI señala a un recurso único que no puede ser modificado de modo que siempre haga referencia a la misma información, en contraparte a una URL que a partir de una dirección web señala a un recurso que puede ser modificado con el tiempo, por ejemplo, la URL www.unam.mx señala al sitio web de la Universidad Nacional Autónoma de México, pero el contenido que hay dentro no es el mismo hoy que el de hace 5 años; una URI en cambio mantiene de forma estática su contenido a pesar del tiempo.

señala quién es el propietario de esa información y así mismo hace referencia a otras fuentes de las que está tomando información de manera que da aún más contexto de vinculación en *Linked Data* y en buscadores semánticos, esta información está estructurada bajo el formato RDF.

Cada una de estas categorías implica diferentes ventajas y desventajas tanto para usuarios como para generadores de información. Bajo la premisa de que la información tiene que ser abierta y pública, las posibilidades con cada rango generan mayores beneficios y a la vez mayores dificultades para el uso de la información.⁴²

En el rango de una estrella, las ventajas para el usuario son que cualquiera puede consultar la información, imprimirla, almacenarla en cualquier dispositivo, modificarla y compartirla libremente; en cuanto a los generadores de información, la ventaja de este rango es que los documentos son fáciles de compartir y que las demás personas fácilmente saben que pueden hacer uso libre de esa información.

En el rango de dos estrellas los usuarios tienen todas las ventajas del rango anterior, además de eso se tiene la ventaja de tener la información en formatos ya estructurados que permiten la modificación con un esquema ya definido de cálculo, visualización, organización, modificación, etcétera, además permiten al usuario exportar la información en otros formatos también estructurados; para los generadores de información la ventaja sigue siendo la facilidad para publicar y compartir archivos públicamente, lo malo es que tal vez los usuarios necesiten algún tipo de software con licencia para consultar lo que hay dentro.

En el rango de las tres estrellas los usuarios tienen todas las ventajas de los rangos anteriores, además de que ahora es posible manipular la información de los archivos de la manera que se desee sin estar restringido por las capacidades de algún software en

⁴² “5-star Open Data”. Consultado el 5 de febrero de 2015. <http://5stardata.info/en/>

particular; como generador de información, la ventaja es que sigue siendo muy fácil de compartir y publicar; sin embargo, la desventaja es que cabe la posibilidad de requerir software o complementos adicionales si se desea exportar la información a software con licencia.

En cuanto al rango de cuatro estrellas los usuarios siguen teniendo todas las ventajas de los rangos anteriores, además de aprovechar las ventajas de tener el contenido en la web tales como vincular esa información de manera web o local, se puede archivar en los marcadores del navegador, se pueden reutilizar solo partes de la información, incluso pueden usarse librerías y herramientas específicas para analizar una parte de la información que se está exhibiendo, además se puede combinar la información que estemos consultando con la de otras fuentes, dado que las URI siguen un protocolo global, conectar la información ayuda a que los datos de cuatro estrellas puedan subir más fácilmente al siguiente rango; la desventaja del rango de cuatro estrellas es que es necesario entender la estructura de los diagramas de RDF, la cual puede ser más complicada de manejar que aquella que se encuentra en formatos más populares como Excel o XML.

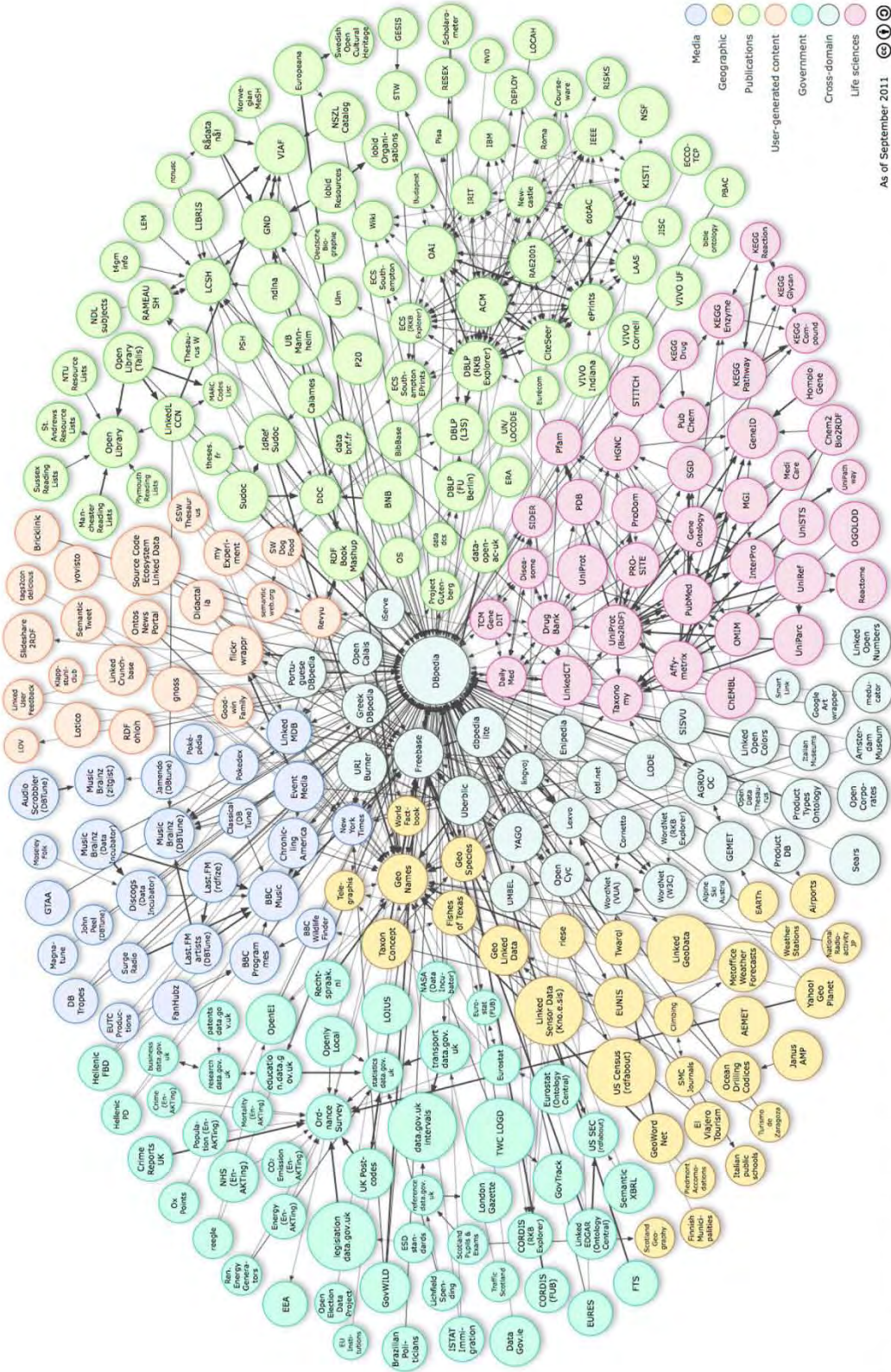
Para quienes generan la información la ventaja es utilizar una URI y tener la información dentro de la web con especificaciones y control sobre lo que se expone así como de la manera en que se va a exponer para los usuarios, al mismo tiempo está la oportunidad de que otros generadores de información pueden vincular información a la suya y ayudar a elevar el rango de estrellas de lo que se expone; por otro lado este nivel requiere más trabajo y constancia que solo publicar los datos, se necesita de una buena organización para representar correctamente la información en las URI que se crean con RDF de modo que se logre un mapa congruente de la información que se está exhibiendo y de con qué está siendo vinculada.

Por último, en el rango de las cinco estrellas, los usuarios tienen todas las ventajas que ofrece el grado anterior, pero al tratarse de datos vinculados correctamente con URI se hace posible descubrir aún más datos vinculados a lo que se está consultando, incluso se puede analizar la manera en que una información está siendo vinculada con otra y saber bajo qué parámetros se está haciendo, sin embargo al estar a un nivel tan detallado también cabe la posibilidad de encontrarse con URI dadas de baja o rotas, igual a como sucede con el error 404⁴³ en las páginas web convencionales, al mismo tiempo hay que tener en consideración que el contenido que se presenta en las URI al final del día es generado por seres humanos, de este modo las confusiones o mal interpretaciones pueden presentarse por lo que la confianza en la fuente que está proporcionando la información y el sentido común del usuario sigue siendo algo necesario.

Para los generadores de datos las ventajas en este último nivel se basan en la posibilidad de expandir de manera enorme sus propias bases de datos al tener la capacidad de descubrir las vinculaciones posibles con otras fuentes, esto al mismo tiempo hace que el valor de sus propias bases de datos crezcan, tanto al nivel de la información que contienen como de la calidad de la información que se les puede extraer, para un generador de datos una buena organización trae como beneficio un mayor número de usuarios que quieran vincular sus propias bases de datos en una gran base de bases de datos; por otra parte el manejo de todos estos recursos vinculados requiere de mayores servicios en la web, por lo que el gasto de tiempo, dinero y esfuerzo puede verse aumentado radicalmente entre más crezcan los datos y las vinculaciones que a ellos se generen, claro, sin contar la constante revisión y reparación de URI que puedan estar mal catalogadas o rotas.⁴⁴

⁴³ Un error 404 indica que nuestro navegador ha sido capaz de comunicarse con el servidor, pero no existe el fichero o recurso que ha sido solicitado.

⁴⁴ Entrevista con Javier Solís González, Gerente de desarrollo de nuevos productos y servicios en INFOTEC de CONACYT, abril, 2014



As of September 2011

Vínculos generados en Linked Data a partir de Open Data para alimentar DBpedia como centro para bases de datos científicas, gubernamentales y de entretenimiento en 2009.

Big Data.

La concentración de grandes volúmenes de información y el uso de *Open Data* y *Linked data* permiten a los expertos en informática generar herramientas dedicadas al análisis de patrones, tendencias e incluso a la realización de predicciones basadas en toda la información que se tiene almacenada y que en muchos casos se continúa generando en tiempo real día a día, es aquí donde entra *Big Data*, una tecnología que conglomerada y traduce toda la información de grandes bases de datos para generar respuestas a preguntas concretas en tiempo real.

Para entender un poco más sobre este tema entrevisté a David Ochoa, uno de los más conocidos periodistas y divulgadores sobre Tecnologías de la Información en México y conductor de “Byte Podcast”⁴⁵, el más reconocido programa de tecnología en habla hispana y más antiguo medio en llevar la cobertura del área tecnología en América latina. David comentó que Big Data es una herramienta que hoy por hoy podemos encontrar de formas de las que ni siquiera somos conscientes en nuestro día a día, desde que salimos a la calle y entramos a centros comerciales o transitamos las calles hasta cuando navegamos por la web o interactuamos en las redes sociales “toda la información que se genera y se captura de manera electrónica puede ser analizada por Big Data, cuando llegamos a un lugar hay contadores que registran nuestra entrada, cuando pedimos algo en el *fast food* esa orden del menú se convierte en información que se registra en una base de datos, incluso cuando vamos a la escuela nuestras calificaciones terminan en alguna base de datos, estos procesos antes quedaban en el papel y se almacenaban como la base de datos de una institución, hoy esos datos se digitalizan y se suben a la web donde se unen a otras bases de datos que permiten que tengamos mayor idea de las cosas acontecen en toda clase de rubros”.

⁴⁵ “Byte Podcast de Dixo en iTunes”. *iTunes*. Consultado el 24 de agosto de 2015.
<https://itunes.apple.com/mx/podcast/byte-podcast/id153006563?mt=2>

Recolectar estos gigantescos volúmenes de información en la web parece algo fácil, pero la realidad es que para poder hacer el análisis de toda esa información se requiere de computadoras extremadamente poderosas, hablamos de volúmenes de información que crecen diariamente en cantidades colosales, para ejemplificarlo más claramente se estima que solo durante un año los seres humanos estamos creando y almacenando información que al guardarla en CD's generaría una gran torre desde la Tierra hasta la Luna y de regreso.⁴⁶

“La generación de grandes volúmenes de información hace que se presenten dificultades para su procesamiento, es por eso que para organizarla se han generado 5 grandes grupos: *web & social media*, *machine to machine*, *big transaction data*, *biometric data* y *human generated data*, todos estos grupos componen las principales fuentes de creación de nuevos datos para *Big Data* y permiten a las máquinas y a los expertos en el manejo de datos a cruzar información con una lógica que responda a las preguntas que se desean plantear para extraer información con el uso esta tecnología”, comenta David.

Web y Social Media.

Es una de las fuentes más variadas y activas de la web, prácticamente cualquier sitio dentro de las redes sociales como Facebook, Twitter, LinkedIn, Tumblr, Snapchat, blogs, etcétera, se encargan de generar analíticas de los contenidos de sus usuarios así como de los usuarios mismos, el conteo de clics o *clickstream Data*, los feeds de twitter, las publicaciones de Facebook y el análisis del contenido de la web son prácticas diarias que se realizan con el objetivo de crear mediciones de lo que ocurre en las redes sociales así como medir el impacto y la temporalidad de la información así como su potencial de distribución en el contexto y tiempo que se genera.⁴⁷

⁴⁶ “¿Qué es Big Data?” CT316, el 18 de junio de 2012. Consultado el 20 de julio de 2014
<https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

⁴⁷ Mayer-Schönberger, Viktor, y Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013, pp 91-94.

Los datos que provienen de esta fuente varían entre estructurados y no estructurados, esto se debe a que se pueden medir con precisión campañas que han sido generadas con objetivos claros desde su inicio y pueden dar información sobre los clicks que se generan en su entorno, la clasificación y características del público que lo consume y comparte, el tiempo que se consume e incluso las preferencias específicas que tiene cada usuario que está recibiendo una información intencionalmente dirigida en una campaña; por otro lado la información no estructurada que se genera se debe a que no se puede dar un significado al contenido de toda la información que los usuarios crean, consumen y comparten, videos, imágenes, “memes”, textos que hacen referencia a situaciones particulares o sin contexto hacen que el análisis estructurado de esta información sea abundante y al mismo tiempo vago, ambiguo y complicado de clasificar con certeza.

Machine-to-machine (De máquina a máquina).

La captura de datos vía Machine-to-machine (conocida también como M2M) hace referencia a todos los instrumentos que logran generar y transmitir datos sin la necesidad de la intervención humana directa, herramientas basadas generalmente en sensores que pueden medir la presión atmosférica, la temperatura, la calidad del aire, los niveles de contaminación del agua y muchas otras cosas más relacionadas con el medio ambiente, también en este ramo se encuentran mecanismos robóticos del área de las manufacturas, la construcción o la agricultura, también el ramo de las telecomunicaciones genera información M2M al contabilizar y clasificar la información del número de personas que se conectan a una red, establecen contacto con una localidad o la cantidad de datos que fluye de un punto a otro, dicho de manera más simple la información generada por vía del M2M es toda aquella que se obtiene a partir la automatización de procesos y de máquinas que recopilan información a partir de analizar el entorno o la actividad concreta para la que están diseñadas, gracias a estos datos se puede conocer cuándo es necesario hacer el mantenimiento de una fábrica, cuándo es el mejor momento para sembrar o cosechar los cultivos o incluso en qué momentos es conveniente aumentar o disminuir la potencia de

la energía eléctrica para satisfacer mejor la demanda de los consumidores de la manera más económica y ecológica posible.⁴⁸ M2M se trata de cómo las máquinas analizan la información que generan otras máquinas.

Big Transaction Data.

Este tipo de información se refiere a todas aquellas operaciones donde se efectúe una relación de intercambio entre un solicitante y un servicio, desde operaciones bancarias, llamadas telefónicas, operaciones de compra venta o de intercambio.

Big Transaction Data ofrece información estructurada conforme a los productos y servicios con los que realiza transacciones a la vez que dentro de las bases de datos es capaz de hacer valer apartados legales y contables entre particulares, instituciones e incluso entre diferentes naciones, el objetivo de esto es lograr bases de datos estructuradas que logren dar consistencia a cantidades masivas de operaciones delimitadas por reglas consistentes sin que importe su procedencia o el tipo de intercambio.⁴⁹

Un ejemplo práctico para *Big Transaction Data* es la creación de bases de datos que den constancia del dinero facturado en un día específico en el comercio global basados en el tipo de cambio de un día o incluso hora específica, de esta manera aunque la paridad de una moneda en la que se realizó una operación de compra-venta cambie repentinamente el día de mañana el registro del valor podrá modificarse y exponerse en tiempo real con relación a las reglas del mercado en todo momento.

⁴⁸ Aguilar, Luis Joyanes. *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor, 2016. pp. 96-105

⁴⁹ Bhatnagar, Vasudha, y Srinath Srinivasa. *Big Data Analytics: Second International Conference, BDA 2013, Mysore, India, December 16-18, 2013, Proceedings*. Springer, 2013, pp 98-105.

El contenido de datos de *Big Transaction Data* es mayormente enfocado a los mercados y las finanzas públicas y su uso se expande a crear predicciones financieras, establecer índices de riqueza o pobreza, así como hacer cruce de información con otras bases de datos para obtener respuestas a asuntos socioeconómicos con reglas consistentes.

Biometrics (Datos biométricos).

Como su nombre lo indica son los datos capturados a partir de la medición de las características biológicas de un individuo, esta información incluye huellas digitales, escaneo de la retina, reconocimiento facial, genética, tipo de sangre, voz, etcétera.

Las mediciones biométricas sirven para generar relaciones fisiológicas entre individuos, estadísticas de salud o consumo, también esta información es de vital importancia para las agencias de seguridad e inteligencia, con ellos se puede utilizar como llave de seguridad la identidad de cada persona, la protección de datos confidenciales así como método de identificación universal y con alto rango de seguridad contra falsificación, al mismo tiempo puede ser utilizada con fines de rastreo de personas desaparecidas o buscadas por la justicia.⁵⁰

El uso de datos biométricos también ayuda a generar relaciones entre estímulos y expresiones faciales, gusto, disgusto, alegría, enojo y demás emociones pueden cuantificarse y relacionarse a partir de la captura estructurada de información biométrica.

Human Generated (Generados por humanos).

Se podría decir que los datos de esta categoría son aquellos que son creados por las personas y que fundamentalmente tratan sobre las personas, las llamadas grabadas en un

⁵⁰ “Biometrics.gov - Introduction to Biometrics”. Consultado el 24 de agosto de 2015.
<http://www.biometrics.gov/ReferenceRoom/Introduction.aspx>.

call center, las notas de voz que mandan por celular o internet, los correos electrónicos, los estudios médicos, los sitios web, documentos, presentaciones, hojas de cálculo.

Esta información en gran medida genera datos no estructurados que son incorporados para sus análisis con el fin de identificar individuos, rastrearlos y generar perfiles personalizados para cada experiencia de usuario, aun así bajo la propia lógica de Human Generated Data la captura no estructurada causa que un enorme volumen de información no pueda ser procesada de manera efectiva y provoque que mucha información quede como una incógnita.

Human Generated Data es valiosa para las agencias de mercadería que buscan ofrecer experiencias de publicidad personalizadas a sus clientes para aumentar sus ventas al ofrecer a cada usuario lo que quiere, necesita o seguramente le puede interesar. También es útil para encontrar perfiles profesionales específicos para tareas o áreas en las que a veces es más valioso conocer los datos personalizados y segmentados por intereses y habilidades de una persona más allá de lo que podría decir un currículum tradicional.

Suena muy bien, pero entonces ¿cómo puedo utilizar todo esto?

Hablar de tecnologías web y de la forma en que están construidas todas las herramientas que millones utilizamos todos los días nos hace ver la complejidad y dificultades que se esconden detrás de los usos cotidianos que damos al internet; aunque no hay que abrumarse, gracias a la apertura y enorme diversidad de la web es posible enterarse de todos los avances sin necesidad de entender demasiados detalles al respecto, más bien los usuarios no especializados en el desarrollo de la web encuentran un entendimiento con los avances en las tecnologías de la web en formas menos abstractas al encontrar todos los avances, herramientas y lenguajes ya ejecutados en forma de aplicaciones, sitios web, portales y servicios en línea con características claras y tangibles.

En el caso de las tecnologías que hemos abordado en este y los capítulos anteriores parece complicado que cualquier persona haga uso de ellas, pero no es así, miles de sitios web en el mundo ya están utilizando la web semántica con el apoyo de big data, linked data y open data de modo que hay muchas veces en las que recibimos datos, estadísticas o predicciones que parecen simples números, pero que en realidad están apoyadas por estas tecnologías que permiten obtener esa información de manera relativamente sencilla.

El ingeniero Javier Solís González de INFOTEC comentó al respecto: “muchas gente no lo sabe pero muchos portales del gobierno de México han apostado desde hace años por actualizar todas sus bases de datos a tecnologías semánticas, sitios como el de la secretaria de turismo,⁵¹ visit México,⁵² portal del empleo,⁵³ pro México,⁵⁴ la Secretaría de Educación Pública,⁵⁵ la Secretaría de Gobernación,⁵⁶ la Secretaría de Desarrollo Social,⁵⁷ el Instituto Mexicano de la Propiedad Industrial,⁵⁸ la Bolsa Mexicana de Valores,⁵⁹ incluso sitios comerciales como el portal de médica sur,⁶⁰ y más de otros 300 sitios tanto de gobierno como particulares con los que INFOTEC ha trabajado aplicando tecnologías semánticas.”

⁵¹ “Secretaría de Turismo | Gobierno | gov.mx”. Consultado el 24 de febrero de 2015.

<http://www.gob.mx/sectur/>

⁵² “Portal Oficial de Turismo en México | VisitMexico”. Consultado el 24 de febrero de 2015.

<http://www.visitmexico.com/>

⁵³ “Portal del Empleo: Bolsa de Trabajo en México”. Consultado el 24 de febrero de 2015.

<http://empleo.gob.mx/inicio.do>

⁵⁴ “ProMéxico”. Consultado el 24 de febrero de 2015. <http://www.promexico.gob.mx/>

⁵⁵ “Secretaría de Educación Pública | Gobierno | gov.mx”. Consultado el 24 de febrero de 2015.

<http://www.gob.mx/sep>

⁵⁶ “Secretaría de Gobernación | Gobierno | gov.mx”. Consultado el 24 de febrero de 2015.

<http://www.gob.mx/segob>

⁵⁷ “Secretaría de Desarrollo Social | Gobierno | gov.mx”. Consultado el 24 de febrero de 2015.

<http://www.gob.mx/sedesol>

⁵⁸ “Instituto Mexicano de la Propiedad Industrial | Gobierno | gov.mx”. Consultado el 24 de febrero de 2015.

<http://www.gob.mx/impi>

⁵⁹ “... Grupo BMV ... Inicio”. Consultado el 24 de febrero de 2015. <http://www.bmv.com.mx/>

⁶⁰ “Médica Sur: Excelencia médica, calidez humana”. Consultado el 24 de agosto de 2016.

<http://www.medicasur.com.mx/>

Por su parte, el periodista David Ochoa aseguró que para 2016 las tecnologías de web semántica apoyadas en Big Data se han vuelto algo cada vez más común, claro, haciendo énfasis en que los portales de grandes empresas y corporativos, instituciones civiles y de gobierno son las más interesadas y con capacidad de utilizar esta tecnología “esto es un tema de lenguajes de programación que pueden resolver las personas de forma individual, pero al mismo tiempo se necesita de recursos materiales y económicos para el procesamiento de la información, por lo que no cualquiera puede implementar estas tecnologías con total libertad”, este obstáculo no quiere decir que las puertas simplemente se han cerrado para el resto del público, ante esta necesidad han surgido empresas que se dedican específicamente a ofrecer servicios de procesamiento de datos con tecnologías semánticas y de big data, “las empresas pequeñas también generan información valiosa de la que se puede extraer conocimiento valioso para su desempeño, las empresas dedicadas al análisis de datos lo saben y por eso ofrecen servicios que van desde asesorías sobre cómo se debe capturar la información correctamente para su uso futuro en big data hasta proporcionar cursos sobre el propio manejo de los datos de la empresa para dar soluciones, propuestas o incluso dar diagnósticos de cosas que están pasando al extraer información de las bases de datos de la empresa sobre temas que ni siquiera estaban sobre la mesa al momento planear el análisis y que surgen como consecuencia de extraer información estructurada”.

Un ejemplo de todo lo anterior en México es Tableau,⁶¹ una empresa que justamente se dedica a analizar datos estructurados para extraer información en tiempo real sobre cientos de temas, “gracias al trabajo que realizan cualquiera de sus clientes pueden enterarse en tiempo real de lo que está ocurriendo con los datos que está generando su empresa, se pueden plantear preguntas específicas sobre esa misma información e incluso se pueden cruzar los contenidos de las bases de datos con información de otras empresas que pertenecen al mismo sector”.

⁶¹ “Tableau Software”. *Tableau Software*. Consultado el 4 de marzo de 2015. <http://www.tableau.com/>

Con el tiempo Tableau ha trabajado con información de muchos países tanto de particulares como de sectores públicos, esto le ha otorgado la capacidad de cruzar información de open data de manera pública y mostrar datos que en realidad son de interés general para todas las personas dentro de su plataforma Tableau Public,⁶² por ejemplo hay un gráfico que expone en tiempo real los lugares en el mundo donde existen leyes sobre el matrimonio igualitario en el mundo, sobre los mejores directores de cine con relación a los sitios más importantes de reseñas, un gráfico sobre el índice de alfabetización de las mujeres en el mundo, un mapa sobre la población de usuarios de facebook en el mundo, un mapa sobre los casos de diabetes que se reportan en cada país y de qué tipo, cosas más geeks como una gráfica que expone el universo de Marvel Comics con todos los personajes que lo conforman y la relevancia y porcentaje de apariciones que ha tenido cada uno de ellos hasta la fecha, el índice de hambre mundial, un índice de los hashtags generados durante la copa del mundo Brasil 2014 y el impacto que tuvo cada uno, una lista de todos los volcanes del mundo con la actividad diaria de cada uno de ellos, una grafica sobre los alimentos más comunes y la cantidad de fotos que la gente toma de ellos en instagram, así como muchas más gráficas con datos sobre gobiernos, instituciones e incluso personas.

Las aplicaciones para las tecnologías semánticas se acentúan con una de las más recientes creaciones de IBM en el campo de los sistemas de cómputo con Watson, un sistema diseñado para crear cómputo cognitivo con la capacidad de resolver preguntas de manera abierta y en lenguaje natural de los seres humanos al hacer consultas en grandes bases de datos con temas específicos para extraer información de grandes volúmenes de datos. El doctor Davod Godnek explica: “Watson no es una inteligencia artificial, es algo que en IBM planteamos como cómputo cognitivo, esto quiere decir que el sistema trabaja la información con la que se alimenta de una manera muy similar a la que hacemos los seres humanos, de esta manera al plantear una pregunta a Watson su sistema observa el

⁶² “Galería”. *Tableau Public*, el 1 de mayo de 2014. Consultado el 4 de marzo de 2015. <https://public.tableau.com/s/gallery>

problema, lo interpreta, evalúa sus posibles respuestas y decide lo que considera más acertado”.

Las aplicaciones que se le han dado al sistema Watson son tan variadas como las bases de datos que se le pueden dar a procesar y en consecuencia todo un campo de investigación se inaugura al utilizar este novedoso sistema de cómputo “en IBM trabajamos con instituciones comerciales y gubernamentales por igual, es por ello que con este sistema podemos analizar información muy variada al mismo tiempo que podemos cruzar información que muchas veces no es posible de manera rápida o accesible, por ejemplo, hemos trabajado con la universidad de Texas, en los Estados Unidos, para analizar y cruzar información entre todas las investigaciones sobre el cáncer que la universidad está generando, el resultado ha sido sorprendente, pues el cómputo cognitivo ha revelado datos sobre nuevos tratamientos a partir de cruzar informes de muchas de las investigaciones, este proceso de manera normal hubiera tardado meses o incluso años en desarrollarse y con la ayuda de Watson hemos obtenido información muy valiosa en cuestión de días”.



Al utilizar tecnologías semánticas es posible analizar bases de datos controladas y estructuradas para la generación de cómputo cognitivo, esto ayuda a generar nuevo conocimiento y a acelerar procesos complicados de investigación.

El cómputo cognitivo de Watson está ayudando a los departamentos de salud a analizar información que requeriría de muchas horas y recursos en ser analizada, con esto en mente científicos de todo el mundo en tan solo 3 años han iniciado proyectos con resultados sumamente interesante, tales como tratamientos contra el cáncer, reorganización de historiales clínicos para la mejora del tratamiento de pacientes en clínicas y hospitales, el análisis de patrones genéticos y su respuesta con medicamentos

específicos con la intención de dar una mejora en los tratamientos, incluso en estudios veterinarios se ha utilizado para poder encontrar tratamientos a padecimientos específicos por especie, raza y descripción de la historia clínica de un animal.

A la vez que el universo médico se ve beneficiado, otras áreas también se benefician con el cómputo cognitivo “debemos tener la mente abierta a las posibilidades que hay allí afuera, en este momento ya estamos trabajando con agencias de viajes que requieren establecer los mejores paquetes para sus clientes a partir de analizar cosas tan elementales como los precios de un lugar hasta las noticias que puedan afectar el desempeño de la visita, incluso se hace un análisis sociológico-cultural de los clientes para otorgar las experiencias que mejor se acomoden a lo que ellos están esperando de su paseo”.

De la misma manera se han encontrado ventajas comerciales al sistema cognitivo de Watson “un ejemplo es el trabajo que realizamos con Red Ant⁶³ donde hemos generado una app que permite a sus vendedores preguntar a qué tipo de clientes es preferible ofrecer las mercancías que tienen a disposición, con esto las personas pueden dirigir publicidad y ofertas directamente a los mercados que más les conviene para aumentar sus ganancias”.

Los sistemas semánticos que están dando sentido a la información de la web a la vez que a la información que generan empresas y gobiernos por igual están develando una fuente de conocimiento que permite crear soluciones a problemas cercanos de manera más rápida, hasta el momento hemos conocido cómo los sistemas de salud, comerciales e incluso de entretenimiento se benefician de estos sistemas, pero aún hay más tareas que pueden ser elaboradas por sistemas de cómputo cognitivos “uno de los proyectos más ambiciosos con Watson son los relacionados a seguridad, hablamos tanto

⁶³ “Red Ant - Connected store experiences for smarter retailers - mobile retail technology to deliver seamless omnichannel shopping experiences”. Consultado el 12 de marzo de 2015. <https://www.redant.com/>

de la parte de seguridad informática como de seguridad para las personas en el mundo real, uno de nuestros clientes es Spark Cognition⁶⁴ con el que hemos creado el primer sistema de seguridad informática basada en Watson, la intención es crear una defensa que se adapte a los ataques informáticos que un sistema puede recibir, para lograrlo hemos introducido una gran cantidad de manuales, documentos de buenas prácticas, informes forenses de casos exitosos de ataques así como mucho contenido similar de manera que Watson tiene información suficiente para generar un sistema de defensa informático que se acople a toda clase de ataque, incluso hemos visto como este sistema ha podido reconocer amenazas que consideramos un “falso positivo”; es decir, situaciones donde un sistema puede detectar una amenaza donde en realidad solo hay un fallo o una mala práctica por parte de los operadores de un sistema.”

La lista de aplicaciones para los sistemas semánticos es extensa, pero cabe con entender que allí donde existan actividades con información para generar y procesar habrá cabida para los sistemas de este tipo que permiten a las personas extraer información que bien puede pasar desapercibida o que puede requerir de muchos recursos y tiempo para ser adquirida, claro ahora de una manera rápida y confiable.

⁶⁴ “Home”. *SparkCognition Inc.* Consultado el 24 de agosto de 2016. <http://sparkcognition.com/>

Capítulo 4. Reflexiones y nociones para el futuro de la web semántica.

Para el año 2015 la web semántica se ha convertido en una tecnología prometedora en el campo de las ciencias computacionales que sirve de apoyo en disciplinas como la inteligencia artificial, las tecnologías web, las ingenierías de software, el comercio digital y la administración pública. Al mismo tiempo el uso de la web semántica ha comenzado a expandirse dentro de nichos específicos concernientes a disciplinas propias de las tecnologías circundantes a la web como en otras áreas del conocimiento que se sirven de esta herramienta para lograr avances útiles basados en la manera en la que utilizamos y organizamos la información.

La Web Semántica pretende crear una web extendida y dotada de mayor significado en la que cualquier usuario en Internet pueda encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a mecanismos de análisis, clasificación, organización y distribución de la información mejor definidos; sin embargo, alcanzar esta tecnología de manera abierta para todo el público y con toda la información que circula en la web aún no es posible, es por eso que esta titánica labor se ha dividido y especializado para en primer lugar establecer las reglas necesarias para que la web semántica sea posible y en segundo lugar para crear las primeras aplicaciones que sirvan de ejemplo para crear una gran red de redes y de bases de datos que ayuden a que cualquier persona pueda encontrar respuestas a sus preguntas.

A partir del desarrollo y evolución creciente de las tecnologías semánticas y de cómputo cognitivo basado en ellas es muy probable que en las próximas décadas veamos cambios importantes en el uso de los buscadores de la web en las que en vez de solo buscar por palabras en un buscador pasemos a en verdad plantear preguntas de manera abierta a un sistema de cómputo y que éste pueda entender y resolver nuestra petición.

Para entender lo que puede venir en el futuro hay que tener en claro nuestro presente, y es que entre las particularidades del uso de la web en esta segunda década del siglo XXI podemos encontrar que la información no sólo se centra en elementos importados del mundo real, también nos encontramos con manifestaciones que se gestan dentro del propio entorno de la web; estas manifestaciones sociales en la web adquieren significado para los usuarios y deberán otorgar también significado para los propios mecanismos semánticos; es en nuestra época donde la existencia de la web semántica mostrará una nueva capacidad de organizar y procesar la información para exponer significados hacia el exterior de un sistema computacional, estos procesos de significación a su vez se resinificarán al interactuar con el mundo real afectándolo de maneras que difícilmente podemos predecir, pero que es muy probable que afecte en el rumbo de importantes decisiones políticas, económicas, sociales y militares.

Este proceso creciente de organización y procesamiento de información acelera los mecanismos de reinterpretación, es decir, se acentúa la creación de un cerebro digital planetario donde incontables fuentes de información se enriquecen, unen y cruzan constantemente para resolver tareas que apoyan a las tecnologías semánticas para la captura, análisis, administración, extracción y generación de información que de otra manera requeriría mucho tiempo y recursos para obtenerse.

Para entender estos problemas de administración y creación de nueva información de manera simple podemos imaginar el modo en que funcionamos los seres humanos al momento de tratar de responder una pregunta gracias a un principio básico del cerebro conocido como heurística de reconocimiento, “escoge lo que conozcas”⁶⁵, con ello las personas podemos inferir lo que es correcto o no a partir de la información más relevante que conocemos (sin importar que esa información no tenga que ver con el cuestionamiento que se esté planteando), este proceso nos permite crear un contexto y

⁶⁵ “Redes para la Ciencia » Redes 122: Estamos programados, pero para ser únicos”. Consultado el 24 de abril de 2015. <http://www.redesparalaciencia.com/7141/redes/redes-122-estamos-programados-pero-para-ser-unicos>

predicciones más acertadas para entender y significar nuestra respuesta o procesar la información que vamos a recibir al obtener la respuesta correcta en caso de habernos equivocado, sin embargo ¿qué pasará con las máquinas y los agentes cognitivos que se están implementando para dar significado a la información de la web?

La solución que plantean sistemas como Watson de IBM, donde un sistema computacional basado en predicciones, probabilidad, análisis y estadística que pueden inferir cuál es la respuesta correcta a cualquier pregunta que se le plantee se basan en un principio similar basado en información estructurada y con significado que se pueda comparar para poder hacer un análisis que lleve a una solución del cuestionamiento que se le realizó.

Tareas como la predicción de mercados, análisis de documentos científicos, rastreo del consumo, la educación, el clima, cómputo capaz de comprender conceptos, relacionamientos de ideas, diálogo congruente, traducciones que obedecen al significado y no solo a la sustitución de palabras, sistemas de relacionamiento de información, analíticas de voz para detectar estados de ánimo, análisis de la personalidad de los usuarios e incluso la capacidad de hacer que los sistemas de cómputo puedan interpretar imágenes por su contenido y no solo por el texto que las describa son solo algunas de las funciones de las herramientas semánticas llevadas a la acción en pleno 2015.

La web semántica ya es una realidad y sus herramientas están impactando a niveles que parecen invisibles para los millones de usuarios que día a día nutren esa vasta red de información al utilizar sus dispositivos electrónicos, incluso el solo actuar con el entorno cotidiano en las grandes ciudades genera cantidades enormes de información con el solo pasar de los días. Cuando hablamos de ciudades inteligentes hablamos de herramientas semánticas, cuando hablamos del internet de las cosas hablamos de tecnologías semánticas, cuando compramos un artículo en las tiendas de auto servicio o en internet están las tecnologías detrás dando evidencia de las transacciones y generando

un análisis que seguro tendrá repercusiones en la toma de decisiones futuras para vendedores, productores e inversores.

Pareciera que el siglo XXI está aquí para romper las barreras entre la ciencia ficción imaginada por décadas al poner frente a la humanidad cada vez más herramientas que le permite una vida donde los sistemas de cómputo, la web y las personas coexisten de manera natural en el día a día de manera que conviven en armonía y de manera necesaria para la preservación del desarrollo de los individuos, grupos y naciones por igual.

Las posibilidades parecen ilimitadas cuando observamos que las tecnologías semánticas ya son aplicadas en áreas de salud, economía, tecnología, educación, entretenimiento, cultura, ciencia, gobierno e incluso en cuestiones de seguridad informática y seguridad militar, ahora solo el tiempo y las personas decidirán cuál será el resultado de conocer, controlar y administrar la información que generamos todos los días para la construcción del futuro de la humanidad.

Bibliografía.

- Aguilar, Luis Joyanes. *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor, 2016.
- Aguilar, Luis Joyanes. *Metodología de la programación*. Mc-Graw Hill, México, 1991.
- Bain, Ken. 2007. *Lo que hacen los mejores profesores universitarios*. Universitat de Valencia. Valencia.
- Begoña Echevarría Llombart, *Las W's del reportaje*, Valencia, Fundación Universitaria San Pablo CEU, 1988, p. 134
- Berners-Lee, Tim, y Mark Fischetti. *Tejiendo la red: el inventor del world wide web nos descubre su origen*. Siglo XXI de España, 2000.
- Cristóbal Cobo Romanió; Hugo Pardo Kuklinski, *Web 2.0. Inteligencia colectiva o medios fast food*, México DF/Barcelona, Planeta Grup de Recerca d'Interaccions Digitals, Universitat de Vic. Flacso México, 2007, pp. 140
- DiBona, Chris, Sam Ockman, Mark Stone. *Open Sources: Voices from the Open Source Revolution*. San Francisco: O'Reilly, 1999.
- Gonzalo Martín Vivaldi, *Géneros periodísticos*, Madrid, Paraninfo ediciones, 1998, pp. 400
- John Hebel, Matthew Fisher, Ryan Blace, Andrew Perez-Lopez. *Semantic Web Programming*. Wyley Publishing. Primera edición, 2009
- Kuklinski, Hugo Pardo. *Geekonomía: un radar para producir en el postdigitalismo*. Edicions Universitat Barcelona, 2010.
- Manuel Castells, *La Galaxia Internet*, España, Areté, 2001, pp. 316
- Mar de Fontcuberta, *La noticia, pistas para percibir el mundo*, Barcelona, Paidós, 2000, pp. 153
- Mayer-Schönberger, Viktor, y Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt, 2013.

- McLuhan, Marshall. 1964. *Understanding media: The extensions of man*. McGraw Hill. Nueva York
- Raúl Trejo Delarbre, *Viviendo en el Aleph*, España, Gedisa, 2006, pp 254
- Terréense W. Pratt, Marvin V, Zelkowitz. *Lenguajes de programación: diseño e implementación*. Prentice-Hall Hispanoamericana, S.A. Tercera edición, México 1998.
- Vicente Leñero y Carlos Marín, *Manual de Periodismo*, México, Tratados y manuales Grijalbo, 1986, pp. 315

Cibergrafía.

- “... Grupo BMV ... Inicio”. Consultado el 24 de febrero de 2015.
<http://www.bmv.com.mx/>
- “¿Qué es Big Data?” CT316, el 18 de junio de 2012. Consultado el 20 de julio de 2014 <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>
- “1 Second - Internet Live Stats”. Consultado el 8 julio 2014.
<http://www.internetlivestats.com/one-second/#traffic-band>
- “5-star Open Data”. Consultado el 5 de febrero de 2015. <http://5stardata.info/en/>
- “Biometrics.gov - Introduction to Biometrics”. Consultado el 5 de febrero de 2015.
<http://www.biometrics.gov/ReferenceRoom/Introduction.aspx>
- “Byte Podcast de Dixo en iTunes”. *iTunes*. Consultado el 24 de agosto de 2015.
<https://itunes.apple.com/mx/podcast/byte-podcast/id153006563?mt=2>
- “DBpedia”. Consultado el 13 de agosto de 2015. <http://wiki.dbpedia.org/Datasets>
- “Digitalismo.com”. <http://digitalismo.com/>
- “Galería”. *Tableau Public*, el 1 de mayo de 2014. Consultado el 4 de marzo de 2015.
<https://public.tableau.com/s/gallery>
- “Google Code Archive - Long-term storage for Google Code Project Hosting.”
Consultado el 12 de agosto de 2015. <https://code.google.com/archive/p/tdwg-rdf/wikis/Beginners6SPARQL.wiki>
- “Home”. *SparkCognition Inc*. Consultado el 24 de agosto de 2016.
<http://sparkcognition.com/>
- “Instituto Mexicano de la Propiedad Industrial | Gobierno | gob.mx”. Consultado el 24 de febrero de 2015. <http://www.gob.mx/impi>
- “Internet users in real time - Internet Live Stats”. Consultado el 8 julio 2014.
<http://www.internetlivestats.com/watch/internet-users/>

- “Japan”. *Wikipedia, the Free Encyclopedia*. Consultado el 13 de agosto de 2015. <https://en.wikipedia.org/w/index.php?title=Japan&oldid=735771550>
- “Médica Sur: Excelencia médica, calidez humana”. Consultado el 24 de agosto de 2016. <http://www.medicasur.com.mx/>
- “Mexico”. *Wikipedia, the Free Encyclopedia*. Consultado el 13 de agosto de 2015. <https://en.wikipedia.org/w/index.php?title=Mexico&oldid=735470769>
- “OWL Web Ontology Language Overview”. Consultado el 15 de agosto de 2015. <https://www.w3.org/TR/owl-features/>
- “Portal del Empleo: Bolsa de Trabajo en México”. Consultado el 24 de febrero de 2015. <http://empleo.gob.mx/inicio.do>
- “Portal Oficial de Turismo en México | VisitMexico”. Consultado el 24 de febrero de 2015. <http://www.visitmexico.com/>
- “ProMéxico”. Consultado el 24 de febrero de 2015. <http://www.promexico.gob.mx/>
- “Red Ant - Connected store experiences for smarter retailers - mobile retail technology to deliver seamless omnichannel shopping experiences”. Consultado el 12 de marzo de 2015. <https://www.redant.com/>
- “Redes para la Ciencia » Redes 122: Estamos programados, pero para ser únicos”. Consultado el 24 de abril de 2015. <http://www.redesparalaciencia.com/7141/redes/redes-122-estamos-programados-pero-para-ser-unicos>
- “Secretaría de Desarrollo Social | Gobierno | gob.mx”. Consultado el 24 de febrero de 2015. <http://www.gob.mx/sedesol>
- “Secretaría de Educación Pública | Gobierno | gob.mx”. Consultado el 24 de febrero de 2015. <http://www.gob.mx/sep>
- “Secretaría de Gobernación | Gobierno | gob.mx”. Consultado el 24 de febrero de 2015. <http://www.gob.mx/segob>

- “Secretaría de Turismo | Gobierno | gob.mx”. Consultado el 24 de febrero de 2015. <http://www.gob.mx/sectur/>
- “Tableau Software”. *Tableau Software*. Consultado el 4 de marzo de 2015. <http://www.tableau.com/>
- “Virtuoso SPARQL Query Editor”. Consultado el 12 de agosto de 2015. <http://dbpedia.org/sparql>
- Barnett, Emma. “Google reveals caffeine: a new faster search engine”, el 11 de agosto de 2009. Consultado el 10 de junio de 2014 <http://www.telegraph.co.uk/technology/google/6009176/Google-reveals-caffeine-a-new-faster-search-engine.html>
- Berners-Lee, Tim. *Tim Berners-Lee acerca de la futura Web*. Consultado el 29 de enero de 2015. http://www.ted.com/talks/tim_bern timers_lee_on_the_next_web?language=es
- Berners-Lee, Tim. *Tim Berners-Lee acerca de la futura Web*. Consultado el 29 de enero de 2015. http://www.ted.com/talks/tim_bern timers_lee_on_the_next_web?language=es
- Burn-Murdoch, John. “Study: Less than 1% of the World’s Data Is Analysed, over 80% Is Unprotected”. *The Guardian*, el 19 de diciembre de 2012, sec. News. Consultado el 11 de agosto de 2014 <https://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume>
- Burn-Murdoch, John. “Study: Less than 1% of the World’s Data Is Analysed, over 80% Is Unprotected”. *The Guardian*, el 19 de diciembre de 2012, sec. News. Consultado el 17 de junio 2014. <https://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume>
- Bush, Vannevar. “As We May Think”. *The Atlantic*, julio de 1945. <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

- Business, Author: Klint Finley Klint Finley. “Open Source Is Going Even More Open—Because It Has To”. *WIRED*. Consultado el 28 de julio de 2015.
<http://www.wired.com/2015/07/open-source-going-even-openbecause/>
- Edrerra-Viedma, José M. Morales-del-Castillo. *Referencias a la web semántica desde la perspectiva de la documentación*. Revista Investigación Bibliotecología, vol 21, num 43, Julio/diciembre, 2007, México. Visto el 10 de octubre de 2014.
<http://www.revistas.unam.mx/index.php/ibi/article/view/4138/3672>
- IBM. *IBM and the Jeopardy Challenge*, 2010. Consultado el 8 de diciembre de 2014
<https://www.youtube.com/watch?v=FC3IryWr4c8>
- Revista Digital Universitaria, 10 de julio de 2005, Vol.6, No.7 ISSN: 1607 – 6079, Publicación mensual. Consultado el 1 de octubre de 2014.
<http://www.revista.unam.mx/vol.6/num7/art72/art72-3.htm#a>
- Stallman, Richard. *Los piratas son los padres. Historias en los albores de la era digital*. Exgae, Barcelona, 2008.
http://exgae.ney/docs/Los_piratas_son_los_padres.pdf
- Tecnológico, INFOTEC Centro Público de Investigación y Desarrollo. “INFOTEC - Web 1.0, Web Social (2.0), La Web Semántica (3) y... ¿qué sigue?”. Consultado el 18 de agosto de 2014.
https://www.infotec.mx/es_mx/infotec/Web_10_Web_Social_20_La_Web_Semantica_3_y_que_sigue
- Tecnológico, INFOTEC Centro Público de Investigación y Desarrollo. Consultado el 18 de agosto de 2014.
http://infotec.com.mx/en_de/infotec/articulo_web_semantica