

## El Valor de los Datos



**German Braun**  
Licenciado en Ciencias  
de la Computación  
Becario Doctoral CONICET

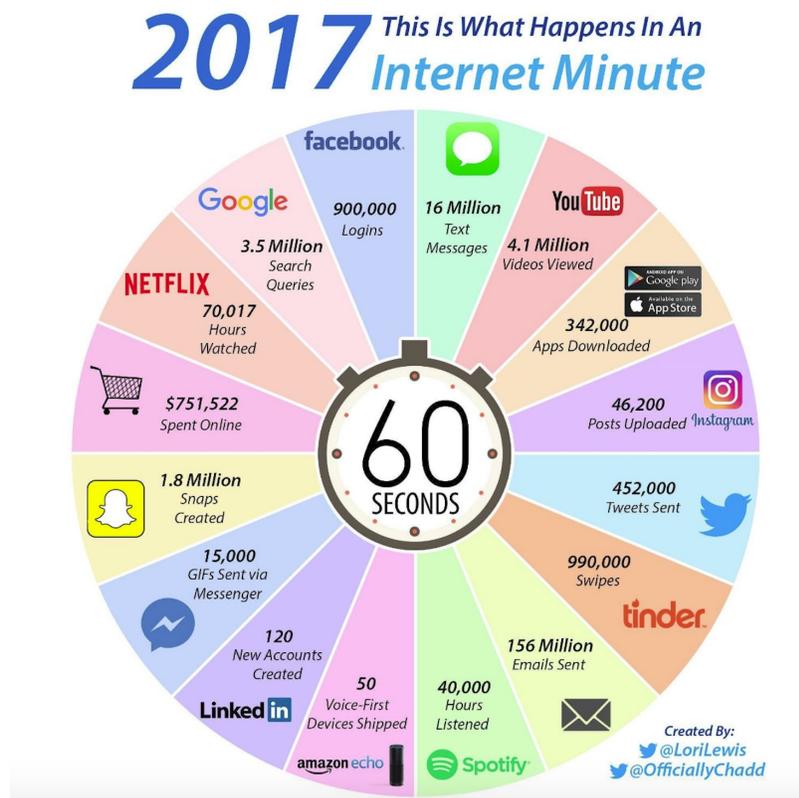


**Laura Cecchi**  
Doctora en Ciencias de la Computación

\*\*Ambos del Grupo de Investigación en Lenguajes e Inteligencia Artificial (GILIA). Facultad de Informática.

*“La Web Semántica es una extensión de la Web actual en la cual la información es presentada con un significado bien definido, entendible por las computadoras para que los humanos trabajen en cooperación”*

Sir Tim Berners-Lee



Internet ha penetrado en nuestras vidas en forma sorprendente, alcanzando no sólo el área del trabajo sino también del entretenimiento, de la salud, de la educación y de las finanzas, entre otras. Gracias a esta tecnología, la Web es accedida en forma sencilla, continua e inmediata, permitiendo a los usuarios ser consumidores de datos e incluso ser generadores de ellos. Si bien, el uso de Internet alcanza a 3800 millones de personas sobre una población total de 7800 millones, según datos de 2017, el 90 % de los datos de la Web actual fueron generados durante los últimos dos años y continúa en

constante crecimiento a una tasa de 2,5 quintillones de bytes de datos por día. Para tener una dimensión real, vamos a suponer que un carácter de un tweet (de 140 caracteres) que se publica en Twitter es un byte y supongamos también que ese byte representa un milímetro cuadrado. Por lo tanto, la masa de tweets generados en un día (~ 70 millones) equivale a "llenar" el Gran Cañón del Colorado, cuya longitud es de 500 kms con una profundidad máxima de 800 metros, con esos datos. ¡Sorprendente!<sup>1</sup>

### Ahora bien ... ¿Para qué sirven estos datos?

Básicamente, los datos nos sirven para tomar decisiones y mientras más "comprendamos" estos datos, mejores decisiones podremos tomar. La combinación de estos datos permite lograr diferentes clases de entendimiento sobre un dominio de estudio. Esto hace que Internet no sólo sea una ventaja, por la cantidad significativa de datos que provee, sino también una responsabilidad en cuanto a la gestión del conocimiento.

Los datos cobran vida cuando se extrae información a partir de ellos. Esta información nos sirve para tener conocimiento sobre un tema particular y adquirir sabiduría. Estos conceptos están formalizados en la pirámide **DIKW** (por sus siglas en inglés, data, information, knowledge and wisdom), que representa la relación entre ellos: *"la información es definida en términos de datos, el conocimiento en términos de información y la sabiduría en términos del conocimiento"*. Desde un punto de vista más formal, la información es inferida a partir de los datos generados por un determinado fenómeno y es el resultado de descubrir el significado, propósito y relaciones básicas entre ellos. En un nivel superior, el conocimiento es la transformación de información en revelaciones que nos indican cómo dicho fenómeno está constituido, cómo funciona y cómo podemos controlarlo. Finalmente, la sabiduría es un concepto más abstracto relacionado a entender cómo los fenómenos que estudiamos están insertos en un contexto real más complejo, cuáles son sus metas y peligros y cómo impactan en la sociedad y en la vida de las personas.

Para poner en valor a los datos, veamos un ejemplo. En 2015, investigadores de Princeton y Harvard, evaluaron el uso de datos móviles para determinar la estacionalidad de la rubéola en Kenia<sup>2</sup>. Usando la ubicación de la torre de enrutamiento y el tiempo de cada llamada y mensaje de texto (**datos**), los investigadores determinaron la ubicación diaria de cada usuario y el número de viajes que estos realizaron entre las regiones cada día (**información**). En total, más de 12 mil millones de comunicaciones de telefonía móvil fueron registradas de forma anónima y vinculadas a una provincia. Luego, comparando el análisis de teléfono móvil con un conjunto de datos muy detallados sobre la incidencia de la rubéola en Kenia, los investigadores encontraron que los patrones de movimiento del teléfono móvil se alineaban con las cifras de incidencia de

---

<sup>1</sup> <https://www.domo.com/learn/data-never-sleeps-5>

<sup>2</sup>

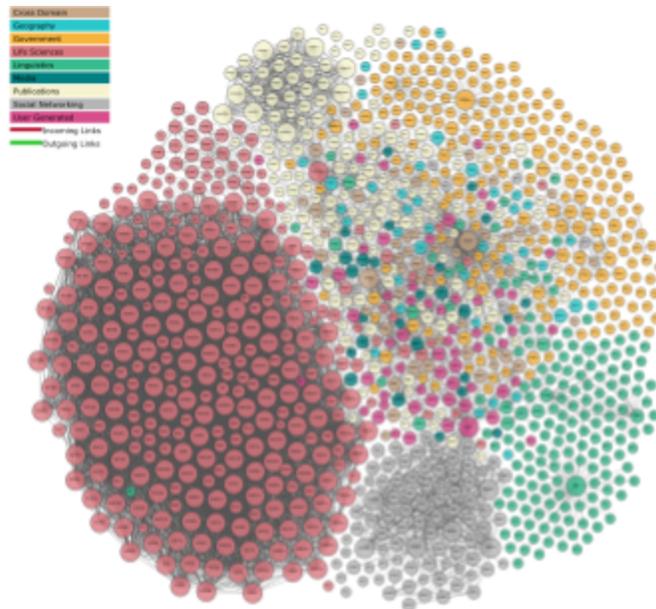
[http://www.tendencias21.net/Datos-de-telefonía-para-predecir-los-picos-de-rubeola-en-Kenia\\_a41032.html](http://www.tendencias21.net/Datos-de-telefonía-para-predecir-los-picos-de-rubeola-en-Kenia_a41032.html)

la rubéola (**conocimiento**). Esto mostró a los investigadores que el movimiento del teléfono móvil puede ser un predictor de la propagación de la enfermedad infecciosa y entender cómo estos fenómenos impactan socialmente (**sabiduría**).

### ¿Qué estamos haciendo para usar los datos?

Las Ciencias de la Computación están inmersas en la búsqueda de soluciones para capturar más y más datos sobre el mundo y en la búsqueda de paradigmas multidisciplinares, formalismos y técnicas computacionales que ayuden a procesarlos y a extraer nuevo conocimiento.

Uno de estos conceptos que hoy en día se encuentra en pleno desarrollo es el de **Web Semántica**. La Web Semántica fue propuesta por Sir Tim Berners-Lee y su finalidad es permitir que las computadoras puedan entender el significado de los datos y sacar conclusiones sobre ellos. En este nuevo contexto, fue necesario definir nuevas bases teóricas para estructurar los datos y enlazarlos entre sí, dando lugar a una **Red de Conocimiento**<sup>3</sup>, el cual es central en el desarrollo de la Web Semántica. Esta red enlaza datos en vez de documentos. Es decir, cada dato que define un concepto del mundo real, es enlazado (mediante un link convencional) con



La imagen muestra el estado actual de la Red de Conocimiento generada a partir de la publicación de datos enlazados en la Web. El tamaño de los círculos corresponde al número de arcos que conectan cada conjunto de datos.

otras definiciones. Finalmente, un software, llamado **agente inteligente**, navega la Web sin intervención humana directa recorriendo esta red de datos enlazados. Estos agentes coleccionan datos, los filtran y los procesan para sacar conclusiones. Mientras que los agentes navegan la Web, interactúan con otros. Para poder comunicarse, los agentes requieren realizar **acuerdos** sobre la denominación y el significado de cada objeto y de las relaciones que existen entre ellos. De este modo, todos los agentes hablan el mismo lenguaje.

En Ciencias de la Computación, estos acuerdos son denominados **ontologías** y su finalidad es definir el conocimiento de un dominio del mundo real, para que los agentes puedan entender el

<sup>3</sup> <http://lod-cloud.net/>

significado de los datos de manera no ambigua. Las ontologías pueden mejorar el funcionamiento de la Web de diversas maneras. Por un lado, son usadas para mejorar las búsquedas dando un significado preciso a los términos empleados y mejorando la exactitud de los resultados, en contraposición a las búsquedas Web actuales basadas en comparaciones sintácticas. Asimismo, aplicaciones más avanzadas de las ontologías también permiten relacionar información de diversas fuentes distribuidas a lo largo del mundo. Finalmente, si bien la computadora no “entiende” realmente la información, las ontologías permiten la manipulación efectiva de terminologías o vocabularios de manera tal que los resultados son entendibles por los usuarios humanos.

### **Tecnologías vinculadas al uso intensivo de datos.**

Actualmente existen diferentes tecnologías vinculadas al uso de grandes cantidades de datos y que impactan fuertemente en la sociedad actual y en su desarrollo:

- **Big Data.** Análisis, administración y manipulación inteligente de grandes cantidades de datos a través de modelos de descripción, predicción y optimización para una mejor y más eficiente toma de decisiones.
- **Internet de las Cosas (IoT).** Es la interconexión de artefactos, embebidos con software, sensores, actuadores y redes de conectividad que permiten la recolección e intercambio de datos entre ellos.
- **Industria 4.0.** Tendencias actuales de automatización e intercambio de datos en tecnologías aplicadas a la manufacturación. Incluye IoT y computación cognitiva y en la nube.
- **Ciudades Inteligentes (Smart Cities).** Se basa en el uso intensivo de Tecnologías de la Información y Comunicación en la prestación de servicios públicos de alta calidad, seguridad, productividad, formación, participación e innovación.

El volumen, la variedad y la velocidad de crecimiento de los datos se presentan como un nuevo desafío, que tiende a un enfoque cultural del conocimiento. Extraer el valor de los datos es el primer eslabón hacia el desarrollo de tecnologías que traten con realidades sociales reales y complejas. Dichas tecnologías promueven también el uso de datos abiertos, impulsando la participación y colaboración social, de manera que el conocimiento contenido en estos datos nos ayude a tomar las decisiones correctas.

### **Bibliografía**

Rowley, Jennifer. ["The wisdom hierarchy: representations of the DIKW hierarchy"](#). *Journal of Information and Communication Science*. 33 (2) 2005

Amy Wesolowski, C. J. E. Metcalf, Nathan Eagle, Janeth Kombich, Bryan T. Grenfell, Ottar N. Bjørnstad, Justin Lessler, Andrew J. Tatem, Caroline O. Buckee. [Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data](#). *Proceedings of the National Academy of Sciences* 2015.

Berners-Lee, Tim; James Hendler; Ora Lassila (May 17, 2001). "[The Semantic Web](#)". *Scientific American Magazine*. Retrieved March 26, 2008