

RICARDO BAEZA-YATES, CIENTÍFICO COMPUTACIONAL

“El diseño de algoritmos tiene una carga ética enorme”

Más de 4.000 millones de personas utilizan internet todos los días. En la era de la posverdad, ¿somos conscientes de los sesgos de la información en internet? ¿En qué consiste la búsqueda semántica? ¿Tienen ética los algoritmos? ¿Cómo va a transformar nuestras vidas el internet de las cosas? El especialista chileno Ricardo Baeza-Yates nos da algunas claves sobre estos temas candentes.

Patricia Luna

2/6/2018 08:00 CEST



Ricardo Baeza-Yates. / SINC

“La web es el conjunto público de datos más importante de la historia que hemos creado entre todos. Tenemos que ser conscientes de en manos de quién está esa información. El control lo tienen los **gobiernos**, las **compañías de teléfono**, los **proveedores de internet**, los **bancos** y nuestras **redes sociales**”, dice a Sinc Ricardo Baeza-Yates (Santiago de Chile, 1961). Este científico computacional, exdirector de investigación de Yahoo, es el actual director de tecnología de [NTENT](#), una de las empresas de EE UU que lidera el

desarrollo de búsqueda semántica inteligente.

El experto en tecnologías de búsqueda y análisis de datos de la web destaca que “el **poder de los datos** está distribuido en la red y todavía no hay un monopolio, pero esto podría cambiar en un futuro cercano, en el que el internet de las cosas transformará de manera drástica el mundo en que vivimos”.

“Tenemos que ser conscientes de en manos de
quién está la información de la web”

Comenta que ahora “muchas personas van a la web a buscar información, pero son pocas las que la aportan. El porcentaje de gente activa en internet es muy pequeño, alrededor de un 5%. Sin embargo con el internet de las cosas todo estará conectado, ya que habrá cientos de sensores enviando señales. Será por lo menos dos órdenes de magnitud mayor de lo que tenemos, lo que supondrá un gran problema para filtrar los datos o procesarlos en tiempo real”.

Más de 4.000 millones de personas usan internet cada día. ¿Cómo manejamos esa cantidad ingente de datos que hay en la web?

Los sesgos de la información de la red

Antes de la era de la posverdad, Baeza estudió durante mucho tiempo los sesgos que pueden afectar a la información, de todo tipo: geográficos (en países centralizados siempre hay más información de la capital que del resto del país), lingüísticos (la mitad de la información de la web está en inglés aunque la mitad del mundo no habla esta lengua) o de género (hay muchas menos biografías femeninas que masculinas en Wikipedia), educacionales o económicos.

“Un tema que me preocupa es que la gente usa datos suponiendo que están completos y son fiables sin siquiera verlos, y muchas veces eso no es cierto. Si uno tiene conciencia del sesgo en la información, la puede analizar de forma correcta. Es importante que sepas en qué lado estás”, subraya.

Sin embargo, agrega, “hay sesgos que son incluso más complicados, que tienen que ver con el mundo de las **recomendaciones**, similares a las que usa Amazon, porque la recomendación que hace el algoritmo afecta a tus decisiones futuras. Es decir, que vas a escoger una de las tres opciones que te doy a elegir y ninguna de las que no te presento”.

“Si no tienes cuidado los algoritmos pueden sesgar tu opinión y llevarte a un autoengaño”

“Son los llamados **sesgos algorítmicos**, que, de manera similar a cuando hacemos predicciones, pueden llevarnos a una especie de autoengaño. Si no tienes cuidado el algoritmo puede sesgar tu opinión y a su vez esa opinión tuya afecta la opinión del algoritmo”, señala.

Es aquí donde surge el trabajo ético de los diseñadores de algoritmos, que han de tener en cuenta este efecto para hacer algo al respecto. ¿Tienen entonces ética los algoritmos, o, en su defecto, es importante que la tengan aquellos que los diseñan?

“Si la persona que escribe el algoritmo es consciente de sus sesgos lo puede hacer equitativo, introducir un equilibrio”, apunta. “Por supuesto, hay mucha carga ética de la que muchos programadores no son conscientes”.

Conversaciones inteligentes con un buscador

Baeza-Yates se dedica a la llamada **búsqueda semántica inteligente**, aquella que permitirá que los buscadores sean capaces de inferir relaciones y responder preguntas, el primer paso de lo que podría ser una conversación inteligente con un buscador.

NTENT usa el aprendizaje de las máquinas para predecir la intención de los usuarios, a partir de una serie de complejas tecnologías y plataformas en tres idiomas inglés, ruso y turco.

“Lo que hacemos –agrega– es tecnología de búsqueda semántica, que permite buscar entendiendo el idioma de la persona y su contexto, por ejemplo, la ciudad donde está y lo que ha hecho antes. Esto permite dar respuestas mucho más aproximadas a lo que se está buscando”.

La diferencia entre un buscador tradicional (o léxico) y uno semántico es que mientras en el primero se usan palabras para dirigir las preguntas del usuario a los documentos relevantes, en el semántico el foco se desplaza a los conceptos. Es decir, el buscador no solo reconoce palabras, sino que también entiende lo que significan y cómo se relacionan con otras en un documento.

“La búsqueda semántica permite buscar información entendiendo el idioma de la persona y su contexto”

“En términos tecnológicos significa que hay que usar una **ontología**, que los grandes buscadores también usan, pero en que en nuestro caso es vital; es una ontología que entiende conceptos y relaciones entre estos, y hacerlo en más de un idioma. Usamos **aprendizaje automático**, *machine learning*, para entender el contexto y mejorar el orden de resultados”, explica el científico.

Muchos de los temas más difíciles de resolver en idiomas no tan populares como el inglés los lleva adelante un grupo de I+D, de investigación y transferencia tecnológica que la compañía tiene [en Barcelona](#) y que dirige el propio Baeza-Yates.

Talento fuera de Silicon Valley

“Fuera de Estados Unidos podemos encontrar expertos en lingüística computacional en idiomas que no son tan frecuentes como el inglés, como el turco. Actualmente, es más fácil encontrar talento fuera, porque la competencia es menor y además Silicon Valley debe de ser ahora el lugar más caro para contratar”, explica.

Destaca que “en los últimos dos años el número de empresas emergentes en

Barcelona ha aumentado muchísimo y es un buen lugar para hacer ciencia de datos. Se está convirtiendo en uno de los centros neurálgicos de Europa en estos temas, después de Londres y Berlín”, destaca.

Entre las aplicaciones futuras de la búsqueda semántica, apunta a la “**búsqueda vertical**, es decir, la que se hace en un ámbito donde las personas pueden expresarse en lenguaje natural y preguntar en voz alta. Por ejemplo, podrías entrar en tu casa y preguntar si ha venido alguien mientras no estabas y el sistema te contestará sí o no. En realidad, tiene aplicaciones en cualquier situación donde en lugar de una persona puedas tener un agente inteligente”.

“Con el internet de las cosas esto va a ser más importante. Imagina la cantidad de sensores que vas a tener, uno no puede estar comunicándose con cada sensor, necesitaremos un mediador entre las personas y los sensores que están conectados a su vida. Esta tecnología podría cumplir ese rol, podría mediar con todos los sensores que están conectados a ti de alguna manera, los de tu móvil, los de tu casa y los de tu coche”, aventura.

Copyright: **Creative Commons**

TAGS

ALGORITMOS | COMPUTACIÓN | ÉTICA | INTERNET |

Creative Commons 4.0

You can copy, distribute and transform the contents of SINC. [Read the conditions of our license](#)