



BERT, el cambio en el algoritmo de Google que mejora las búsquedas, llega a más de 20 países

En septiembre Google dio a conocer un cambio importante en su algoritmo que permite optimizar las búsquedas, por medio de la técnica denominada BERT, que favorece la interpretación de las frases en contextos. Hoy se anunció que esta herramienta, que hasta ahora sólo estaba disponible en inglés, comenzará a estar disponible en español, portugués entre otros y desembarca, además, en más de 20 países.

Pero para entender el impacto de este cambio primero hay que hacerse una pregunta clave: ¿cómo hace el buscador de [Google](#) para procesar tanta información? [El secreto está en](#)

[el algoritmo](#), esa fórmula (o “fórmulas”) que permite procesar las millones de consultas que recibe a cada instante.

MÁS SOBRE ESTE TEMA: [La fórmula detrás del algoritmo de Google, el secreto mejor guardado de la compañía](#)

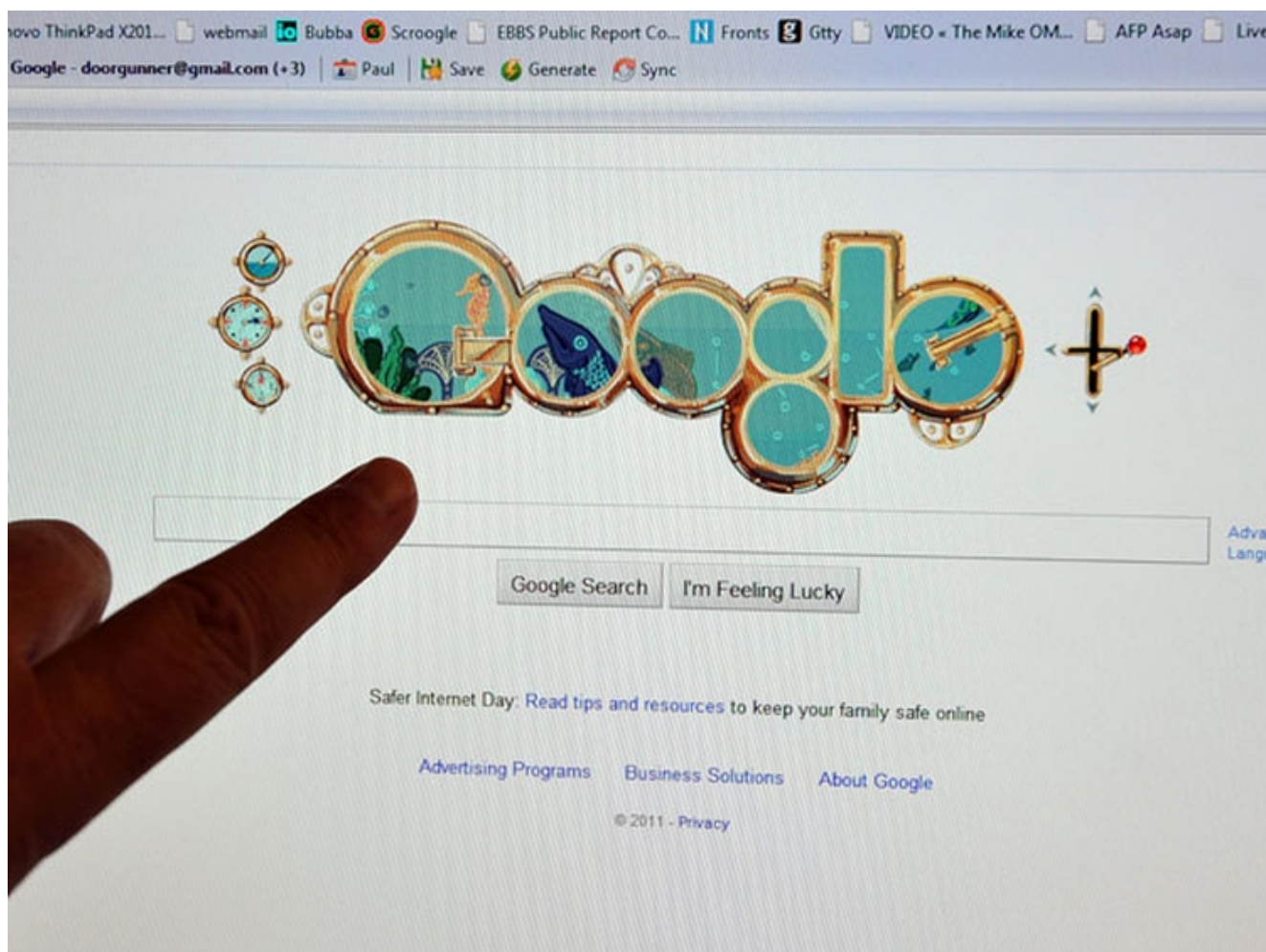
Este algoritmo, que dio inicio a un de las compañías más grandes del planeta, se fue optimizando a lo largo de los años. Y uno de los cambios más grandes que atravesó desde su creación fue el anunciado en septiembre. Con esta nueva optimización se busca lograr mejoras en el procesamiento de la información para lograr cada vez respuestas más precisas.

Después de todo, el buscador dice tener como misión organizar la información la información del mundo y que esté disponible. Para ir tras este objetivo, a lo largo de los años se fueron generando múltiples avances de la mano del **machine learning** que permite una mejor comprensión del lenguaje, teniendo en cuenta el contexto y la forma de los enunciados de búsqueda.

En este sentido, hace unas semanas **Google anunció que abriría a la comunidad el acceso a una nueva técnica** para el pre entrenamiento en el procesamiento de lenguaje natural llamada Representaciones de codificador bidireccional de transformadores, o **BERT**.

El gran aporte de BERT es que permite una interpretación bidireccional, es decir que para interpretar un término en contexto se tiene en cuenta tanto la palabra que le antecede como la que tiene a continuación.

Así, por ejemplo, en la oración “accedí a la clave bancaria”, el concepto “clave” es interpretado teniendo en cuenta el fragmento “accedí a” y “bancaria”. A esto se refiere el concepto de interpretación bidireccional.



El buscador procesa miles de millones de búsquedas al año (AFP)

BERT permite poder resolver algunas interpretaciones que, de otro modo, podrían resultar ambiguas. Si, por ejemplo, se escribe en el buscador “2020 turista a Tailandia necesita certificado vacuna” se entenderá con mayor precisión que la búsqueda es de información para saber si es condición tener este documento.

BERT también aprende a modelar las relaciones entre oraciones

mediante el entrenamiento previo en una tarea que se puede generar a partir de cualquier corpus de texto. Así es que dadas dos oraciones A y B, ¿es B la siguiente oración real que viene después de A en el corpus, o simplemente un oración al azar?

Ejemplo:

Oración A: La mujer fue a la tienda de ropa

Oración B: compró una camisa

En este caso, BERT entiende que estas dos oraciones están vinculadas. En cambio, si la oración B decía “los perros tienen cuatro patas” el sistema aprende que se trata de dos frases no vinculadas.

En un principio esta nueva técnica estaba disponible sólo en inglés pero la novedad es que a partir de ahora comenzará a estar disponible (en un roll out progresivo) en otros idiomas: español, alemán, francés, portugués, holandés, danés, árabe, japonés, coreano, ruso, hindi, italiano, vietnamita, ruso y tailandés. A su vez desembarcará en más de 20 países entre los cuales están Australia, Brasil, Canadá, China, Egipto, India, Italia, Japón, Corea, México, Nigeria, Rusia, Arabia Saudita, Suiza, Reino Unido y Vietnam, entre otros. **Es decir que las mejoras en las búsquedas llegan a todos esos sitios, en todas esas lenguas.**

La apertura del código a la comunidad



El buscador tiene en cuenta, a la hora de mostrar los resultados, la relevancia y autoridad de la fuente (Foto: Pixabay)

“Con este lanzamiento, cualquier persona en el mundo puede entrenar su propio sistema de respuesta de preguntas de última generación (o una variedad de otros modelos) en aproximadamente 30 minutos en una sola TPU en la nube, o en unas pocas horas usando una sola GPU”, explicaron desde **Google** en su blog oficial, en septiembre.

El lanzamiento incluyó el código fuente creado sobre **Tensor Flow** y una serie de modelos de representación lingüística previamente entrenados.

“Decidimos abrir el código BERT para que cualquiera pueda utilizar las herramientas que hemos creado. Vemos enormes beneficios para el abastecimiento abierto de nuestra tecnología, y no solo es algo que creemos fundamentalmente, sino que permite que el ecosistema avance a un ritmo mayor, lo

que es positivo para toda la industria. Hemos visto una adopción significativa por parte de los usuarios y estamos muy emocionados de ver cómo la comunidad continúa interactuando con la tecnología”, explicó un vocero de la compañía a **Infobae**.

Esta mejora es particularmente importante para las búsquedas en las cuales anteriormente no se entendía la intención detrás de la consulta que hacía el usuario y por lo tanto, no necesariamente se mostraban los resultados más relevantes.

“Con esta mejora, podemos comprender mucho mejor el lenguaje y proporcionar resultados más útiles, que pueden provenir de cualquier tipo de sitio, incluidos los sitios de noticias”, aclararon desde la empresa.

Si bien la idea venía dando vueltas hace tiempo, recién hace poco se logró utilizar **BERT para pre entrenar una red neuronal**. Esta modificación del algoritmo permite mejorar los resultados que recibe el usuario, logrando así cumplir con el objetivo esencial del buscador. De todos modos, todavía queda mucho por hacer.

“Este trabajo sigue en pie: el año pasado realizamos más de 3.00 actualizaciones de nuestros algoritmos, y continuamos midiendo y mejorando nuestros resultados para asegurarnos de ofrecer una experiencia confiable y de alta calidad”, remarcaron desde **Google**.

Fuente: [Infobae](#)