



Microsoft anunció hoy que abrió una pieza clave de lo que hace que sus servicios de búsqueda en Bing puedan devolver rápidamente los resultados a sus usuarios.

Haciendo que esta tecnología sea abierta, la compañía espera que los desarrolladores puedan crear experiencias similares para sus usuarios en otros dominios donde los usuarios realizan búsquedas en grandes almacenes de datos, o incluso, en el comercio minorista.

La pieza de software de la compañía que pasó a código abierto es hoy en día una biblioteca desarrollada por Microsoft para hacer un mejor uso de todos los datos que recopiló y los modelos de AI que construyó para Bing.

«Hace solo unos años, la búsqueda web era simple. Los usuarios escribieron algunas palabras y leyeron páginas de resultados. Hoy, esos mismos usuarios pueden, en cambio, tomar una foto en un teléfono y colocarla en el cuadro de búsqueda o utilizar un asistente inteligente para hacer una pregunta sin tocar físicamente un dispositivo. También pueden escribir una pregunta y esperar una respuesta real, no una lista de páginas con respuestas probables», dijo la compañía.

Con el Algoritmo de Árbol y Partición del Espacio (SPTAG) que se encuentra en el núcleo de la biblioteca de fuentes abiertas de Python, Microsoft puede buscar miles de millones de datos de información en milisegundos.

La búsqueda de vectores en sí no es una idea nueva. Sin embargo, lo que Microsoft ha hecho es aplicar el concepto al trabajar con modelos de aprendizaje profundo. Primero, el equipo toma un modelo pre-entrenado y codifica esos datos en vectores, donde cada vector representa una palabra o píxel. Utilizando la nueva biblioteca SPTAG, genera un índice vectorial. A medida que entran las consultas, el modelo de aprendizaje profundo convierte ese texto o imagen en un vector y la biblioteca encuentra los vectores más relacionados en ese índice.



Microsoft ofrece el algoritmo crucial de Bing como open source

«Con la búsqueda de Bing, el esfuerzo de vectorización se ha extendido a más de 150 mil millones de datos indexados por el motor de búsqueda para mejorar la comparación tradicional de palabras clave. Estos incluyen palabras sueltas, caracteres, fragmentos de página web, consultas completas y otros medios. Una vez que un usuario busca, Bing puede escanear los vectores indexados y entregar la mejor coincidencia», dijo Microsoft.

La biblioteca ahora está disponible bajo la licencia MIT y proporciona todas las herramientas para construir y buscar estos índices vectoriales distribuidos.