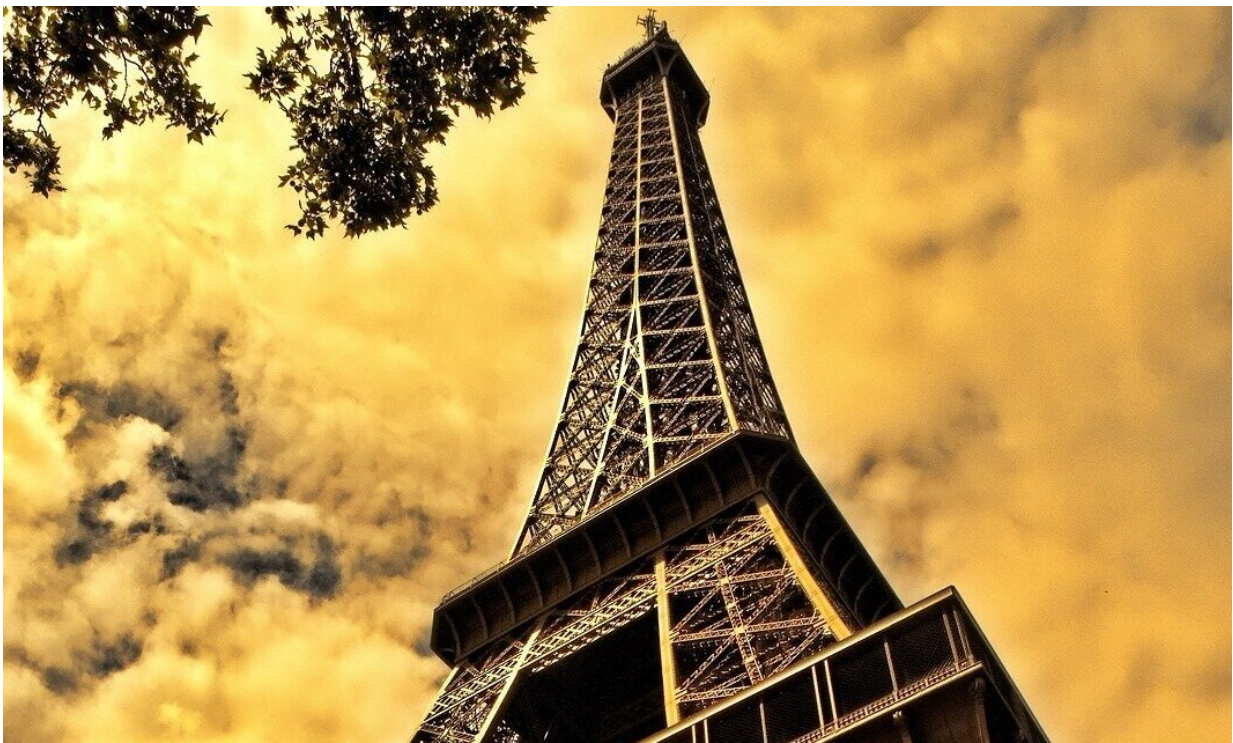


# 'How tall is the tower in Paris?' How vector search knows you're asking about the Eiffel Tower

May 16 2019, by Charlie Waldburger

---



Credit: CC0 Public Domain

Only a few years ago, web search was simple. Users typed a few words and waded through pages of results.

Today, those same users may instead snap a picture on a phone and drop it into a [search box](#) or use an intelligent assistant to ask a question without physically touching a device at all. They may also type a question and expect an actual reply, not a list of pages with likely answers.

These tasks challenge traditional [search](#) engines, which are based around an inverted index system that relies on keyword matches to produce results.

"Keyword search algorithms just fail when people ask a question or take a picture and ask the search engine, 'What is this?'" said Rangan Majumder, group program manager on Microsoft's Bing search and AI team.

Of course, keeping up with users' search preferences isn't new—it's been a struggle since [web search](#)'s inception. But now, it's becoming easier to meet those evolving needs, thanks to advancements in [artificial intelligence](#), including those pioneered by Bing's search team and researchers at Microsoft's Asia research lab.

"The AI is making the products we work with more natural," said Majumder. "Before, people had to think, 'I'm using a computer, so how do I type in my input in a way that won't break the search?'"

Microsoft has made one of the most advanced AI tools it uses to better meet people's evolving search needs available to anyone as an [open source project on GitHub](#). On Wednesday, it also released user example techniques and an accompanying video for those tools via [Microsoft's AI lab](#).

The algorithm, called Space Partition Tree And Graph (SPTAG), allows users to take advantage of the intelligence from deep learning models to

search through billions of pieces of information, called vectors, in milliseconds. That, in turn, means they can more quickly deliver more relevant results to users.

Vector search makes it easier to search by concept rather than keyword. For example, if a user types in "[How tall is the tower in Paris?](#)" Bing can return a natural language result telling the user the Eiffel Tower is 1,063 feet, even though the word "Eiffel" never appeared in the search query and the word "tall" never appears in the result..

Microsoft uses [vector](#) search for its own Bing search engine, and the technology is helping Bing better understand the intent behind billions of web searches and find the most relevant result among billions of web pages.

## Using vectors for better search

Essentially a numerical representation of a word, image pixel or other data point, a vector helps capture what a piece of data actually means. Thanks to advances in a branch of AI called deep learning, Microsoft said it can begin to understand and represent search intent using these vectors.

Once the numerical point has been assigned to a piece of data, vectors can be arranged, or mapped, with close numbers placed in proximity to one another to represent similarity. These proximal results get displayed to users, improving search outcomes.

The technology behind the vector search Bing uses got its start when company engineers began [noticing unusual trends in users' search patterns](#).

"In analyzing our logs, the team found that search queries were getting

longer and longer," said Majumder. This suggested that users were asking more questions, over-explaining because of past, poor experiences with keyword search, or were "trying to act like computers" when describing abstract things—all unnatural and inconvenient for users.

With Bing search, the vectorizing effort has extended to over 150 billion pieces of data indexed by the search engine to bring improvement over traditional keyword matching. These include single words, characters, web page snippets, full queries and other media. Once a user searches, Bing can scan the indexed vectors and deliver the best match.

Vector assignment is also trained using deep learning technology for ongoing improvement. The models consider inputs like end-user clicks after a search to get better at understanding the meaning of that search.

While the idea of vectorizing media and search data isn't new, it's only recently been possible to use it on the scale of a massive [search engine](#) such as Bing, Microsoft experts said.

"Bing processes billions of documents every day, and the idea now is that we can represent these entries as vectors and search through this giant index of 100 billion-plus vectors to find the most related results in 5 milliseconds," said Jeffrey Zhu, program manager on Microsoft's Bing team.

To put that in perspective, Majumder said, consider this: A stack of 150 billion business cards would stretch from here to the moon. Within a blink of an eye, Bing's search using SPTAG can find 10 different business cards one after another within that stack of cards.

## **Uses for visual, audio search**

The Bing team said they expect the open source offering could be used for enterprise or consumer-facing applications to identify a language being spoken based on an audio snippet, or for image-heavy services such as an app that lets people take pictures of flowers and identify what type of flower it is. For those types of applications, a slow or irrelevant search experience is frustrating.

"Even a couple seconds for a search can make an app unusable," noted Majumder.

The team also is hoping that researchers and academics will use it to explore other areas of search breakthroughs.

"We've only started to explore what's really possible around vector search at this depth," he said.

Provided by Microsoft

Citation: 'How tall is the tower in Paris?' How vector search knows you're asking about the Eiffel Tower (2019, May 16) retrieved 2 May 2024 from <https://phys.org/news/2019-05-tall-tower-paris-vector-youre.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.