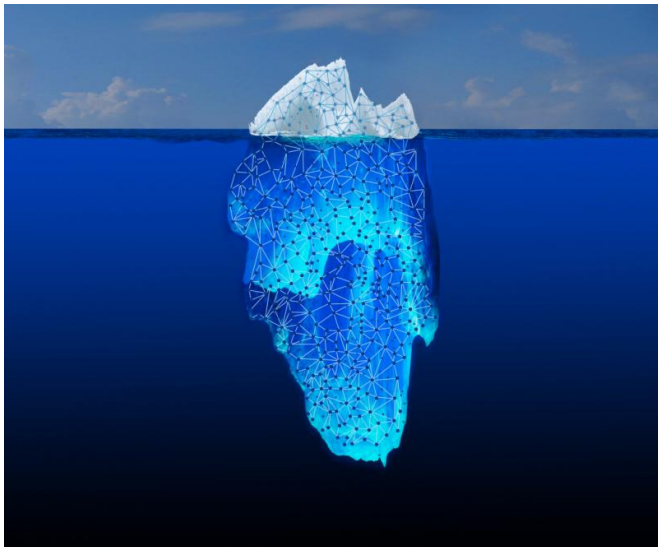


'Deep web search' may help scientists

25 May 2015, by Elizabeth Landau



What you see when you do a basic Web search is only the tip of the iceberg. Most of the information is buried in the "Deep Web." JPL is collaborating on a DARPA initiative called Memex, which explores the connections between bits of information hidden in this vast ocean of content. Image credit: Credit: NASA/JPL-Caltech

When you do a simple Web search on a topic, the results that pop up aren't the whole story. The Internet contains a vast trove of information—sometimes called the "Deep Web"—that isn't indexed by search engines: information that would be useful for tracking criminals, terrorist activities, sex trafficking and the spread of diseases. Scientists could also use it to search for images and data from spacecraft.

The Defense Advanced Research Projects Agency (DARPA) has been developing tools as part of its Memex program that access and catalog this mysterious online world. Researchers at NASA's Jet Propulsion Laboratory in Pasadena, California, have joined the Memex effort to harness the benefits of deep Web searching for science. Memex could, for example, help catalog the vast amounts of data NASA spacecraft deliver on a

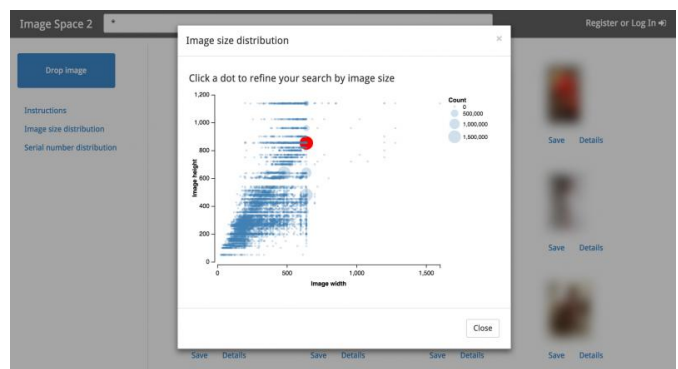
daily basis.

"We're developing next-generation [search](#) technologies that understand people, places, things and the connections between them," said Chris Mattmann, principal investigator for JPL's work on Memex.

Memex checks not just standard text-based content online but also images, videos, pop-up ads, forms, scripts and other ways information is stored to look at how they are interrelated.

"We're augmenting Web crawlers to behave like browsers—in other words, executing scripts and reading ads in ways that you would when you usually go online. This information is normally not catalogued by search engines," Mattmann said.

Additionally, a standard Web search doesn't get much information from images and videos, but Memex can recognize what's in this content and pair it with searches on the same subjects. The search tool could identify the same object across many frames of a video or even different videos.



This screen shot from the Image Space and ImageCat applications shows how the search tool can organize a large amount of information and look for relationships between data hidden in images. Credit: NASA/JPL-Caltech/Kitware/Continuum

The video and image search capabilities of Memex could one day benefit space missions that take photos, videos and other kinds of imaging data with instruments such as spectrometers. Searching visual information about a particular planetary body could greatly facilitate the work of scientists in analyzing geological features. Scientists analyzing imaging data from Earth-based missions that monitor phenomena such as snowfall and soil moisture could similarly benefit.

Memex would also enhance the search for published scientific data, so that scientists can be better aware of what has been released and analyzed on their topics. The technology could be applied to large NASA data centers such as the Physical Oceanography Distributed Active Archive Center, which makes NASA's ocean and climate data accessible and meaningful. Memex would make PDF documents more easily searchable and allow users to more easily arrive at the information they seek. Awareness of existing publications also helps program managers to assess the impact of spacecraft data.

All of the code written for Memex is open-source. JPL is one of 17 teams working on it as part of the DARPA initiative.

Memex is related to DARPA's previous Big Data initiative called XDATA, managed by DARPA Program Manager Wade Shen. That research effort is also aimed at processing and analyzing large amounts of data, with defense, government and civilian applications. JPL was one of 24 groups involved.

"We are developing open source, free, mature products and then enhancing them using DARPA investment and easily transitioning them via our roles to the scientific community," Mattmann said.

Provided by NASA

APA citation: 'Deep web search' may help scientists (2015, May 25) retrieved 20 November 2019 from <https://phys.org/news/2015-05-deep-web-scientists.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.