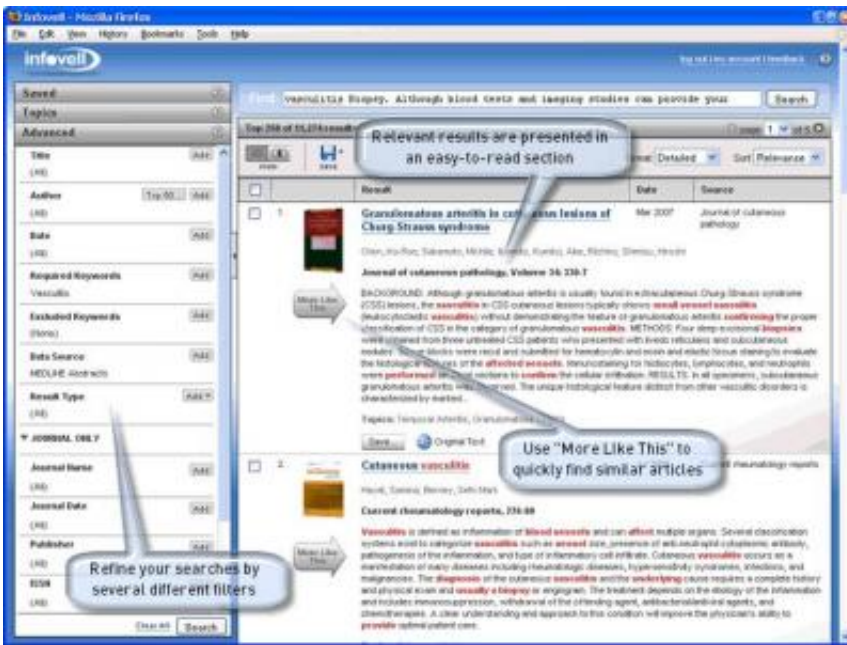


Infovell's 'research engine' finds deep Web pages that Google, Yahoo miss

September 8 2008, by Lisa Zyga



With Infovell, users search with key phrases up to 25,000 words long, rather than keywords. Image credit: Infovell.

According to a study by the University of California at Berkeley, traditional search engines such as Google and Yahoo index only about 0.2% of the Internet. The remaining 99.8%, known as the "deep Web," is a vast body of public and subscription-based information that traditional search engines can't access.

To dig into this "invisible" information, scientists have developed a new

search engine called Infovell geared at helping researchers find often obscure data in the deep Web. As scientists working on the Human Genome Project, Infovell's founders designed the new searching technology based on methods in genomics research. Instead of using keywords, Infovell accepts much longer search terms, and in any language.

"There are no 'keywords' in genetics," explains Infovell's Web site. "New unique and powerful techniques have been developed to extract knowledge from genes. Now, through Infovell, these techniques have, for the first time, been applied to language and other symbol systems, shattering long-held barriers in search and leapfrogging the capabilities of current search providers to deliver the World's Research Engine."

While keywords may work fine for the general public looking for popular and accessible content, they don't often meet the needs of researchers looking for specific data. As information in the deep web continues to grow, Infovell explains that a one-size-fits-all approach to searching will make academic searching even more challenging.

One reason is the nature of deep Web sites themselves. While many popular Web sites are specifically designed to be search-engine friendly, a lot of deep Web content is unstructured, making it difficult for keyword-based search engines to index. Further, the deep Web does not receive much traffic, meaning these pages don't have many incoming links and therefore aren't ranked highly by systems such as Google's PageRank. And for private sites, barriers such as registration and subscription requirements also make it difficult for search engines to access them.

Searching with keywords also presents a trade-off between being too general and getting millions of irrelevant results, or being too specific and not getting any results at all. After getting results, users then have to

sift through many pages looking for what they need.

But with Infovell, users search with "KeyPhrases," from paragraphs to whole documents or even sets of documents up to 25,000 words. Because it's born out of the world of genomics, Infovell is also language-independent. Users can search in English, Chinese, Arabic, or even mathematical symbols, chemical formulas, or musical notes. "The key requirement is that the information is in digital format, and it can be stored in a linear, sequential and segregated manner," according to Infovell's site.

Infovell's technology allows users to locate the most current and comprehensive documents and published articles from billions of pages, with topics including life sciences, medicine, patents, industry news, and other reference content.

Currently, some researchers use advanced search options provided by individual sites to try to get around keyword search engines. However, these search engines require users to learn special syntax, and only work for the site they're at. The advantage of Infovell is that it doesn't require special training (and it doesn't use Boolean operators, taxonomies or clustering); rather, it is easy to use and can search everything at once.

Although Infovell is not the first attempt at a search engine for crawling the deep Web, its developers hope that researchers will benefit from Infovell's advantages more in the future, especially as the deep Web continues to grow.

Infovell is being demonstrated at DEMOfall08, a conference for emerging technologies taking place in San Diego on September 7-9. Users can sign up for a 30-day risk-free trial at Infovell's Web site, and Infovell is initially available on a subscription basis. Later this year, Infovell will release a free beta version on a limited basis without some

of the advanced features in the premium version.

More information: www.infovell.com

Via: www.networkworld.com

Citation: Infovell's 'research engine' finds deep Web pages that Google, Yahoo miss (2008, September 8) retrieved 19 July 2024 from <https://phys.org/news/2008-09-infovell-deep-web-pages-google.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.