

free newsfeeds | open forums | free newsletters | access your settings



INNOVATION

Explore: NewsFactor Sites by Topic

"THE INTERNET'S 24-HOUR E-BUSINESS & TECHNOLOGY NEWS NETWORK"

= Section Updated Today

- Front Page
- Special Reports
- E-Business
- Enterprise
- Technology
- Trends
- Opinion
- Cybercrime
- Innovation
- Macintosh
- Worldwide Tech
- Sci-Tech

Bringing a Much Bigger Internet to Light

By Brian McDonough
[NewsFactor Sci::Tech](#),
 Part of the NewsFactor Network
 July 17, 2002

- [Send this Article](#)
- [Talkback](#)
- [Print this Article](#)
- [Related Stories](#)

What may prove valuable about deep-Web mining is the ability to interpret individual pieces of data that might not otherwise be of much use.

[COMPLETE STORY](#)

advertisement

FREE WHITE PAPER: eService Best Practices. In a tough economy, you can't afford to replace customers so your service needs to be excellent. But you must also control costs. eService lets you do both. [Get this paper](#) to learn about eService.

Deep mining of online data could transform the way information is collected and analyzed. It promises companies an easier and more effective way to keep up with their rivals and manage their brands, and it will help researchers develop a deeper understanding of social and economic trends. It also might alter the way the average citizen is targeted and analyzed, bringing more personal information into the glare of the spotlight.

"Deep Web" is a term coined by [BrightPlanet.com](#), a company that offers corporate and institutional clients a way to get **at more and better information than any Web browser can deliver.**

The problem with Web browsers is that they can catalog only those Internet sites that have static, unchanging URLs. However, a lot of data available online is in databases that only generate a temporary Web page in answer to a specific query. The deep Web is a 500-times-bigger data repository than the area of the Internet that is targeted by typical search-engine queries, BrightPlanet president Thane Paulsen told NewsFactor.

In the Deep End

The deep Web includes phone directories and "a lot of government data that's available online," said Joseph Hellerstein, a computer science professor at the University of California at Berkeley who has mined the deep Web on his own in a number of research projects.

A [Google](#) search will not go into every database on the Web and enter search terms into appropriate fields to retrieve relevant data. BrightPlanet's Deep Query Manager will.

ADVERTISEMENT

TODAY
 July 21, 2002
SEARCH

 Special: Tech Innovation
 A Week on This Site
 A Week on NewsFactor
 Most Popular

- NEW!** Are Hacking Defenses Winning the War?
- NEW!** New 'Instant Ethernet' Powerline Device Unveiled
- NEW!** Ask Jeeves Ousts Overture, Goes Google in \$100M Deal
- NEW!** Gateway Gets Aggressive with Pricing
- NEW!** Microsoft Reports Whopping \$1.53 Billion Profit
- NEW!** AOL's Pittman Steps Down Amid Ad Sales Revelations
- NEW!** Wireless Tech Stocks To Watch
- NEW!** Will Alternative Web Browsers Ever Unseat Microsoft?
- NEW!** Yahoo! Security Measure Alters E-Mail Text
- NEW!** Fujitsu Unveils New Disk Drives, Server Processors
- NEW!** Apple's .Mac Challenges Microsoft's .NET
- NEW!** Study: Web Security Spending To Surge
- NEW!** Shutting Microsoft Out of the Wireless World
- NEW!** Jobs: Apple's 'Jaguar' Set To Pounce

"The trend is toward more content going into databases rather than onto the surface Web," said Paulsen. The

"surface Web" is what the company calls the millions of static-URL sites that will show up in response to a standard search-engine query.

Adds CEO Bill Shelander, "Content is exploding on the Internet in every direction, but the biggest [growth] is in databases."

Bringing Data Home

So, just as the Nexis-Lexis service allows users to scour a wide variety of publications and news sites at once, BrightPlanet offers a tool that allows users to automatically search the shallow Web -- it even uses popular search engines -- while also getting into relevant databases that offer public access, but would be too time-consuming for a user to visit individually ... assuming the user could even find them.

To date, the customers who subscribe to the company's ASP-model services -- for about US\$120 per month, per seat -- include corporations, government agencies and research institutions.

Giving Data a Purpose


One principal use of deep-Web mining, Paulsen and Shelander say, has been in brand management and gathering competitive information. A car manufacturer, for example, might want to aggregate and retrieve all the information available about its luxury sedan. It might also want all the data it could get about its competitor's luxury sedans.

Another use: Comprehensive, deep research. Corporate research and development departments need a lot of information in planning products that have multiyear development cycles. Industry analysts require the best information they can get to put together complete pictures of their fields.

Medical researchers are another group Paulsen said has found a need for this kind of deep research. Say a researcher is looking into "chronic daily headache syndrome." Deep research of various medical databases, in addition to a well-filtered sweep of the surface Web, might yield hundreds of highly valuable search results. A search of a regular search engine, on the other hand, could produce thousands of links -- many of them useless to a researcher.

How You Look at It

What may prove really valuable about deep-Web mining is the ability to interpret various individual pieces of data that might not otherwise be of much use. At UC Berkeley, Hellerstein has worked at mining the deep Web in collaboration with social science researchers trying to find new ways to draw answers from the increasing collections of disparate data available online.

"The first demo we did for presentation was during the last presidential campaign, when the big story was campaign finance," Hellerstein told NewsFactor. "We looked at data from the Federal Election Commission on donors to Bush and Gore, and correlated it with data on home purchases from Yahoo! (Nasdaq: YHOO) , and crime data and other demographics."

The researchers were able to take the donors' addresses and get a picture of who was giving money to each campaign. They found -- true, it is not much of a surprise -- that Bush was getting most of his money from people in very affluent, low-crime neighborhoods, while Gore was drawing funds from a wider range of donors.

Practical and Personal

Currently, Hellerstein is working with Hal Varian, dean of UC Berkeley's School of Information Management and Systems, on research into worker migratory patterns.

"We're always trying to look for leading indicators or forecasts of different economic variables," said Varian. The question at hand is how long someone will look for a job in a region that has suffered an economic hit before leaving the area.

Taking Texas in the wake of the Enron scandal as a model case, Varian, Hellerstein and others plan to create a picture of how workers weather economic storms. They intend to scour online resume postings and job-oriented databases, as well as standard economic data and a range of diverse factors -- freeway traffic patterns, for example.

So, corporations and industry analysts can study their markets. Researchers can try to better

Purchasers



Full Story

NEW! When Will Broadband Rule the Market?

▷ [Full Story](#)

NEW! U.S.: States Might Be Hindering E-Commerce

▷ [Full Story](#)

NEW! iPod for Windows: Apple's Trojan Horse

▷ [Full Story](#)

NEW! Intel Misses Estimates, Plans 4,000 Job Cuts

▷ [Full Story](#)

NEW! Will XML Kill HTML?

▷ [Full Story](#)

NEW! Thursday's CyberCrime and Security Report

▷ [Full Story](#)

NEW! U.S., Business Agree on Internet Security Standards

▷ [Full Story](#)

NEW! Microsoft Changes MSN Browser Strategy

▷ [Full Story](#)

NEW! AOL Time Warner: Who's the Boss?

▷ [Full Story](#)

NEW! Florida Politician Guilty of Internet Auction Scam

▷ [Full Story](#)

NEW! The Changing Face of Online Stock Trading

▷ [Full Story](#)

NEW! Apple Results Lackluster But Profitable

▷ [Full Story](#)

NEW! Amazon Jumps on Web Services Bandwagon


▷ [Full Story](#)

OPTIONS

▷ [See more news](#)

▷ [Get news by e-mail](#)

▷ [Visit open forums](#)

understand how aspects of how society works. All good news. But according to some, there is a downside: Better mining of the data hidden below the surface of the Web means better ways to uncover more personal information  about an individual.

Stay Tuned for Part 2: Plumbing the Personal Data Depths



Talkback: [Click here to add your comment about this story...](#)

See Related Stories

[IBM Data Mining Research Tackles Privacy Dilemma](#)

(31-May-02)

[Journey to the Internet's Unknown Regions](#)

(24-Apr-02)

[Text-Mining Technology Sorts Data the Smart Way](#)

(13-Mar-02)

[Interface Innovation: The Future of Information Access](#)

(19-Dec-01)

[Feds To Draw 'Map' of Internet](#)

(07-Dec-01)

Sponsored Links

- ❖ Free Guide: Bulletproof Your Network. [Click here.](#)
- ❖ [Click here for FREE WHITE PAPER: eService Best Practices.](#)
- ❖ [Click here for a FREE Infrastructure development white paper.](#)
- ❖ [Click to improve employee efficiency, increase customer satisfaction.](#)
- ❖ [Reach Thousands of Internet Pros Everyday with NewsFactor Newsletters!](#)
- ❖ [If you know anything at all about I.T. you better see this!](#)
- ❖ [Free Windows and Linux software downloads. Click here.](#)

NEW! Software Updates brought to you in partnership with [DaveCentral](#), part of OSDN - the Open Source Development Network.

Windows

RECENT UPDATES



1. [GuildFTPd 0.999.5](#) in [FTP](#) - [Servers](#)
2. [AdvaCrypt Suite 2 2.59ds](#) in [File Utilities](#) - [Encryption Tools](#)
3. [Enhanced HTML 2002 4.4.55](#) in [Web Authoring](#) - [Text Editors](#)
4. [SheerPower 4GL 3.0](#) in [Database](#) - [Programming](#)
5. [602Pro PC SUITE 2001](#) in [Office Applications](#) - [Suites](#)

Also, check out today's shareware review of [Hacktrap](#) at DaveCentral.

Windows

TOP 5 PROJECTS



1. [MIDI Made Music Jukebox](#) (215,751 hits) in [Audio](#) - [Players](#)
2. [3D Custom Screensaver](#) (122,403 hits) in [System Utilities](#) - [Misc. Utilities](#)
3. [Zero Popup](#) (115,821 hits) in [Web Surfing](#) - [Browser Utilities](#)
4. [Ultimate Boot Disk](#) (104,970 hits) in [System Utilities](#) - [Misc. Utilities](#)
5. [Auto Mouse](#) (80,876 hits) in [Web Surfing](#) - [Browser Utilities](#)

NewsFactor.com

[Front Page](#) | [Special Reports](#) | [Worldwide Tech](#) | [E-Business](#) | [Monster Deals](#) | [Tech Stocks](#) | [Technology Trends](#) | [Opinion](#) | [CyberCrime](#) | [Culture](#) | [Editorial Corrections](#)



Other NewsFactor Network Sites

[NewsFactor Portal](#) | [E-Commerce Times](#) | [TechNewsWorld](#) | [Linux Insider](#) | [Wireless NewsFactor](#)
[osOpinion](#) | [TechExtreme](#) | [CRMDaily](#) | [CRM Buyer Magazine](#)

[FreeNewsFeed](#) | [Free Newsletters](#)

[Business Development](#) | [How To Contact Us](#) | [About NewsFactor Network](#)
[How To Advertise](#) | [Article Reprint Information](#)



© 1998-2002 Triad Commerce Group, LLC. All rights reserved. See [Terms of Use](#) and [Privacy](#) notice.

