

Hocus Pocus: (Un)veiling the (In)visible Web

Over the past three years, the Invisible Web was a topic rich and intriguing—especially at the local Internet bar where the talk tried to climb above the heads of the average human and into an almost SearchTrekian space where no one else would dare to go. Early adventurers such as Michael Bergman, Chris Sherman, and Gary Price originally charted the Deep Web as about 500 billion individual documents. But in those early days, PDF's and other formats were not retrieved by the convention search tools such as Google, Yahoo, AltaVista, All the Web or metasearchers like Ixquick, Vivísimo, and SurfWax. While those search tools often accessed more a couple billion pages in their databases, a large portion of available information was and still is difficult or impossible to search.

The reality is many information specialists as well as the general public use the Invisible Web already. Most Web surfers have accessed an Invisible Web site at one time or another. However, they use only a portion--typically found in three general forms:

- 1) The Fee Group - paid databases such as EBSCO, Wall Street Journal, OVID, ProQuest, Medline, etc. These are databases that have a cost associated with use.
- 2) The Free Group - Government databases such as the Census, AskERIC, PubMed, the Currency Converter, FindArticles, library online catalogs, etc. These databases are free for anyone to access.
- 3) The Hybrids - UnCoverWeb, online newspapers like the New York Times, Wall Street Journal, etc. currently take this form. These databases have free portions and fee portions.

What is the Invisible Web?

- Content on Internet that is not directly indexed by conventional search tools
- Data primarily found in databases
- Requires a direct query to retrieve information
- Free, Fee, and Hybrid (part free) Databases
- Other terms: Deep Web, Opaque Web, Special Databases, Searchable Databases

If it is Invisible, Why Bother?

- Generally better quality information
- More specific information
- Finding information is faster and more efficient
- Information that is more relevant to evaluating resources

The Not So (In)visible, Deep Web

While the (In)visible Web has been the next big thing recently, learning to apply the principle of library and information science to search the Internet will ultimately prevail upon algorithms, placing, ranking, and other principles of computer science. Now PDFs, Microsoft Word, Excel, and PowerPoint files, searchable databases, and other formatted information are becoming available on major search tools—especially Google. Other information such as a book from my local library catalog is just a click or two deeper. True, I may not be able to find out how much \$179 dollars is worth in the currency of Egypt, but I can easily find a currency converter. So it is a matter of knowing what you are looking for and continuing to be versatile and persistent. Using Google, the Invisible Web, or any other search tool will not miraculously transform Internet searching into anything other than what it is—a simple search. Applying principle of systematic research such as horizontal searching across databases, catalogs, and search tools will bring Internet research into the 21st century as what it really is—an art.

Strategies for finding information on the Invisible Web

- Move beyond keyword searching on Google
- Read the Search tips or Help menus on search tools to discover the possibilities
- Try alternative strategies for searching using Invisible Web sites when appropriate:
 - Invisible-web.net - <http://www.invisible-web.net/>
 - The Invisible Web - <http://www.invisibleweb.com/>
 - CompletePlanet - <http://www.completeplanet.com/>
 - Direct Search - <http://www.freepint.com/gary/direct.htm>
 - Subject Directories such as:
 - Infomine - <http://infomine.ucr.edu/>
 - Academic Info - <http://www.academicinfo.net/index.html>
 - Best Information on the Net - <http://library.sau.edu/bestinfo/>
 - Internet Public Library - <http://www.ipl.org/>
 - Librarians' Index to the Internet - <http://lii.org/>
 - Scout Report Archives - <http://scout.wisc.edu/archives/>
- Think of the Invisible Web as part of the whole instead of separate
- Apply concepts of Library and Information Science not Computer Science
- Horizontal Searching—incorporate the Web as integral to a search strategy
 - Search the Web titles of articles
 - Look for bibliographies that can then be incorporated into searches for books and journal articles
 - Search for authors from books and journals
 - Search for organizations and government reports
 - Follow citations onto the Web

Resources for Staying Current on the Web

Search Engine Watch - <http://www.searchenginewatch.com/>

Free Pint - <http://www.freepint.co.uk/>

The Scout Report - <http://scout.cs.wisc.edu/report/sr/current/index.html>

ResearchBuzz - <http://www.researchbuzz.com/>

FOCUS on the BEST on the NET - <http://www.focusbest.net/join.html>

Netsurfer Digest - <http://www.netsurf.com/nsd/subscribe.html>

Byte.com Newsletter - <http://www.byte.com/newsletter>

The contents of this presentation are posted at the following site:

<http://home.sou.edu/~vidmar/onlinew2003>

Biography:

Dale Vidmar is an associate professor and the Instruction Librarian at Southern Oregon University Library. He teaches a graduate Internet Research and Web Design class and a host of other classes on searching the Internet and research in general. He maintains an Internet Searching Tools Web site at: [www.sou.edu/library/searchtools]. He has published articles and given several presentations on Internet searching because it has become as much an obsession as a passion in his profession. A portion of every day trying to learn something new about the Web. To his satisfaction, every day generally offers something that he did not know yesterday.

vidmar@sou.edu

<http://www.sou.edu/library/dale>