



Power Searching: The Hidden Web

Information Sheet

Prepared by Rita Gavelis

LibraryAssist.com

What is the Hidden Web?

The Hidden Web, also commonly known as the invisible or deep web, are pages that have not been indexed by search engines. According to Michael K. Bergman in his paper: "The Deep Web: Surfacing Hidden Value."¹, the content in the deep web is over 500 times greater than what is visible to conventional search engines. He compares searching on the Internet to dragging a net across the surface of the ocean: A great deal of information will be caught in the net, leaving a wealth of information deep below the surface.

Below is a list of the different reasons why a page might not appear in a traditional search engine, thus making it invisible:

- The page does not have many links pointing to it, so it is considered unimportant by a search engine.
- The page is an orphan, with no links leading to it, so it cannot be found by a search engine's web crawler.
- The page has a temporary URL, meaning the content of that page is created as a result of a database query. That URL disappears when you enter a new database search. The only way to retrieve that same page is to enter the same exact query in the database. So most databases are inaccessible to traditional search engines.
- The site is password protected and therefore inaccessible to a web crawler.
- Audio and video content is difficult to index, so many files with pertinent information are missed.
- Some websites place html code in their site instructing web crawlers not to index their entire site or only specific pages from their site.
- The page might consist of a file format that has not been programmed into the search engine's algorithm.

¹ Bergman, Michael K. "The Deep Web: Surfacing Hidden Value." Brightplanet.com. 24 Sept. 2001. Web. 7 Apr. 2013. <<http://brightplanet.com/wp-content/uploads/2012/03/12550176481-deepwebwhitepaper1.pdf>>.

What are Databases?

Databases are websites or online services that have compiled articles, data, images, or other files and organized them. They offer built-in search features allowing you to find the information while you are in the database but prevent outside search engines such as Google and Bing access to that content.

Some databases are free to use, such as the Library of Congress site or must be subscribed to such as Ebsco's Novelist database.

Fee vs. Free

There are thousands of free databases available through government, educational, and non-profit organizations that have gone un-indexed by the popular search engines. They take a great deal of time to find. Once you do come across one, be sure to book mark it in your browser or social bookmarking services such as Delicious, StumbleUpon, or Pinterest.

Below are some tips to help determine when to use a subscription database vs. those found in the Hidden Web.

When to Pay:

- The information is needed immediately and you don't have time to search a directory or hidden web.
- You need a specific article from a journal that is only available for a fee.
- The information needed must be from a peer reviewed journal.
- Some paid databases have features enabling users to save searches, create citations, and organize articles into folders.

When to Probe the Deep Web

- You do not have access to a paid database.
- The information is in a format that cannot be accessed via a subscription: charts, images, audio, video are not part of the database package.
- The information is so specific, it can only be found within certain sites that must be searched independently.
- The information needed was created or compiled by a government agency or service: Census Bureau, Government Publication Office, U.S. Geological Survey, National Institute of Health.
- You can find the same information in a subject directory.
- The information is student-created: search the college web site.
- The information required is specialized or obscure.