



MicroModule: Invisible

REVIEW Page

Below is the entire module on one page.



There are parts of the Internet that search engines simply can't *see*. The search engine's robotic "crawlers" either miss or are locked out of these areas on the Internet. Behind the barriers lie treasure troves of quality information. Collectively this information is called the Invisible Web. There are many categories of *invisible* information missed by the popular search engines. The most common are:

- webpages that have been intentionally skipped by search engine crawlers
- webpages that are dynamically assembled from online database content
- password protected webpages
- non-HTML resources like image, audio, animation, and PDF files

Intentionally skipped: Webpages that have been missed or skipped by the crawlers are sometimes called the nearly visible or opaque web. These pages could be indexed but are skipped intentionally to save the search company time and money. Because crawling and indexing is expensive, search engines limit the number of pages they copy from each site. This can leave hundreds of pages out of the search index, but still available to the site's users. (For more on this topic see the IMSA Module: Opaque Web.)

Dynamic Material created on demand: Online databases create HTML pages to match your criteria. These pages are *dynamically* assembled when you query the database. Search engines do not index the contents of online databases, and they cannot index dynamically assembled pages that don't exist until the user creates them. Searching for a book on Amazon.com is one example of using the Invisible Web. Amazon's database will assemble a unique page to match your requests.

Password protected information: Many sites have password-protected webpages. Search engine spiders can reach the front door, but can't crawl in. On the other side of the barrier is quality information developed and categorized by professionals. Before you search these pages you must first establish an account. Some sites are free, others charge a fee. Regardless the materials beyond the password barrier can't be reached by search engines and remains invisible until you establish an account, obtain the key, and login to the website.

Non-HTML formats: Search engines were originally designed to comb through HTML text pages and create an index of keywords. Non-HTML file formats that didn't contain much text were routinely skipped. These file formats include image, audio, animation, and PDF files. Recently some of the commercial search engines have added image and PDF files to their indexes. Additionally specialized search engines are available to help you find these and other types of files. (For more on this topic see the IMSA Module: File Formats.)

Vast Resources await: The Invisible Web is a vast resource estimated to be from 2 to 500 times the size of the easily accessible information on the public web. The materials found on the Invisible Web are often more focused, current, and professionally relevant than what you can find on the public web using search engines. Knowing how to use Invisible Web resources will make you a more efficient and powerful researcher. (For more on this topic see the IMSA Module: How Many Pages Are There On The WWW?)

FAQs



What is the Invisible Web?


There are parts of the Internet that search engines simply can't see. The search engine's robotic crawlers are locked out of these areas on the Internet. Behind the barriers lie treasure troves of quality information. Collectively this information is called the Invisible Web. Luckily, you can use search engines to find the front door of many Invisible Web Resources. Think of it as a two step process. First you locate an Invisible Web Database. Second you go directly to that database and begin investigating the contents.

Why would you want to search the Invisible Web?

The quality of resources on the Invisible Web is often better than what you will find on the public web. Professionals with specific subject matter expertise maintain many database collections. Materials on the fee-accessed parts of the Invisible Web are likely to be more current, and professionally prepared. Also the Invisible Web is thought to be from 2 to 500 times larger than the public web. If you don't want to miss what might be the very best sources of information on your topic, use the Invisible Web.

Are most Invisible Web resources free?

Yes. The majority of Invisible Web resources are free for the finding. Many web sites only require that you set up a free account to access their materials. However, premium information often costs money. Consider the value of the information, your time, and needs. Often paying for access is a good choice!



21st Century Information Fluency Project
IMS A Illinois Mathematics and Science Academy

[contact us](#) | [site map](#) | [IMS A home](#)

Create your own free account to access invisible web content!

You are not logged in [Login](#) [Join](#)

Home

[Wizard Tools](#)

[Tutorials](#)

[Resources](#)

[Events](#)

[About Us](#)

[Site Map](#)

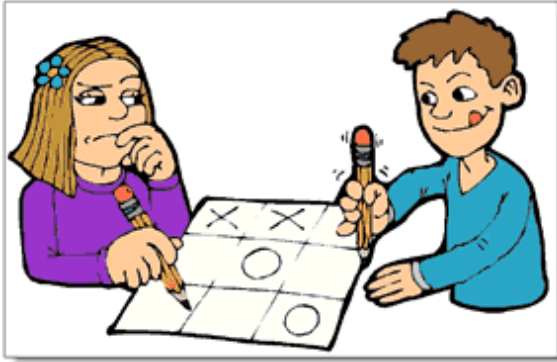
Registration Form

Personal Details

Full Name
Enter full name, eg. John Smith.

User Name ■
Enter a user name, usually something like 'jsmith'. No spaces or special characters. Usernames and passwords are case sensitive, make sure the caps lock key is not enabled. This is the name used to log in.

E-mail ■
Enter an email address. This is necessary in case the password is lost. We respect your privacy, and will not give the address away to any third parties or expose it anywhere.



What search strategies should I use when approaching the Invisible web?

Searching the Invisible Web for information is a two-step process. First you search for the right online resources likely to hold the desired information. Next you search the site itself. Without a search strategy, the size of the Invisible Web creates information overload. To search the web you need a plan that helps you quickly find the most likely sources of quality information. If you develop a personal search strategy that is both flexible and focused, you will become a more efficient researcher

Strategies to Consider: If you know a site that has the kind of information you are looking for, go directly to the source. If you were looking for the population of a US city, for example, instead of using a traditional search engine, search the database available at <http://www.census.gov>.

When searching for information on the Invisible Web, use a search engine to search by subject to find an Invisible Web database that might contain the information.

Use websites that list Invisible Web collections and provide easy access to search forms.

- When visiting a site with Invisible Web resources use the site's search tools to refine your search to the specific information you are looking for.

How can I use a search engine to find Invisible Web databases?

You can use your favorite search engine to find Invisible Web resources; just add keywords like database, archive, or repository to your query. Once you find a site with database resources, go directly to it and use the site's own tools to investigate further.

Let's experiment by combining the Boolean AND command with these keywords: database, repository, archive.

Examples:

- "lesson plans" AND database
- "invisible web" AND repository
- "university library" AND archive

Google Web Images Groups News Froogle more »

"invisible web" AND repository Search [Advanced Search](#) [Preferences](#)

The "AND" operator is unnecessary -- we include all search terms by default. [\[details\]](#)

Web Results 1 - 100 of about 967 for "invisible web" AND repository. (0.57 seconds)

The UIUC Web Integration Repository
... IWRandom: This dataset contains 33 original query interfaces randomly sampled from 16 top level domains from **Invisible-Web.net**. ... Using This **Repository**. ...
metaquerier.cs.uiuc.edu/repository/ - 6k - [Cached](#) - [Similar pages](#)

[Randomly Sampled Query Interface from Invisible-Web.net](#)
... ssa, Interface Page. ukonline, Interface Page. unesco, Interface Page. wings, Interface Page. Randomly Sampled Query Interface from **Invisible-Web.net**.
metaquerier.cs.uiuc.edu/repository/datasets/iwrandom/browsable.html - 6k - [Cached](#) - [Similar pages](#)
[[More results from metaquerier.cs.uiuc.edu](#)]

[Those Dark Hiding Places: The Invisible Web Revealed](#)
Those Dark Hiding Places: The **Invisible Web** Revealed. Robert J. Lackie, Assistant ... a collection of documents and becoming a multidimensional **repository** for sounds ...
library.rider.edu/scholarly/rlackie/Invisible/Inv_Web_Main.html - 25k - Jun 20, 2004 - [Cached](#) - [Similar pages](#)

[DeepWeb.com - Your source for deep web information](#)
... Also known as the '**invisible web**', the 'deep web' is a vast **repository** of web pages, usually generated by database-driven websites, that are available to web ...

How can I use advanced search syntax to find Invisible Web resources?

Most search engines provide advanced search syntax that will help you find Invisible Web resources. By *field searching* for Invisible Web related keywords in the **URL**, **title**, or within likely **domains**, you can uncover vast numbers of Invisible Web resources.

If you know specialized search terms you can construct advanced queries that target your search. Most search engines provide advanced feature forms that will help you construct an advanced query. Think of this as detective work that will help make the 'invisible' accessible!

Advanced Search forms allow you to combine domain restriction with specialized field search terms like:

- **inurl:**
- **intitle:**

Advanced Web Search

You can use the options on this page to create a very specific search. Just fill in the fields you need for your current search.

Yahoo! Search

Show results with

all of these words in the URL of the page

the exact phrase in the title of the page

any of these words any part of the page

Learn to recognize URLs leading to invisible web databases any part of the page

Tip: Use these options to look for an exact phrase or to exclude pages containing certain words. You can also limit your search to certain parts of pages.

Updated

Site/Domain

Any domain

Only .com domains Only .edu domains

Only .gov domains Only .org domains

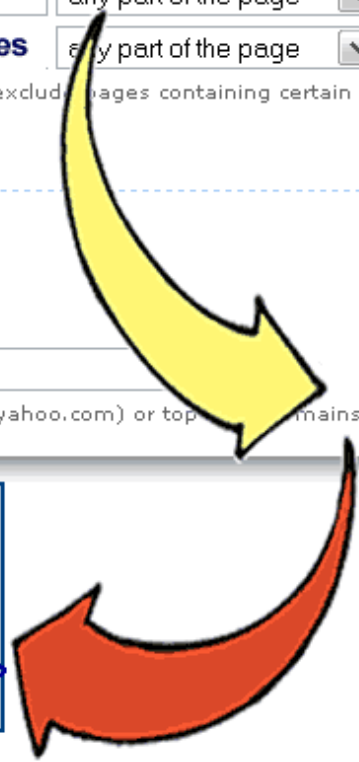
only search in this domain/site:

Tip: You can search for results in a specific website (e.g. yahoo.com) or top-level domains (e.g. .com, .org, .gov).

Search the Library

UT Southwestern **Library** is an academic medical **library** that supports the educational, research and clinical needs of the UT Southwestern faculty, staff and students. ... **Search the Library Results. for Databases** ... Alcohol and Alcohol Problems Science **Database** (NIAAA) <http://etoh.niaaa.nih.gov/> ...

swmed.edu/cfdocs/library/search/search.cfm?subject=Databases§ion=2 - 251k - [Cached](#) - [More from this site](#) [Save](#) [Block](#)



Do some websites have both visible and invisible content?

Yes, many websites have both Invisible Web content as well as resources indexed by search engines. Invisible Web materials might be locked behind a password barrier. Other pages might have been skipped by the search engine's crawler. Still those pages are easily available to you using the site's own search tools, or site map.



Do Invisible Web resources ever become visible?

What's invisible today may become visible tomorrow. This is especially true when it comes to resources in non-HTML formats. Search engines are changing their policies about which formats they index. Google's decision to index images and PDF files created market pressure that has spurred competitors to add other formats to their search indexes. At the moment, search engines do not index database contents. This too could change as technology advances.

Are Invisible Web resources the same for every search engine?

No, what is invisible to one search engine might be indexed by another. *This is one of the reasons you should use more than one search engine when seeking information.* Search engines use different types of crawlers and don't always index the same pages or websites. Policies vary on which non-html file formats to index. At present, the dynamic, on-demand information generated by online databases remains beyond the reach of all search engines.

Are there good collections of Invisible Websites available on the public Internet?

There are many websites that specialize in collecting and presenting access to Invisible Websites. The websites listed below are a good place to start your investigation of the Invisible Web. These [Invisible Web Resources](#) are in the Resources, Annotated Links area of the IMSA Portal:

Chris Sherman and Gary Price. The Invisible Web Directory [Website]. :2001 [cited 22 February 2003]. Available from World Wide Web: <<http://invisible-web.net/>>.

A directory of some of the best resources of the Invisible Web. This site was developed by reference librarian Gary Price and Chris Sherman, Associate Editor of SearchEngineWatch.com, as a companion to their 2001 book *The Invisible Web: Uncovering Hidden Internet Resources Search Engines Can't See*.

Librarians' Index to the Internet: "Invisible or Hidden Web" [Web Directory]. :Lii.org The Library of California:2003 [cited 24 June 2003]. Available from World Wide Web:<[http://lii.org/search?query=\(invisible+or+hidden\)+Web;searchtype=keywords](http://lii.org/search?query=(invisible+or+hidden)+Web;searchtype=keywords) >

The Librarian's Index to the Internet is a searchable, annotated subject directory of more than 11,000 Internet resources selected and evaluated by librarians for their usefulness to users of public libraries. The Librarian's Index to the Internet (lii.org) is a reliable and efficient guide to Internet resources. This link will take you to the most current results for the keywords: (invisible or hidden) Web.

Authored by Dennis O'Connor 2003-2005



[1](#) ... [3](#) [4](#) [5](#) [6]

End of Micromodule - Invisible.

Return to Micromodule [List](#)