



## *MicroModule: opaque*

### **What Is The Nearly Invisible Web Or Opaque Web? How Can I Search It?**

#### **REVIEW Page**

Below is the entire module on one page.



The public web is open and freely available to search engines. The invisible web also part of the Internet, but inaccessible to the robotic web-crawling technology search engines use to automatically build and update their indexes. (For more on this topic, see the IMSA Micro Module: What Is the Invisible Web?) In this module we will consider information that bridges the public and invisible webs, the 'nearly-visible' or 'opaque' web.

Think of the nearly visible or opaque web as web pages that are just one click beyond the reach of a search engine. The website itself has been visited and some of its pages are copied into the search engine Index. However, due to storage limitations, not all pages on a site are visited by every search engine. The opaque or nearly visible web is information on a public website that has not been indexed by the robotic 'crawlers' or 'spiders' sent out by the search engine. Indeed the information is 'indexible' but it hasn't yet been indexed.

Why would this happen? Crawling the web is expensive because storage is expensive. For this reason search engines impose limits on the number of pages they record at any given site. With a limited 'depth of crawl' the robotic spider might copy 150 to 300 pages from a site, and leave 700 pages out of the index for that site. This un-indexed information is said to be part of the nearly invisible or opaque web. The information is out there, but you'll have to find your way to it indirectly by following links on the website. You can click to the web pages once you are on the site, you just won't see the pages showing up on a search engine hit list.

Sometimes a Webmaster may choose to 'hide' a page from search engine crawlers using special html code that instructs crawlers to skip pages or sub-directories of information. This code is placed in a file called robots.txt. Additionally the NOINDEX meta tag can be added to a page, which will then be automatically skipped by a search engine crawler. The html NOFOLLOW meta tag allows a page to be indexed, but blocks the spider from following links on that page. While these codes make the information invisible to crawlers, you can still see and use the pages when you are visiting the website.

Keep in mind that each search engine has its own unique index. What is opaque to one search engine might be indexed and highly visible to another search engine. This is another good reason to always use three different search engines when looking for information. Also, it's hard to know how long a page will remain hidden. Search engines are constantly updating and revising their index systems. What's opaque today may be visible tomorrow.

So how can you find information if it doesn't appear on a search engine's hit list? By knowing that important information may be hidden behind the next click on a web page, you'll be more disposed to look deeply into the sites you visit. If you find a good website, spend time exploring it at depth. If the website has a sitemap use it to dig into the information, who knows, you may unearth an opaque gem of information that will shine when held up to the lens of your research! (For more on these topics see the IMSA Micro Modules: How Can You Search An Individual Web Site In Depth? & What is a Sitemap?)

## **FAQs**



### **What is the nearly invisible or opaque web?**

The opaque or nearly visible web is information on a public website that has not been indexed by the robotic 'crawlers' or 'spiders' sent out by the search engine. The information is 'indexible' but it hasn't yet been indexed. Crawlers may just miss the page because they have limited the number of pages they index from each site. Also webmasters can exclude pages by using special HTML codes.

### **How can I search the opaque web?**

By knowing that important information may be hidden behind the next click on a web page, you'll be more disposed to look deeply into the sites you visit. If you find a good information source, spend time exploring it at depth. If the website has a sitemap use it to dig into the information. If a site provides a search box, use keywords to quickly find what you are looking for. (For more on these topics see the IMSA Micro Modules: How Can You Search An Individual Web Site In Depth? & What is a Sitemap?)

### **How does search engine 'depth of crawl' create the opaque web?**

Some engines impose limits on the number of pages they record at any given site. With a limited 'depth of crawl' the robotic spider might copy part of a site, while leaving other pages out of the index for that site. If a website has a thousand pages, but only 100 are crawled and indexed, the depth of crawl has created a good deal of 'opaque web' content.

### **Are their ways to make a webpage intentionally 'opaque'?**

A webmaster may choose to 'hide' a page from search engine crawlers by using special html code that instructs crawlers to skip pages or sub-directories of information. This code is placed in a file

called robots.txt. Additionally the HTML NOINDEX meta tag can be added to a page, which will then be automatically skipped by a search engine crawler. The HTML NOFOLLOW meta tag allows a page to be indexed, but blocks the spider from following links on that page.

### **Are the same pages nearly invisible or opaque to all search engines?**

Each search engine has its own unique index. What is opaque to one search engine might be indexed and highly visible to another search engine. This is another good reason to always use three different search engines when looking for information. Also, how long a page will remain hidden is hard to determine. Search engines are constantly updating and revising their index systems. What's hidden today may be visible tomorrow.

[Authored by Dennis O'Connor 2003](#)



[1](#) [2](#) [[3](#)]

*End of Micromodule - opaque.*

*Return to Micromodule [List](#)*