



Coffee Break Training - Info Search and Web 2.0

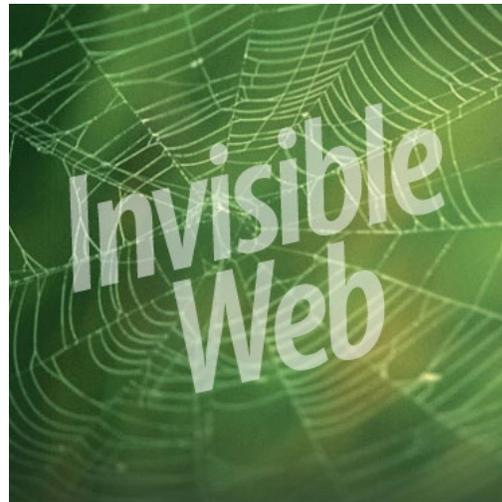
Exploring Invisible Web Resources: Part I

No. IS-2009-6 May 6, 2009

Learning Objective: The student shall be able to define what the Invisible Web is and identify several significant first responder resources not otherwise accessible via Google or other search engines.

It's hard to imagine that a search engine returning many thousands of hits on a topic you're exploring might not actually have access to all that's out there on the Web. Experts estimate that actually some 80 percent of authoritative content on the Web is invisible to search engines. The Invisible or Deep Web is many times larger, in fact, than the "surface Web" we have access to via search engines like Google. To better grasp what we mean by the term "**Invisible Web**" let's first consider what is **visible**.

Basically, the visible or "surface" Web refers to any online content that can be found and accessed by a search engine. Search engines send out robot programs or spiders, as they are sometimes called, to crawl the Web looking for pages to add to a search engine database. These pages then get passed to the search engine's indexing program which stores the content in its database. This database in turn is what you're actually searching when you use a Web search engine.



Search engine spiders have gotten a lot better in recent years in crawling non-HTML files (like pdf or .doc files), but some online resources, particularly databases (or other sites with dynamically generated content), often are still inaccessible to them. So what happens then when a spider encounters a database as opposed to a simple Web page?

Well, a search engine spider can't enter a username or password. It can't enter a keyword search and pass the results to its database either. Most commercial databases restrict their content (via password or IP recognition) just to paying subscribers. In short, this means that many types of online resources with highly authoritative content such as digital libraries or repositories, journal article databases, and statistical databases have to be searched separately. Their content is "invisible" to search engines. That's why it's not enough to be good at ferreting out information on Google, or the other search engines. While the size of their databases is large, they by no means provide access to all or even most of the information available to you.

This is as true for the fire and emergency services domain as it is for other subject areas. If you're not already, you should become familiar with important first responder-related databases and what each has to offer. In subsequent Coffee Break Bulletins we'll take a closer look at several prominent "Invisible Web" resources for the first responder community including the:

Lessons Learned Information Sharing (LLIS) <https://www.llis.dhs.gov/>
Homeland Security Digital Library (Restricted Collection) <https://www.hsdl.org/>
NETC Learning Resource Center Online Public Access Catalog (OPAC) <http://www.lrc.fema.gov/index.html>
Responder Knowledge Base <https://www.rkb.us/>
National Firefighter Near-Miss Reporting System <http://www.firefighternearmiss.com/>

You'll want to bookmark these and consider them as your "virtual" reference library to be used for particular needs and alongside traditional Web search engines.