

Those Dark Hiding Places: The Invisible Web Revealed

- Robert J. Lackie, Associate Professor-Librarian, Rider University



Note: Site no longer being updated (Final Version can be found in PDF form [here](#))

"If only I had known!" was the bitter cry of the searcher who relied just on search engines to search the Web. Although many popular search engines boast about their ability to index information on the Web, more of it (dynamically-generated pages, certain file formats, and information held within numerous databases) has become invisible to their searching spiders. Much of the Web is hiding information from us, but we can access this hidden content! Learn how you can reveal the secrets of these dark, hiding places.



Hidden Content on the Web

Note "The Web," according to Chris Sherman, Internet search expert and Associate Editor of SearchEngineWatch.com, "is increasingly moving away from being a collection of documents and becoming a multidimensional repository for sounds, images, audio, and other formats." Because much of this information is not accessible to many general search engines' software spiders, we need to look for specific search tools that will lead us to this hidden content. Some of these tools include *directories*, *searchable sites*, *free Web databases*, and a *few general and many specialized search engines*. Begin searching with...

- [Directories and Portals](#) when you:
 - have a broad topic
 - want selected, evaluated, and annotated collections
 - prefer quality over quantity
- [Invisible or Deep Web](#) [searchable sites and databases] when you:
 - are looking for information that is likely in a database
 - are looking for information that dynamically changes in content
- [Search engines](#) [general and specialized] when you:
 - have a narrow topic
 - want to take advantage of the newer retrieval technologies



Directories

Note Directories are Web sites that provide a large collection of links, arranged according to a classification scheme that enables browsing by subject area. I really like directories, but what I want to point out right away is that I am not against using search engines. I consider directories to be complements to search engines, not their replacements. However, there is a trend

developing toward the use of directories because, in addition to their classification, their content is pre-screened, evaluated, and annotated by humans. Sometimes, though, this annotation and classification process makes the information not as timely as it could be. This is usually true in very large directories, so look at several, large and small. Let's look at a few smaller, more selective directories that can also lead you to some of the Web's hidden content.

- [Librarians' Internet Index \(http://lii.org/\)](http://lii.org/) - *Websites You Can Trust*: LII offers a searchable and browsable collection of over 20,000 quality websites, "maintained by librarians and organized into 14 main topics and nearly 300 related topics," in addition to an excellent weekly newsletter [they have over 40,000 subscribers in many countries], available by email or RSS, of high-quality Websites related to current events, holidays, and popular and important issues. New features added with their Fall 2005 upgrade include icons following the titles allowing you to view more details, make comments about, or e-mail the site. Of course, LII can also lead you to Invisible Web databases by typing in a broad topic and adding the words: "and databases" (i.e., *biology and databases*).
- [FindLaw \(http://www.findlaw.com/\)](http://www.findlaw.com/) - "FindLaw is the web's premier free legal information site, reaching hundreds of thousands of unique visitors daily. FindLaw incorporates case law, legal news, cutting-edge commentary, legal technology trends, practice tips, message boards, RSS feeds, over 60 newsletter titles, and much more, to create a vibrant online community for today's legal professional. FindLaw was founded in 1995 as a repository for free legal information on the web and has grown to become the award-winning standard for legal websites." **NOTE:** To find an annotated list of free databases on many law-related topics, from their main page, click on the "Visit our professional site" link in the top right corner (or at the very bottom information section under "For Lawyers", click on the words "Visit our professional site"). Then under the "Research the law" section click on the "View all by practice area" link under the "browse by practice area" section, pick a practice area/legal subject heading (i.e., "Health Law"), and then look for "Databases" under the "Web Guide" for that legal subject heading.
- [InfoMine \(http://infomine.ucr.edu\)](http://infomine.ucr.edu) - This scholarly resource collection includes tens of thousands of sites, grouped into 9 annotated, indexed categories (databases) for easy retrieval. This librarian-built "virtual library of Internet resources [is] relevant to faculty, students, and research staff at the university level," while also very useful for higher-level high school and professionals, too. "It contains useful Internet resources such as databases, electronic journals, electronic books, bulletin boards, mailing lists, online library card catalogs, articles, directories of researchers, and many other types of information."
- [About.com \(http://www.about.com/\)](http://www.about.com/) - This portal, visited each month by more than 29 million people, neatly organizes, thousands of topics, including Invisible Web, with good news and commentary. Try typing "Invisible Web" as a phrase in quotes to find many links to hidden content on the Web, including the "Invisible Web: The Cloaked Internet," "Visible versus Invisible Web," and their new, "The "Cloaked" or "Deep" Web, Explained," from their Internet for Beginners guide, and "Invisible Web Gateways." You will see links to other pertinent articles, too--all worth reading & exploring.



Invisible Web Searchable Sites

Note

Chris Sherman states that "vast expanses of the Web are completely invisible to general purpose search engines," but there are ways "to find the hidden gems search engines can't see."

Some Recommended Links to Invisible Web Databases:

• [ResourceShelf \(http://www.resourceshelf.com/\)](http://www.resourceshelf.com/) - Gary Price, MLIS, of *Gary Price Library & Internet Research Consulting*, one of the foremost authorities on invaluable Invisible Web resources, has assembled a massive collection at his [Direct Search \(http://www.freepint.com/gary/direct.htm\)](http://www.freepint.com/gary/direct.htm) found on his "ResourceShelf" Weblog & Newsletter site for information professionals and online researchers. Other well-known Web research tools, including "Price's List of Lists," are included on the left-hand list of links (below the list of dates).

• [CompletePlanet \(http://www.completeplanet.com/\)](http://www.completeplanet.com/) - BrightPlanet Corporation's site contains "over 70,000 searchable databases & specialty search engines." Check out their [Help/FAQs](#) section "[How to search CompletePlanet](#)", "[CompletePlanet FAQs](#)", "[Deep Web FAQs](#)", and the "[Largest Deep-Web Sites](#)". This site is "A comprehensive listing of dynamic searchable databases. Find databases with highly relevant documents that cannot be crawled or indexed by surface web search engines."



Some Invisible Web Databases

Note

Although there are thousands of Invisible Web databases available to us for free on the Web, below I have listed a few of my favorites:

- [AnimalSearch](http://animalsearch.net/) (<http://animalsearch.net/>) - A database for family-safe animal-related sites, you can search by group, type, and geographic regions, and "each site is reviewed, prior to inclusion, for content relevancy and safety" It also has low cost animal wallpaper and e-cards.
- [Educator's Reference Desk](http://www.eduref.org/) (<http://www.eduref.org/>) - This site contains 2000+ lesson plans, 3000+ links to value-added online education information, and 200+ question archive collected on the award-winning AskERIC site during the past decade. This site also provides access to the ERIC database--the world's largest source of information on education research & practice, including free, full-text expert digest reports, and it also links you to the Gateway to Educational Materials (GEM), which "provides quick and easy access to over 40,000 educational resources found on various federal, state, university, non-profit and commercial Internet sites."
- [NatureServe Explorer](http://www.natureserve.org/explorer) (<http://www.natureserve.org/explorer>) - This online encyclopedia provides authoritative "information on more than 70,000 plants, animals, and ecosystems of the United States and Canada. Explorer includes particularly in-depth coverage for rare and endangered species."
- [Nuclear Explosions Database](http://www.ga.gov.au/oracle/nuclear-explosion.jsp) (<http://www.ga.gov.au/oracle/nuclear-explosion.jsp>) - Geoscience Australia's database provides location, time, & size of explosions worldwide since 1946. Click on "databases" under "Online Tools" to see a list of other searchable online mapping tools & databases.
- [On-Line Encyclopedia of Integer Sequences](http://www.research.att.com/~njas/sequences/) (<http://www.research.att.com/~njas/sequences/>) - "Type in a series of numbers and this database will complete the sequence and provide the sequence name, along with its mathematical formula, structure, references, and links."
- [PubMed](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) - Provides access to over 17 million MEDLINE citations, including links to full text articles & related resources. You will also want to explore PubMed Central (PMC) in the "Related Resources" section on bottom left. This site is an e-archive of free, full text articles from almost 400 life sciences journals, as well as Bookshelf, "a growing collection of [full text] biomedical books (70+) that can be searched directly." They now offer a "new global NCBI 'Entrez' search engine" where you can search across their many life sciences databases, too.
- [FindArticles](http://www.findarticles.com/) (<http://www.findarticles.com/>) - The FindArticles database is an updated replacement of their original free, searchable article Web archive, with the current service now searching 10 million+ articles "from the back issues of over 900 magazines, journals, trade publications and newspapers." according to Alexa.com , "The Web Information Company" that does Traffic Rankings and Overviews of websites and search engines on the web.
- [MagPortal.com](http://magportal.com/) (<http://magportal.com/>) - MagPortal.com is another site for finding freely available magazine articles on the Web, using keyword searching or category browsing methods. Indexing a close to 200 magazines and only the articles that are freely accessible online are indexed. This index's focused content allows it to update with new articles within days of them becoming available. The material is of good quality, and their Hot Neuron Similarity software package allows them to measure the similarity between articles, linking similar articles to each other.
- [Directory of Open Access Journals](http://www.doaj.org/) (<http://www.doaj.org/>) - Launched in May 2003, Sweden's Lund University Libraries Head Office hosts this "one-stop shopping" open access directory, providing no-cost access to the full text of over 4170 journals, with over 1,509 journals are searchable at article level (over 280,420 articles available). "This service covers free, full text, quality controlled scientific and scholarly journals. We aim to cover all subjects and languages."
- [HighWire Press: Free Online Full-text Articles](http://highwire.stanford.edu/) (<http://highwire.stanford.edu/>) - Launched in early 1995, Stanford University Libraries' HighWire Press hosts the largest repository of high impact, peer-reviewed content, with 1,245 and 5,884,561 full text articles from over 140 scholarly publishers. HighWire-hosted publishers have collectively made over 1.8 million articles free. With our partner publishers we produce 71 of the 200 most-frequently-cited journals. I like how it also provides very quick full-text access to your institution's journal subscriptions to HighWire-affiliated journals via IP

address recognition when using a computer workstation within your library/institution--journals to which you probably did not even know that you had access! (Click on "For Institutions" tab on the top and follow the directions). You can also browse by topic or alphabetically on this page--you will be impressed!

Note By the way, if you like viewing accompanying Web sites from excellent books on Web research, you may also want to visit the [Super Searchers Web Page](http://www.infotoday.com/supersearchers/) (<http://www.infotoday.com/supersearchers/>), which "features a growing collection of links to subject-specific Web resources recommended by the world's leading online searchers" in global business, primary research, mergers/acquisitions, news, writing, health/medicine, investment, business, entrepreneurial research, & legal information resources. The books and their Web sites can lead researchers to a wealth of hidden resources.



Search Engines

Note Some general and specialized search engines, like those listed below, can help you locate specific information or certain file formats, so I like to go to them first. I do use several search engines for research, but they are not all created equal when it comes to uncovering data in the Invisible Web domain. A great site for keeping up-to-date on search engines is [Search Engine Watch](http://www.searchenginewatch.com/) (<http://www.searchenginewatch.com/>). Another great site on search engines is [Search Engine Showdown](http://www.searchengineshowdown.com/) (<http://www.searchengineshowdown.com/>). Let's explore these two sites and general & specialized search engines that allow us to find some Invisible Web data. Immediately below are a few interesting specialized search engine services/sites.

- [Google News](http://news.google.com/) (<http://news.google.com/>) - This award-winning automated (no Google editors) version scours the Web every 15 minutes, capturing news from 25,000. Recently, Google News added a new feature: a "Top Stories" section that allows us to select the top news stories from several different countries. They even have an "Advance Archive Search," an "Advanced News Search", and a "Blog search". Note: Yahoo! News, Topix.net, and Daypop are also impressive news-aggregating services with special features, too.
- [Scirus](http://www.scirus.com/srsapp/) (<http://www.scirus.com/srsapp/>) - This science search engine, with over 450 million science-specific Web pages, offers excellent advanced search options for a wide variety of information types and sources of materials on the Web, including journals. Scirus has become pretty successful at pinpointing science-specific data, reports, articles, and relevant scholarly Web pages--a considerable recent improvement. It also allows researchers to search for not only journal content but also scientists' homepages, courseware, pre-print server material, patents and institutional repository and website information. Check out their NEW Scirus Topic Pages with the new brand name [SciTopics](http://www.scitopics.com) (<http://www.scitopics.com>) that launched on 20th Jan 2009.
- [UFOSeek: The UFO and Paranormal Search Engine](http://www.ufoseek.com/) (<http://www.ufoseek.com/>) - "Yes, Mulder, the truth is really, um, out there, and you can find it using this paranormal/UFO search engine," currently indexing 209,032 Paranormal, Spiritual and UFO sites in the their system.

Note We know that information on some sites is presented in formats other than static HTML, which gives search engines a problem. Adobe Portable Document Format (PDF) has been an example of this. If HTML text that accompanies the PDF file describes the file well, you may find the site, but if the site provides unhelpful headings or titles, then the file is pretty much "invisible." This is also true for Flash files, for instance. Fortunately for us, a few general search engines are more easily bringing some PDF, Flash, and other non-HTML files to our desktops.

- [Google](http://www.google.com/) (<http://www.google.com/>) - Still the most popular general purpose search engine on the Web, Google allows you to go to the page as it is currently on the Web, or go to a cached copy Google stored when it retrieved the page (nice when the current page won't connect). In addition, Google allows you to find those Invisible Web documents: PDF files. You can also view them in HTML (nice when you have a slow connection or the PDF is so large that you don't want to wait to display). From Google's Advanced Search, you will see that in addition to allowing you to limit your search to finding PDF files, you can limit or exclude other file formats, such as Postscript; Microsoft Word, Excel, or PowerPoint; & Rich Text formats. Check out their "Google Web Search Features." Note: Google claimed (in August 2005) to track 11.3 billion objects--which consist of the some 8.2 billion Web pages and 2.1 billion images, as well as material from its group discussions--it no longer lists figures on its main pages.
- [Yahoo! Search](http://www.yahoo.com/) (<http://www.yahoo.com/>) - Google's biggest competitor since dropping them as a partner, Yahoo! (selected in spring 2005 by Search Engine Watch as the "2004 Outstanding Search Service Winner") also provides cached

copies and locates Word, Excel, PowerPoint, PDF, and RSS/XML files. Yahoo! also has full Boolean searching capability after purchasing the AlltheWeb and AltaVista search engines, so it looks like Google is going to be keeping an eye on Yahoo!'s continued aggressive progress. Check out their interesting "Yahoo! Shortcuts" (<http://tools.search.yahoo.com/newsearch/resources>) for fun ways to quickly find everyday information, as well as their Yahoo! Search Subscriptions (<http://search.yahoo.com/subscriptions>), which enables you to search access-restricted content such as news and reference sites that are normally not accessible to search engines. **Note:** Yahoo! (in August 2005) stated that its index covered 20.8 billion online objects, made up of about 19.2 billion documents and 1.6 billion images--partly because of a 2005 upgrade--like, Google, figures are not listed on Yahoo's main pages.

- **Gigablast** (<http://www.gigablast.com/>) - An interesting up-and-coming search engine, "Founded in 2000, Gigablast was created to index up to 200 Billion pages with the least amount of hardware possible." Gigablast also locates Word, Excel, PDF, and other non-HTML files, and like Google and Yahoo!, it provides cached (most recent "archived copy") of these files. It also links you to multiple "older copies" via The Internet Archive Wayback Machine. In addition, it also provides full Boolean searching, so keep an eye on Gigablast, too.



Note

FYI: Below are a few of my recent articles on the invisible/hidden web (and other education-related topics) for your review; other articles/presentations can be found at my [Robert J. Lackie's Selected Online Materials](http://www.robertlackie.com/rlackiepub.html) (<http://www.robertlackie.com/rlackiepub.html>) page:

- Klapperstuck, Karen J., Robert J. Lackie. "[Cool Tools for Content Creation: More Than Blogs or Wikis.](#)" (cover article). *MultiMedia & Internet@Schools: The Media and Technology Specialists' Guide to Electronic Tools and Resources for K-12 Education* 16.2 (2009): 12-15. Note: An HTML full text version of the article is available via the MMISchools.com Website.
- Lackie, Robert J., and Robert D. Terrio. "[Mashups and Other New or Improved Collaborative Social Software Tools.](#)" *MultiMedia & Internet@Schools: The Media and Technology Specialists' Guide to Electronic Tools and Resources for K-12 Education* 14.4 (2007): 12-16. (PDF article). Note: A Google Docs Presentation on mashups is available [here](#).
- Lackie, Robert J. "[Web 2.0 and Its Technologies for Collaborative Library Communication.](#)" (cover article). *MultiMedia & Internet@Schools: The Media and Technology Specialists' Guide to Electronic Tools and Resources for K-12 Education* 13.6 (2006): 9-12. (PDF article). Note: An HTML full text version of the article is available via the MMISchools.com Website.
- Lackie, Robert J. "[Social Networking Sites and Personal Information Search Engines.](#)" *Internet Librarian 2006: Collected Presentations, Monterey, CA, 23-25 Oct. 2006*. Comp. Carol Nixon and Jennifer Burmood. Medford: Information Today, 2006. 147-52. (PDF article and Website).
- Lackie, Robert J. "[The Changing Face of the Scholarly Web: Finding Free, Quality, Full-Text Articles, Books, and More!](#)" *MultiMedia & Internet@Schools: The Media and Technology Specialists' Guide to Electronic Tools and Resources for K-12* 13.4 (2006): 8-12. (PDF article). Note: An HTML full text version of the article is available via the MMISchools.com Website.
- Lackie, Robert J., and Robert J. Congleton. "[Free and Fee-Based Online Science Resources for the K-12 Community.](#)" *MultiMedia & Internet@Schools: The Media and Technology Specialists' Guide to Electronic Tools and Resources for K-12* 11.6 (2004): 10-16. (PDF article). Note: An HTML full text version of the article is available via the InfoToday Website.
- Lackie, Robert J. "Internet@Schools: [Finding Free Educational Resources for Teachers & Librarians.](#)" *Internet Librarian 2004: Collected Presentations, Monterey, CA, 15-17 Nov. 2004*. Comp. Carol Nixon and Jennifer Burmood. Medford: Information Today, 2004. 296-302. (PDF article and PowerPoint).
- Lackie, Robert J. "[The Evolving 'Invisible Web': Tried-and-True Methods and New Developments for Locating the Web's Hidden Content.](#)" *College & Undergraduate Libraries* 10.2 (2003): 65-71. (PDF article).
- Lackie, Robert J. "[Science and Math Web Resources for the Higher Education Community.](#)" *Online: The Leading Magazine for Information Professionals* 27.6 (2003): 35-39. (PDF article).



Send comments or questions about this workshop and/or Web site to [Robert J. Lackie](mailto:rlackie@rider.edu) (rlackie@rider.edu), including if you would like permission to link to [Those Dark Hiding Places: The Invisible Web Revealed](http://www.robertlackie.com/invisible/index.html) (http://www.robertlackie.com/invisible/index.html).



NicheUSA, L.L.C. As a consultant for *NicheUSA* with its ZoomerOne (software tool for finding [best web resources](#)) product, I help with educational website recommendations. If you are interested in quality Web sites, directories, and portals for social studies, science, math, and language arts for kids (grades 3 to 12), then visit my recommended listings housed on the [NicheUSA' Education ZoomerOne links homepage](#) (http://eduzoomerone.wikispaces.com/).



This site was selected as a Hot Site in the June 11, 2001 edition of *USATODAY.com*, a free, highly popular Web news service. Check out other Hot Sites by clicking on their logo.



This site was selected as Reference Site of the Day on June 12, 2001, by *Refdesk.com*, "The single best source for facts on the Net; a one-stop site for all things Internet." Click on their logo for other Sites of the Day.



This site was also selected on July 5, 2001, for inclusion in *Librarians' Internet Index*, a searchable and browsable collection [maintained by librarians] of over "tens of thousands" of quality websites related to "current events, holidays, and popular and important issues." Click on their logo to search lii.org.



This site was selected as the "Internet Site of the Week" in the IT (Database) Section of the February 16, 2005 edition of the *Bangkok Post*, "The World's window to Thailand and the region," and one of Thailand's leading English-language newspapers.

Those Dark Hiding Places: The Invisible Web Revealed is produced by Robert J. Lackie, Associate Professor-Librarian at Rider University, Lawrenceville, New Jersey, where he co-leads the Franklin F. Moore Library's Instruction Program and serves as Library Liaison to the Biology, Chemistry & Physics, Mathematics, Teacher Education, and Graduate Education & Human Services Departments. He received his Master of Library and Information Science at the University of South Carolina and his Master of Arts in Curriculum, Instruction, & Supervision at Rider University. In April 2004, he was selected by the [New Jersey Library Association](#) as the [2004 Librarian of the Year](#), and in May 2004, he was chosen as a recipient of the [2004 Rider University Award for Distinguished Teaching](#). In 2005, he was honored to be selected for inclusion in the 60th Diamond Anniversary (2006) Edition of [Who's Who in America](#), and in June 2006, he received the American Library Association's [2006 Ken Haycock Award for Promoting Librarianship](#). (Click [here](#) for detailed information on [Robert J. Lackie's](#) seminars/workshops, curriculum vitae, short biography, selected publications/presentations, etc.).

Note

Many of the spider gifs found on this site are credited to Lisa Konrad at [Animation Arthouse: Spiders](http://www.animation.arthouse.org/spider.html) (http://www.animation.arthouse.org/spider.html). Special thanks to William A. Lackie for his technical advice and design assistance with this Website. Also, many thanks to Anne Clyde, Laura Cohen, Greg Notess, Gary Price, Chris Sherman, Danny Sullivan, and Wei-hsing Wang for their valuable information and research.



