

2000

The Invisible Web: Navigating the Web outside Traditional Search Engines

Karen R. Diaz

Follow this and additional works at: https://researchrepository.wvu.edu/lib_faculty



Part of the [Library and Information Science Commons](#)

Digital Commons Citation

Diaz, Karen R., "The Invisible Web: Navigating the Web outside Traditional Search Engines" (2000). *University Libraries Faculty Scholarship*. 9.

https://researchrepository.wvu.edu/lib_faculty/9

This Article is brought to you for free and open access by the University Libraries at The Research Repository @ WVU. It has been accepted for inclusion in University Libraries Faculty Scholarship by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.

Reference and User Services Quarterly, Vol. 40, No. 2, 2000, pp.131-134.

(Print: 0033-7072) (Electronic: 1094-9054)

<http://www.ala.org>

<http://find.galegroup.com/itx/infomark.do?contentSet=IAC->

[Documents&docType=IAC&type=retrieve&tabID=T002&prodId=AONE&docId=A69653577&userGroupName=c](http://find.galegroup.com/itx/infomark.do?contentSet=IAC-Documents&docType=IAC&type=retrieve&tabID=T002&prodId=AONE&docId=A69653577&userGroupName=c)

[olu44332&version=1.0&searchType=PublicationSearchForm&source=gale](http://www.rusq.org/)

<http://www.rusq.org/>

© American Library Association

The Invisible Web Navigating the Web outside Traditional Search Engines

Karen R. Diaz, User Education Librarian for Internet Instruction, Ohio State University, Columbus, OH

You've heard the pitch about how big the Web is—some 380 million pages and counting. But did you know that there is a virtual treasure trove of information that is not visible through the traditional search engines? Did you know that there are thousands of searchable databases, archives, and other information sources delivering highly targeted information for a much improved searching experience?—*The Invisible Web Catalog on the Lycos Network*

Search engines find Web sites, Web pages, and in some cases, specified types of documents. Some provide more comprehensive results while others provide more precise results. However, what most search engines rarely find are records in databases. Increasingly, valuable sorts of information reside in databases and remain out of reach of standard search tools. This reality is becoming more apparent and getting increasing press among would-be power searchers. Web sites are popping up to address the issue and point researchers in the right direction. Conferences are offering workshops on "The Invisible Web."

What Is the Invisible Web?

Plain and simply, the Invisible Web consists of the component parts of the Web that are hard or impossible to find through prominent search engines and directories. While these items remain outside the scope of traditional Web search tools, they still reside "on" the Web. Generally, the Invisible Web is comprised of records in databases. The term itself is by no means an official term or one that is used by everyone addressing the issue. Other terminology used includes "searchable databases" or "special databases."

Why Use Databases?

Databases can be used for many purposes. The readers of *Reference & User Services Quarterly* are overwhelmingly likely to understand this point. Consider your own reference habits. When you want to know if your library has a book, you consult your online catalog, a database, via the Web. If you are trying to verify a bibliographic citation, you consult one of a variety of bibliographic databases to which your library has access. If you want to find a phone number for someone across the country, you consult a Web database (such as WhoWhere? or 411.com) of phone numbers. Your reference habits are such that you likely already have considered, either consciously or unconsciously, the notion of compartmentalized information on

the Web. The reasons for using databases can be summed up as follows:

1. Use a database to get the exact type of information you need quickly. For example, to find the address of Microsoft, do a quick search in a business database. You would waste a lot of time going through hundreds of hits (news releases, articles, personal Web sites, etc.) to locate an address via a search engine.
2. Some information on the Web is only available in a database. A Web index might find a database site for you, but may not find the contents of the database made available within that site. Sometimes there are riches "below the surface" of the Web.
3. Use a database to get expected results. Databases will search a defined, understandable, and (hopefully) logical collection of items. There are times you are pleased with serendipitous and unexpected results, and times you are not.
4. Use a database to get to information more quickly or to information that is more current than what is available in traditional print resources. Examples of databases that can meet this need include FedStats and the Universal Currency Converter. FedStats (www.fedstats.gov) provides access to statistics maintained by various agencies of the federal government. You only need to know your topic; FedStats will find where the statistics are reported regardless of agency. The Universal Currency Converter (www.xe.net/ucc) updates rates once per minute, which is certainly more current than yesterday's newspaper.

Types of Databases

Fee or Free?

Some databases are accessible to anyone without cost, while others are available only via a subscription rate. Traditional library tools and the vendors we do business with usually suffice to keep us abreast of fee-based services. However, it is important to learn the tools to access the increasing number of valuable databases available for free.

Hybrid Databases

Some databases provide a hybrid of content that is free and content that is available for a fee. For instance, a database might allow free access to bibliographic records and charge for full text articles retrieved. Or a database might provide free access to current records and charge for archival records. Two examples of these sorts of databases are UnCoverWeb (<http://uncweb.carl.org>) and Northern Light's Special Collection (www.northernlight.com).

Bibliographic Databases

Of course, bibliographic databases represent the backbone of library databases, originating with library catalogs. Consider also the online index and abstracting services to which your library subscribes. There are more such services freely available on the Web. Two examples are Jake and the U.S. Trademark Electronic Search System. Jake (<http://jake.med.yale.edu>) is a database that indicates where about 23,000 prominent journals are indexed. The U.S. Trademark Electronic Search System (www.ospto.gov/web/menu/tm.html) is a bibliographic database of registered trademarks.

Full Text Databases

Libraries also are increasingly becoming accustomed to purchasing full text databases (databases that contain, and not merely point to, information resources). Many libraries bought, but now have free access to, Britannica Online (<http://britannica.com>). Other examples include Bartleby Online (www.bartleby.com), and Recipe Search from Martha Stewart (www.marthastewart.com/channels/cooking/recipesearch/search.asp).

Other Databases

Other types of databases include the following types of data:

- ◆ Directory: addresses, phone numbers, people, businesses, places, etc. Examples include the Roller Coaster Database (www.rcdb.com) and ZDNet Company Finder (www.zdnet.com/companyfinder).
- ◆ Products: clothing, books, toys, flowers, other items for sale. Examples include Amazon.com (www.amazon.com) and Land's End (www.landsend.com).
- ◆ Statistical: numeric data such as demographics, voting statistics, employment, income, etc. Examples include the Community Health Status Indicators Project (CHSI) (www.communityhealth.hrsa.gov) and BestPlaces.net (www.best-places.net/index.html).
- ◆ Raw Data: numbers or text representing survey results, scientific studies, etc., which remain meaningless until downloaded and put into the proper software. Examples include Raw Weather Data (www.stormchase.com/Raw.html) and Capital Punishment Survey—Raw Data (www.globaldialog.com/-theoacme/cappun4.html).

Effective Use of Databases

You can become a power user of any database by extrapolating rules from the years of training we've had working with more traditional databases.

Scope

Scope describes the type and level of information in a database. Understanding scope helps determine what type of material is and is not available for retrieval. Some common types of scope in bibliographic databases are dates of coverage and publication types (articles only, book chapters, etc.). Department stores might make clothing available for online purchase, but might not include furniture in an online database.

If an item is out of the scope of a database, it might exist, but will never be found in that database. Look for an "About the database" or "Scope" link on a database site if you have questions about what you can expect to find.

Records

In addition to understanding the scope of a database, it is useful to know what fields a

database record contains and which of those fields are searchable. For instance, if you want to find a doctor in your town who can tell you if you have allergies, you need to know that the record contains "medical field" or "area of specialty" along with location information. You would also need to know that these fields are searchable. Some databases provide you with an up-front sample record to explain record structure, others provide "help," some provide browsing by fields in addition to searching. It is often worth (perhaps necessary) spending five minutes getting to know a database and its records before giving up on yielding desired results.

Controlled Vocabulary and Thesauri

Just because a database is free doesn't mean it doesn't also carry some of the hallmarks of quality fee-based databases. Creators of a database using controlled vocabulary define which subject terms are acceptable and assign only those terms as subject headings for each record. The resulting list of terms is often referred to as a thesaurus. When done thoroughly, a thesaurus not only will list acceptable headings, but also indicate related terms, broader terms, and narrower terms for a concept.

Two databases invoking the use of sophisticated and well-developed controlled vocabulary are familiar to librarians: ERIC (www.accesseric.org/searchdb/searchdb.html) and PubMed, the freely available version of Medline (www.ncbi.nlm.nih.gov/PubMed).

Search Engines with a Twist

It is true that even as we attempt to define what search engines will and won't retrieve, new technologies emerge and foil those definitions. Some items have remained outside the scope of traditional search tool retrieval, but are being brought into focus by new sorts of search tools. Following are two relatively new tools that can best be defined as search engines, but the sources from which they draw are limited in scope and provide a more "database like" retrieval. Now it is difficult to define if these items remain "invisible." Newer technologies will likely appear, causing us to make new distinctions and recategorize these finding tools.

Xrefer (<http://wl.xrefer.com>) searches more than thirty resources, including encyclopedias, dictionaries, thesauri, and books of quotations from the world's leading publishers. The site boasts, "Facts, words, concepts; people, places, sayings, and more. All cross-referenced, all reliable, and all in one place." Search Adobe PDF (<http://searchpdf.adobe.com>) claims to search through more than a million summaries of Adobe7 Portable Document Format (PDF) files on the Web. Formerly, search engines were not able to access this type of document.

How to Find Databases

So, what's out there? How do you know if there is a database available for your specific information need? Web indexes and directories remain the overarching tools for finding databases on the Web. Here are some rich sources for finding and learning about databases:

The BigHub.com. www.thebighub.com

Enter your search term in the search box and click on the "specialty sites" button, or browse through die "Specialty Sites" hierarchy available just below the search box.

Chris Sherman's Guide to the Invisible Web

<http://websearch.about.com/internet/websearch/msubmenu120.htm>.

Links to articles about the invisible Web, gateways to databases, and lists of databases on the Web.

Direct Search <http://gwis2.circ.gwu.edu/~gpiice/dirccet.htm>

The site claims to provide a "compilation of links to the search interfaces of resources that contain data not easily or entirely searchable/accessible from general search tools like Alta Vista, Google, and Hotbot"

Librarian's Index to the Internet www.lii.org

Consults the "databases" section retrieved from any broad subject selected.

LITA Toolkit for the Expert Web Searcher

www.lita.org/committe/toptech/toolkit.htm.

Includes information on subject guides, search engines, the invisible Web, and more.

Lycos: Searchable Databases http://dir.lycos.com/Reference/Searchable_Databases

Browse through the broad topics to get to searchable databases of interest to you.

Net. TUTOR/Specialized Databases

<http://gateway.lib.ohio-state.edu/tutor/databases>

An online tutorial for students learning to do research on the Web, developed by the Ohio State University Libraries.

Search.com www.search.com

Allows you search by keyword across several databases on the Web. Or, drill down by topical area and discover which databases are available. Search just one or several at a time.

Yahoo! www.yahoo.com

In the search box, enter a broad subject and "Web directories" (for example, photography and "Web directories"). The annotations of your results will indicate which are searchable databases.

Invisible Web?

The dramatic term "invisible" is meant to underscore the importance of realizing there is more to the Web than even a Meta-crawler search might reveal. In reality, the resources discussed here are not invisible, but merely out of the mainstream of accepted search strategies. The more we learn about the compartments on the Web and the various tools that exist to access these compartments, the less invisible the resources become, and the richer and more productive is our experience using the Web for research.