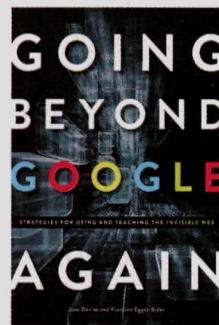# Going Beyond

Strategies for using and teaching the Invisible Web

I t seems unlikely that people will give up their reliance on general-purpose search engines or their practice of beginning a search using Google or one of its competitors. But people should be encouraged to use other research tools when needed, such as databases and more specialized search engines—otherwise known as the Invisible Web.

What makes each of the suggested research tools "invisible" is their ability to uncover resources that general-purpose search engines cannot. Some of the tools do require a subscription or fee. (What follows is just a sample of the tools featured in *Going Beyond Google Again*, published by Neal-Schuman.)

1. Basic Research Tools

The exploration of basic tools can begin with databases, which are at the heart of the Invisible Web. Proprietary databases and those that provide answers dynamically are Invisible Web resources. Databases offer vetted resources that must conform to an editorial standard, and their content is not included in general-purpose search engines.

Users do have to work through the database's own search functions, which may not be as simply laid out as the Google search box.

Student Research

Voice of the Shuttle (VoS):
Website for Humanities Search
vos.ucsb.edu

Sponsored by the English Department of the University of California at Santa Barbara, Voice of the Shuttle (VoS) was started by Professor Alan Liu in 1994. It is a dynamic database of online resources on literature, the humanities, and cultural studies. VoS includes both primary and secondary resources and offers links to course materials, author websites, literature in English and other languages, and ebooks.

By Jane Devine and
Francine Egger-Sider

BizNar: Deep Web Business Search
biznar.com

BizNar scans all kinds of resources, including periodicals such as *Advertising Age,* government resources such as USA .gov and the Bureau of Labor Statistics, news sources *Businessweek* and the *Wall Street Journal,* and social networks like LinkedIn and Wordpress. BizNar represents an example of federated search, which is an approach to accessing materials related to a broad subject, in this case business, by targeting a range of specially selected databases and search engines.

Data.gov: Empowering People, an Official Website of the United States Government
data.gov

The US federal government, which collects all kinds of data, is a good source for student reports or business or general information purposes. They are often invisible because they are buried deep in massive websites and because the government is the unique source for important research and information. Statistical data can be found across many government department websites, but it may be hard to track down with a search.

Internet Archive
archive.org

Home of the Wayback Machine, the Internet Archive is a large repository dedicated to preserving the digital world and offering access to its past through a publicly accessible library. While not fully comprehensive, the collections do a great job of preserving much of the early web for posterity. These resources do not appear on Google search results lists; users must navigate the Internet Archive's own search function. Users can search by subject, by the URL of the site that has disappeared from the web, and by media type.
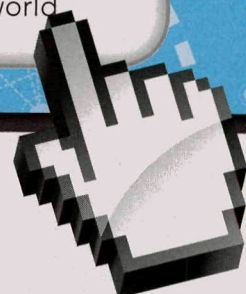
Everyday Life Aids

Pipl: People Search
pipl.com

Pipl, sponsored by the search engine company of the same name, identifies itself as the most comprehensive people search on the web. It claims its success is due to the fact that it taps Invisible Web resources, where a lot of people information is kept; however, it does not list the sources that it relies on.

MedNar: Deep Web Medical Search
mednar.com

MedNar's free search engine utilizes federated search for medical resources. The advanced search option lists specific sites, including government sites, medical societies, and some commercial databases. Search results show the particular collections searched and the number of results

found in each collection. A search can be limited to one collection. Results include an article link and source information. A topic breakdown for the search shows the number of articles available for each related topic.

Topsy: Real-Time Search for the Social Web (Twitter, Google+, Video)
topsy.com

Created by Topsy Labs, an indexing technology company, Topsy can be searched in several languages in general or more specifically by links, tweets, photos, videos, experts, and what it calls "trending." A user can receive subject search results sorted by time periods that range from the last hour, day, week, 30 days, or more. An "expert" search allows the user to look for a subject and get results for people who have been posting on that topic, along with analytics on how often the search terms appear in an individual's postings.

Yummly: Every Recipe in the World
yummly.com

Yummly offers the searcher the opportunity to look for recipes and select from various search result options that include ingredients, cooking time, directions, and the source for the recipe. Additional ways to search include by national cuisines, allergies, and holidays. Yummly searches not only for keywords but for context and intent, utilizing a semantic approach.

2. Second-Layer Tools: More academic

Beyond popular and basic reference tools, resources in this second layer vary from proprietary databases to fee-based resources that cover in depth almost any subject area, especially in the sciences.

WorldWideScience.org: The Global Science Gateway
worldwidescience.org

This database is the product of an alliance among international scientific institutions and is operated by the US Department of Energy. It can be searched in many languages and offers translations. It also offers a list of all the institutional collections included and is a focused federated tool that searches across all of these holdings. The materials offered include conference papers, articles, and other documents not readily found on the surface web.

New York Public Library Digital Gallery
digitalgallery.nypl.org

More than 700,000 digitized images from the New York Public Library, including historical documents, photographs, art pieces, and maps, are featured. The collection can be searched by keyword or browsed by subject. Users can print out images or order high-quality reproductions. The library provides information on how to get copyright approval for images that require it.

DeepDyve
deepdyve.com

DeepDyve is a fee-based resource that offers access to articles and journals to the general public. It can be searched for content as a database and returns results with author, title, source, a line about the purpose of the article, and the cost to access. Holdings can also be browsed by subject, journal title, and publisher. Selecting a journal title brings up all of its contents by volume and issue including the full run of many titles.

DOAJ: Directory of Open Access Journals
doaj.org

This resource offers access to full-text articles from online scholarly publications covering all subjects, favoring research and scientific pubs. While it calls itself a directory, DOAJ is really a database of articles that can be searched using keywords. A browsing option permits users to search journals by subject area or to go directly to specific titles. DOAJ is maintained by Lund University in Sweden, and the service is financed by sponsors and members.

BASE—Bielefeld Academic Search Engine
base-search.net

BASE, sponsored by the Bielefeld University Library in Germany, covers material not readily found by commercial search engines. It seeks "intellectually selected resources" that meet academic quality standards and "web resources of the 'Deep Web,'" including more than 30 million documents in several languages.

Scitation
scitation.aip.org

Sponsored by the American Institute of Physics, the world's largest publisher of physics journals, this database offers

all things physics. Journals can be browsed by title, publisher, and subject category. Keyword and advanced searching is available. Browsing titles brings up listings, links, and availability of full text. Most items are open access, but anyone can purchase articles from the subscription journals.

3. Third Layer: Research tools for people engaged in very specialized fields

E-Print Network—Energy, Science, and Technology for the Research Community
osti.gov/eprints
The E-Print Network offers scientific and technology-related resources collected from more than 35,000 databases worldwide, including materials on basic and applied sciences, physics, chemistry, biology and life sciences, materials science, nuclear sciences and engineering, energy research, and computer and information technologies. Keyword search results include title, author, date, a summary, and source.

Plants Database
plants.usda.gov
Sponsored by the US Department of Agriculture, Natural Resources Conservation Service, this database covers anything to do with plants. A user can search under the common or scientific name of a plant, by characteristics, or by region.

Fold 3: The Web's Premier Collection of Original Military Records
fold3.com
This tool searches US military records, covering American conflicts from the Revolutionary War through the present, and offers photographs and digitized records.

FindSounds: Search the Web for Sounds
findsounds.com
FindSounds finds sound effects on the web. It offers searching in several languages, including English, German, French, and Chinese. Enter a textual description or approximation to produce a list of sources of the sound, which can be downloaded and listened to, along with information on file type and properties. (A search under "cat" found more than 200 cat sounds.)

Yovisto: Academic Video Search
yovisto.com
Yovisto is a video search engine specializing in educational video content, including online lectures. A search returns video screenshots and titles, duration, number of views, and other information, along with a link to each video and related subject tags that are in turn linked. Yovisto utilizes semantic search so that the "user has not only access to keyword-based search results but will also be guided by content-based associations to enable serendipitous discovery" (*Towards Exploratory Video Search Using Linked Data*, Waitelonis and Sack, 2011, p. 646).

FindThatFile: Finds What Nobody Else Does
findthatfile.com
FindThatFile claims to be the most extensive file search tool on the internet, covering 47 different file types. Google's advanced search offers only 10. A search can be conducted for all file types, or the user can select from documents, videos, audio files, fonts, software, and compressed file formats.

Making of America
quod.lib.umich.edu/m/moagrp
The Making of America Project has been a long-term effort to create a digital collection of primary documents related to American history. It is a collaborative endeavor among libraries, principally the University of Michigan Library and the Cornell University Library.

Social Science Research Network
ssrn.com
The Social Science Research Network is a worldwide collaborative sponsored by Social Science Electronic Publishing. The site, which supports dissemination of social science research, offers nearly half a million scholarly abstracts and nearly as many full-text papers. Its network covers subject areas such as accounting and other business fields, music, philosophy, literature, and politics. ∎

JANE DEVINE *(left) has been chief librarian and department chair for the LaGuardia Community College Library, part of City University of New York, since 2004. Before that, she served as LaGuardia's periodicals/government documents/electronic resources librarian and also worked for the New York Public Library as a reference librarian.* FRANCINE EGGER-SIDER *(right) has been coordinator of technical services at LaGuardia Community College since 1989. Previously, she worked at the French Institute/Alliance Française in New York City.*