

>appendix

Bibliographic Database Searches

Searching a Bibliographic Database

In a *bibliographic database*, each record is a bibliographic citation to a book or a journal article. In your university library, the online catalog is an example of a bibliographic database.

There are several bibliographic databases available to business researchers. Some of the more popular and comprehensive business bibliographic databases are:

- Proquest (from Proquest Information and Learning, formerly *ABI/Inform*).
- Business and Industry (from Gale Group).
- Business Source Complete (from EBSCO).
- Dow Jones Factiva.
- Lexis-Nexis Universe (from a division of Reed Elsevier).

Most of these databases offer numerous purchase options both in the amount and type of coverage. Some include abstracts, short summaries of the articles cited. Nearly all of these databases include the contents of around two-thirds of the indexed journals in full text, although the amount and the specific titles may vary widely from database to database. Full-text options vary from an exact image of the page to ASCII text only or text plus graphics. Search options also vary considerably from database to database. For these reasons, most libraries supporting business programs offer more than one business periodical database.

The process of searching bibliographic databases and retrieving results is basic to all databases.

1. Select a database appropriate to your topic.
2. Construct a **search query** (also called a *search statement*).
 - Review and evaluate search results.
 - Modify the search query, if necessary.
3. Save the valuable results of your search.
4. Retrieve articles not available in the database.
5. Supplement your results with information from Web sources.

Select a Database Most of us select the most convenient database without regard to its scope, but considering the database contents and its limitations and criteria for inclusion at the beginning of your search will probably save you time in the long run. Remember that a library's online catalog is a bibliographic database that will help identify books and perhaps other media on a topic. While journal or periodical titles are listed in a library's online catalog, periodical or journal articles are rarely included. Use books for older, more comprehensive information. Use periodical articles for more current information or for information on very specific topics. A librarian can suggest one or more appropriate databases for the topic you are researching.

Save Results of Search While the temptation to print may be overwhelming, remember that if you download your results, you can cut and paste quotations, tables, and other information into your proposal without rekeying. In either case, make sure you keep the bibliographic information for your footnotes and bibliography. Most databases offer the choice of marking the records and printing or downloading them all at once or printing them one by one.

Retrieve Articles For articles not available in full text online, retrieval will normally require the additional step of searching the library's online catalog (unless there is a link from the database to the catalog) to determine if the desired issue is available and where it is located. Many libraries offer a document delivery service for articles not available. Some current articles may be available on the Web or via a fee-based service.

World Wide Web Search Process

The World Wide Web is a vast information, business, and entertainment resource that would be difficult, if not foolish, to overlook. Millions of pages of data are publicly available, and the size of the Web doubles every few months.¹ But searching and retrieving reliable information on the Web is a great deal more problematic than searching a bibliographic database. There are no standard database fields, no carefully defined subject hierarchies

> **Exhibit Ba-1** Web Search Process Compared to Bibliographic Search Process

Bibliographic Search Process	Web Search Process
<ol style="list-style-type: none"> 1. Select a database appropriate to your topic. 2. Construct a search query. <ul style="list-style-type: none"> • Review and evaluate search results. • Modify the search query, if necessary. 3. Save those valuable results of your search. 4. Retrieve articles not available in the database. 5. Supplement your results with information from Web sources. 	<ol style="list-style-type: none"> 1. Select a search engine or directory.* 2. Determine your search options.* 3. Construct a search query.* <ul style="list-style-type: none"> • Review and evaluate search results. • Modify the search query, if necessary. 4. Save those valuable results of your search.* 5. Supplement your results with information from non-Web sources.*

*Denotes a variation.

(called *controlled vocabulary*), no cross-references, no synonyms, no selection criteria, and, in general, no rules. There are dozens of search engines and they all work differently, but how they work is not always easy to determine. Nonetheless, the convenience of the Web and the extraordinary amount of information to be found on it are compelling reasons for using it as an information source.

As you can see in **Exhibit Ba-1**, the basic steps to searching the Web are similar to those outlined for searching a bibliographic database. As you approach the Web, you start at the same point: focusing on your management question. Are you looking for a known item (e.g., IBM's website or that of the American Marketing Association)? Are you looking for information on a specific topic? If you are looking for a specific topic, what are its parameters? For example, if your topic is managed health care, are you hoping to find statistics, marketing ideas, public policy issues, accounting standards, or evidence of its impact on small business?

There are perfectly legitimate reasons to browse for information, and with its hypertext linking system, the Web is the ultimate resource for browsing. The trick is to browse and still stay focused on the topic at hand. In the browse mode you do not have any particular target. You follow hypertext links from site to site for the sheer joy of discovery. This is somewhat analogous to window shopping at the mall or browsing the bookshelves in the library. It may or may not be fruitful. And browsing is not likely to be efficient. Researchers often work on tight deadlines, because managers often cannot delay critical decisions. Therefore, researchers rarely have the luxury of undirected browsing.

Below we detail those steps in the Web search process that pose altered behavior for searches.

Select a Search Engine or Directory A search for specific information or for a specific site that will help you solve your management question requires a

great deal more skill and knowledge than browsing. Start by selecting one or more Internet search engines. Web search engines vary considerably in the following ways:

- The types of Internet sources they cover (http, telnet, Usenet, ftp, etc.).
- The way they search Web pages (every word? titles or headers only?).
- The number of pages they include in their indexes.
- The search and presentation options they offer.
- The frequency with which they are updated.

Furthermore, some information publicly indexable via the Web is not retrievable at all using current Web search engines. Among the material open to the public, but not indexed by search engines, are the following:²

- Pages that are proprietary (i.e., fee-based) and/or password-protected, including the contents of bibliographic and other databases. Some password-protected databases, such as the Thomas Register, are actually free and available to the public after initial registration.
- Pages accessible only through a search form (databases), including such highly popular Web resources as library catalogs, e-commerce catalogs (such as Lands' End, Amazon.com, and similar offerings), and the Security and Exchange Commission's EDGAR catalog of SEC filings. If you want to find the price of a book at Amazon.com, you first will have to find the Amazon.com page and then search the database for the title.
- Poorly designed framed pages.
- Some non-HTML or nonplain-text pages, especially PDF graphics files, for which no text alternative is offered. These pages cannot be retrieved using any current search engine.

- Pages excluded by the Robots Exclusion Standard (usually implemented with a robots.txt file). This standard is used by Web administrators to tell indexing robots that certain pages are off-limits. An outstanding example of this is the U.S. government's extensive information resource called GPO Access (described later).

The search engine, portal, or directory you select may well be determined by how comprehensive you want your results to be.³ If you want to use some major sites only, then start with a directory such as Yahoo! If, however, you are interested in gathering comments and opinions that are the focus of usenet groups, then use a more inclusive search engine such as Northern Light (<http://www.northernlight.com>). At least within the publicly indexable pages, one approach emphasizes selectivity, and the other, comprehensiveness. If you are interested in comprehensiveness, use more than one search engine. You are likely to yield very different and perhaps better results using additional search engines.

What is the difference between a search engine, a portal, and a directory? Directories rely on human intervention to select, index, and categorize Web contents. Subject directories build an index based on Web pages or web sites, but not on words within a page. Presenting a series of subject categories that are then further subdivided, Yahoo! (<http://www.yahoo.com/>) was the first Web subject directory and is still one of the most popular choices for finding information on the Web. The reason is that most users are satisfied with a few good sites rather than a long list of possibilities.

A search engine's different software components allow it to search and retrieve Web pages. These include:

- Software that automatically sends robots, sometimes called spiders, out to comb the Web, going from server to server to build an index of the words, pages, and files that are publicly indexable.
- Algorithms that determine how those pages will be selected and prioritized for display.
- User interface software that determines the search options available to the user.

Robots may be sent to roam the Web on a daily basis or on a six-week basis, so it is possible that some newer pages may not be included in the index developed by a particular search engine. Most search engines try for at least a monthly update of their indexes. Some robots may search only the upper-level pages and totally ignore valuable pages buried within a site.

The algorithm used by the search engine can have an enormous impact on the type and quantity of information retrieved. Search algorithms determine whether every word is to be included or only the top 50 or so words,

whether more weight will be given to words in metadata or titles or in words used frequently, and so on. The possibilities are limitless and are the major reason that results from one search may vary considerably from those of another search.

A **portal** is, as the name suggests, a gateway to the Web. A portal often includes a directory, a search engine, and other user features such as news and weather. Most Internet service providers (ISPs) are portals to the Web. The AOL home page is an example. This portal uses information based on past user search behavior to determine what to offer on the opening screen. Therefore, some valuable search engines, indexes, directories, and more may be relegated to an "other search aids" category. If as a researcher your behavior differs from the majority pattern, you have to be more knowledgeable about search strategies to bypass the frontline strategies offered by the portal. Several ISP portals now offer subscribers the option of customizing the portal with user-chosen search engines and secondary sources. Most of the major search engines are now actually portals to the Web that include their own search engine. Specialized portals are increasingly popular. One such portal is the Annual Report Service, a collection of links to a large number of company annual reports (www.annualreportservice.com).

Determine Search Options Nearly all search services have a Help button that will lead you to information about the search protocols and options of that particular search engine. How does the search engine work? Can you combine terms using Boolean operators (AND, OR, NOT) or other connectors? How do you enter phrases? truncate terms? determine output display? limit by date or other characteristic? Some search engines provide a basic and an advanced search option. How do they differ?

Construct a Search Query and Enter Search Term(s) The Web is not a database, nor does it have a controlled vocabulary. Therefore, you must be as specific as possible, using the keywords in your management question and any variations you can think of. It is up to you to determine synonyms, variant spellings, and broader or narrower terms that will help you retrieve the information you need. This may involve some trial and error. For instance, a general term (such as *business*) would be useless in a search engine that purports to index every word in every document.

Save Results of Search If you have found good information, you will want to keep it for future reference so that you can cite it in your proposal or refer to it later in the development of your investigative questions. If you do not keep documents, you may have to reconstruct your

search. At a future time, given that some portion of the Web is revised and updated daily, those same documents may no longer be available.

Supplement Results with Information from Non-Web Sources There is still a great deal of information in books, journals, and other print sources that is not available on the Web. Although many novice researchers start and end here, the more sophisticated researcher knows a Web search is just one of many important options.

Searching for Specific Types of Information on the Web

Once you have defined your topic and established your search terms, you need to determine whether you are looking for a specific site (known item) or an address of a person or institution (who), a geographic place name and location (where), or a topic (what).

Known-Item Searches In the same way that search protocols for the library's online catalog vary between a known-item (author or exact title) and a more general keyword search, the way you query the Web for an exact item also varies from a more general query. A trend among search engines is to establish algorithms that will yield more precise results. One of the first to follow this trend was the search engine Google (<http://www.google.com/>), which debuted in 1998. Google and others like it help you retrieve the most precise results from known-item searches by creating an algorithm that interprets a link to a site as a vote for that site. The sites that receive the most votes (links) rise to the top of the results list in a known-item search. The Google system also emphasizes the importance of the linking page in its algorithm.

Who Searches In the *who* searches, you are looking for an e-mail address, a phone number, a street address, or a Web address of a person or institution. For this type of information, you will first need to identify a database containing the information you need and then search that database according to its search protocols. At this writing, almost all Web search engines and portal sites partner with infoUSA (<http://www.infoUSA.com/>) to supply the phone number databases for their white and yellow page services.⁴

Where Searches A *where* option comprises the mapping services that help you locate an address on a map or discover the route from one place to another. Mapping services are databases tied to geographic information systems (GISs). One popular site is MapQuest (<http://www.mapquest.com>).

What Searches As we have already noted, search engines vary considerably in the way they work and in their size. If you are searching for a very unusual term, select one of the more comprehensive search engines, such as Northern Light (<http://www.northernlight.com/>) or AltaVista (<http://altavista.com/>). Generally, it is more efficient to start with a directory such as Yahoo! or one of the more specialized directories on the Web. Some especially good specialized sites are the Argus Clearinghouse (<http://www.clearinghouse.net/>), featuring subject guides on dozens of topics prepared by librarians and other specialists, and INFOMINE Scholarly Internet Resource Collections from the University of California housed at University of California–Riverside (<http://infomine.ucr.edu/>). See **Exhibit Ba-2** for a selection of business-related websites.

Since the Web was introduced to the world in 1992, Web technology has been seeking ways to make the contents more accessible. The dynamic nature of the Web, its lightning growth rate, the ephemeral nature of some Web pages, the different skill levels and interests of users, and the lack of standards make this an enormous challenge. Trends indicate that the Web will continue to grow and that we will continue to apply new technologies to tapping the information available on this vast and unique resource.⁵ Already, some search engines are better able to identify key resources. Efforts are under way to adopt standardized metatags in the coding to describe the contents of Web pages. Some search engines are using expert systems to learn more about the information requester. This is already being used extensively to target advertisements and is being used more frequently to “select” from among several options the information source that will be delivered to the requester. More efforts are being made to index a larger portion of the Web content; at the same time, efforts are also under way to improve the relevancy of the results delivered for any one search.

Government Information

Government publications, especially those of the U.S. government, are mandatory resources for many business research projects. The agencies of the U.S. government, considered as a whole, are the largest publishing body in the world. The government collects and provides access to a wide variety of social, economic, and demographic data. In addition, government laws and regulations, court decisions, policy papers, and studies all have a potential impact on business. Additionally, the government provides directories, maps, and other information sources. Specialists are available throughout the government to provide individual assistance.

Searching for government information is a complicated task that usually requires some knowledge of how

Exhibit Ba-2 Selected Websites for General Business Research

SITE ID	SITE URL	Sponsor
Annual Reports Online	www.zpub.com/sf/ar/	Annual Reports Library, San Francisco
BizLink, Your On-line Business Resource	www.plcmc.org/bizlink/	Public Library of Charlotte and Mecklenberg County
Business and Economics Resources Online	http://lcweb.loc.gov/rr/business/beonline/	Library of Congress
ENTERWeb (Enterprise Development Website)	www.enterweb.org/ Went offline in late 2018	ENTERWeb
Guide to Labor-Oriented Internet Resources	www.irle.berkeley.edu/library/	Institute of Industrial Relations Library, University of California-Berkeley
DnB Hoover's	www.hoovers.com	Hoover's, Inc.
MSU-CIBER International Business Resources on the WWW	http://globaledge.msu.edu/	Michigan State University Center for International Business Education and Research
Rutgers Accounting Web	http://accounting.rutgers.edu/	Accounting Research Center, Rutgers University
Tax and Accounting Sites Directory	www.taxesites.com/ Restricted access 2018	University of NC- CH
Thomas Register of American Manufacturers	www.thomasnet.com	Thomas Publishing Co.
Working Knowledge	https://hbswk.hbs.edu/	Harvard Business School Business Research for Business Leaders

government functions. In recent years, the U.S. government has been working aggressively to make its information available, not only through the Depository Library Program but also through the development of electronic resources. As a result, information that used to be tucked into the farthest corners of the library is now readily available and searchable on the Web. In many libraries, the entire government documents collection is included in the library's catalog, with links to Internet resources. In the remainder of this appendix, we examine three of the most useful government information types. A list of selected government resources on the Web is included in **Exhibit Ba-3**.

Government Organization Two of the most useful resources regarding government organization are:

- U.S. Government Manual (published annually); <http://www.gpoaccess.gov/gmanual/> (updated annually).
- Congressional Directory (published annually); <http://www.gpoaccess.gov/cdirectory/> (updated regularly).

The U.S. Government Manual describes the functions of every government agency and is particularly useful for identifying key personnel and agency contacts,

Exhibit Ba-3 Selected Government Sources

Source	Internet
Economic Indicators	https://www.census.gov/economic-indicators/
US by the numbers	https://usafacts.org/
EDGAR Database of Corporate Information (SEC filings)	https://www.sec.gov/edgar.shtml
SAM.gov(Federal Business Opportunities) (a listing of government procurement, sales, and contract awards)	https://beta/sa.gov
USA.gov	https://www.usa.gov/statistics
Small Business	www.usa.gov/Business/
U.S. Government Information	https://www.govinfo.gov/
U.S. Congress on the Internet	www.congress.gov
U.S. Bureau of Economic Analysis	www.bea.gov
U.S. Bureau of Labor Statistics	http://stats.bls.gov
U.S. Census Bureau	www.census.gov
U.S. Department of Commerce	www.commerce.gov
U.S. Department of Labor	www.dol.gov
U.S. Patent and Trademark Office	www.uspto.gov
U.S. Small Business Administration	www.sba.gov/

including those at the local and regional levels. Generally very knowledgeable and helpful, these people are invaluable in any research project for their ability to cut through red tape and to answer questions pertinent to their agencies.

The *Congressional Directory* lists members of Congress and congressional committees, as well as key personnel throughout the U.S. government.

Laws, Regulations, and Court Decisions Government information regarding these key areas and other legal information can be obtained by consulting GPO Access (<http://www.gpoaccess.gov/>). GPO Access is the government's official and real-time site for finding government information. Included at this site is the complete *Monthly Catalog of U.S. Government Publications*. Use it to identify the full range of government publications, printed and electronic, issued in or after 1994.

Especially valuable on GPO Access are the databases covering laws, regulations, and congressional debates and publications. The key databases for laws and regulations are:

- *Congressional Bills* —provide texts of varying versions of bills. Only a small portion of this collection ever becomes law, but the topics can reveal trends of interest to business researchers.

- *Public Laws*—provide the texts of laws as they are passed; covers 1994 to the present. The printed version of this source is called *U.S. Statutes at Large*.
- *U.S. Code* —a codification of laws currently in effect and as revised over time. It is also available in libraries as a printed document.
- *Federal Register* —“the official daily publication for rules, proposed rules, and notices of federal agencies and organizations, as well as executive orders and other presidential documents;”⁶ covers 1995 to the present. It is also available in libraries as a printed or microfiche document.
- *Code of Federal Regulations (CFR)*—“a codification of the general and permanent rules originally published in the *Federal Register* by the executive departments and agencies of the federal government.”⁷ A printed version is available in libraries.

GPO Access includes other valuable databases, including the Supreme Court decisions from 1937 to the present, the *Economic Report of the President*, the U.S. Budget,

Commerce Business Daily (CBDNet), and GAO reports. New databases are added regularly.

Many libraries have created local gateways to GPO Access that help speed information retrieval (<http://www.gpoaccess.gov/>). GPO Access offers dozens of fields to search, and they can be searched independently or together. Searching each database independently provides more flexibility and more precise searching options because some fields are unique to a particular database. In general, search options are similar to those for other databases. Use the search hints for each file for more details and special search possibilities.

Government Statistics Information regarding government statistics may be obtained by consulting the following sources:

- *Statistical Abstract of the United States* (<http://www.census.gov/prod/www/abs/statab.html>).
- FedStats (<http://www.fedstats.gov>).
- U.S. Bureau of the Census (<http://www.census.gov/>).

The government collects statistics on just about every topic imaginable—from crimes to hospital beds, from teachers to tax revenue, from steel production to flower imports. For any statistical inquiry, start with the *Statistical Abstract of the United States*. This annual compendium compiles statistics issued by nearly every government agency as well as additional data from selected nongovernment organizations. Many are time-series tables covering several years or even decades. All tables indicate the source of the statistics. These sources can then be consulted if desired for even more comprehensive data. Check the library's catalog or ask the librarian to find these more specialized resources. Some may be available via FedStats.

FedStats is an online compilation of statistics provided by more than 70 U.S. agencies, including the Census Bureau, the Bureau of Labor Statistics, and the Federal Bureau of Investigation. Use the search option or the directory option to find the needed statistical tables. An especially useful feature of FedStats is the state and regional statistical data option.

No discussion of government statistical information would be complete without an examination of the U.S. Bureau of the Census. The Census Bureau is probably most well known for the Decennial Census of Population and Housing. The first such census was taken in 1790 to meet the constitutional mandate for apportioning seats in

Congress. It has been taken in every year ending in zero since that date. Now it is used not only to apportion seats in the House of Representatives but also for allocation of federal aid to states and for a myriad of other purposes. The decennial census asks a certain core of questions of everyone. These are known as the "100 percent questions." While these questions may vary slightly from census to census, data on age, race, gender, relationship, and Hispanic origin are fairly constantly collected. A longer questionnaire is sent to a sample of the population. Data from its additional questions, used in conjunction with the data from the 100 percent questions, are used by government agencies at all levels; local planners; business and industry; schools; and social service agencies, among others, for planning; grant writing; economic development; and many other purposes.

To make census information easier to understand, the Census Bureau, in cooperation with local planning agencies, has created a multilevel mapping system. The entire country is mapped into small units called *blocks*. Data from the 100 percent questions are available for all blocks, but sample data are not. Both 100 percent–question data and sample data are available for census tracts (groups of blocks) and for larger mapped units such as cities, metropolitan statistical areas, counties, and states. Tracts are especially valuable to local-level researchers because their boundaries remain mostly constant from census to census, thus allowing comparison. In cases where there is population growth, tracts may be split from one census to another and therefore may need to be added together to achieve comparable statistics. Metropolitan statistical areas, defined by the Office of Management and Budget, consist of a large population nucleus together with adjacent communities having a high degree of social and economic integration with that core. Metropolitan statistical areas comprise one or more entire counties, except in New England, where cities and towns are the basic geographic units.

In addition to the decennial census, the Census Bureau conducts the economic census in years ending in two and seven, covering all areas of the economy from the national to the local level. Both the decennial census and the economic census are supplemented by numerous survey reports, including the new American Community Survey, initiated to provide more up-to-date information on American communities. In fact, the Census Bureau has proposed using the American Community Survey—instead of the long (sample) questionnaire used through 2000—in the next decennial census. For an overview of the many report topics available, see the "Subjects A–Z" listing on the Census Bureau website (<http://www.census.gov/>).

Reference Notes

1. Good sources for Web size estimates are the studies by Steve R. Lawrence and C. Lee Giles, “Searching the World Wide Web,” *Science* 280 (April 1998, pp.98–100, and “Accessibility of Information on the Web,” *Nature* 400 (July 8, 1999), pp. 107–109, with updated summary data at www.wwwmetrics.com.
2. Michael Dahm, “Counting Angels on a Pinhead: Critically Interpreting Web Size Estimates,” *Online* 24 (January-February 2000), pp.35–44. This article further interprets the pioneering research by authors Steve R. Lawrence and C. Lee Giles (op cit.).
3. The May-June 1999 issue of *Online* focuses on search engine technology. See, for example, Danny Sullivan, “Crawling under the Hood: An Update on Search Engine Technology,” *Online* 23 (May-June 1999), pp. 30–38. See also Danny Sullivan’s “Search Engine Watch” (<http://searchenginewatch.com/>) and Greg Notess’s “Search Engine Showdown” (www.notess.com/search/) for current information about search engines and their features.
4. Greg R. Notess, “Duplicative Databases: Yellow Pages from infoUSA,” *Database* 22 (February-March 1999), pp. 73–36.
5. See for example, the June 22, 1999 issue of *PC Magazine* for a special report detailing “Ten Trends That Are Defining the Future” (Introduction. p. 100).
6. “Federal Register Online via GPO Access,” http://www.access.gpo.gov/su_docs/aces/aces140.html.
7. “About the Code of Federal Regulations,” <http://www.access.gpo.gov/nara/about-ctr.html#page1>.

NOTE: All URLs were active as of March 13, 2021.