


Searching the Web

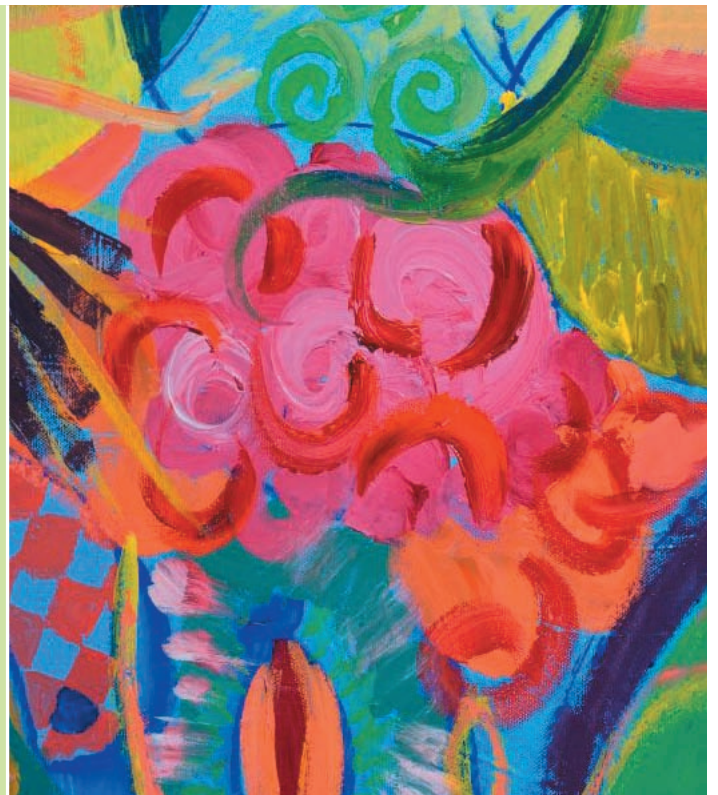
Files You Will Need:

No files needed.

You can use the Web to access millions of Web pages, which contain information on a virtually unlimited number of topics. To find the information you want among all these Web pages, you need to learn searching methods and tools.  Nancy Shand and Ranjit Singh, staff writers at the *Georgetown Journal*, a top-rated daily newspaper that serves the Georgetown metropolitan area, have hired you as their assistant. They want you to use the Web to help them gather information for their stories.


OBJECTIVES

- Define search engines
- Develop a Web search strategy
- Use a search engine
- Use a Web directory
- Use a metasearch engine
- Use a subject guide
- Understand advanced searching techniques
- Conduct a search using the Advanced Search page
- Conduct a search using Boolean operators



**UNIT
D**
Internet

Defining Search Engines

A **search engine** is a Web site (or part of a Web site) that finds Web pages containing the word or phrase you specify and displays the pages as a list of links. For example, you could enter the word “Louisiana” into the appropriate location in a search engine, and then click the Search button to get a list of Web pages that might contain information about Louisiana. One of the most popular search engines is Google, but there are many others available.  Before you accept your first research assignment from Ranjit and Nancy, you decide to learn some of the terms associated with Web searching and search engines.

DETAILS

The following terms are associated with searching the Web:

- **Search Expression and Keyword**
A **search expression** or **query** is the words or phrases you enter when you conduct a search. A search expression can be composed of one or several words; each word in a search expression is called a **keyword**.
- **Search Engine Database**
A search engine does not search the Web to find a match to the search expression you enter; it searches only its own database of Web content that it has catalogued. Each search engine database indexes the information it collects from the Web differently. Some search engines store the entire content of every Web page they index; other search engines collect information only from a Web page’s title, description, keywords, or HTML tags; and others read only a certain amount of the text in each Web page. If the terms you use in your search expression are not in the part of the Web page that the search engine stores in its database, the search engine will not return that page. Therefore, if you enter the same search expression into different search engines, you will get some results that are the same and some that are different because each search engine contains a different set of information in its database and each search engine uses different procedures to search its database.
- **Web Robot**
Most search engines use a Web robot to build their database. A **Web robot**, also called a **bot** or a **spider**, is a program that automatically searches the Web to find new Web sites and updates information about Web sites that are already in the database. A Web robot also deletes information in the database when a Web site no longer exists.
- **Hit**
A **hit** is a Web page that is indexed in the search engine’s database and contains text that matches your search expression.
- **Results Pages**
All search engines provide a series of **results pages**, which include links to Web pages that match your search expression. Figure D-1 shows a results page using the AlltheWeb search engine for the search expression *modern art*; Figure D-2 shows search results for the same search expression in the Lycos search engine. Notice that suggested related search expressions are listed on the right. The Bing search engine does something similar by listing categories that you can click to narrow the search.
- **Narrowing the Search Results**
Often, a search will yield millions of hits. To narrow your search results, create a search expression that contains more than one keyword; most search engines will search for pages that contain all of the words in your search expression, although in no particular order. If you want to search for pages that contain a specific phrase, enclose the phrase in quotation marks in the search expression. For example, a search for pages that contain the keyword *rice* returns over 96 million hits in one search engine, and the pages cover topics ranging from a page for Rice University to a page about someone whose last name is Rice to pages having something to do with rice, the grain. But a search using the search expression *brown rice* decreases the number of hits to about 21 million; and a search with “*brown rice*” (enclosed in quotation marks) decreases the number of hits to just under 2 million. Finally, if you add the keyword *recipe* so that the complete search expression is “*brown rice*” *recipe*, the number of hits decreases again to about 1 million, and all of the hits are about recipes with brown rice. You can refine the search further by adding additional ingredients to the search expression. You get the best results by using keywords that do not have multiple meanings and are not articles or prepositions.

QUICK TIP

Even the best search engine sites cannot keep their databases completely updated, so you will occasionally click a link in the list of results only to find that the page to which it is linked no longer exists.

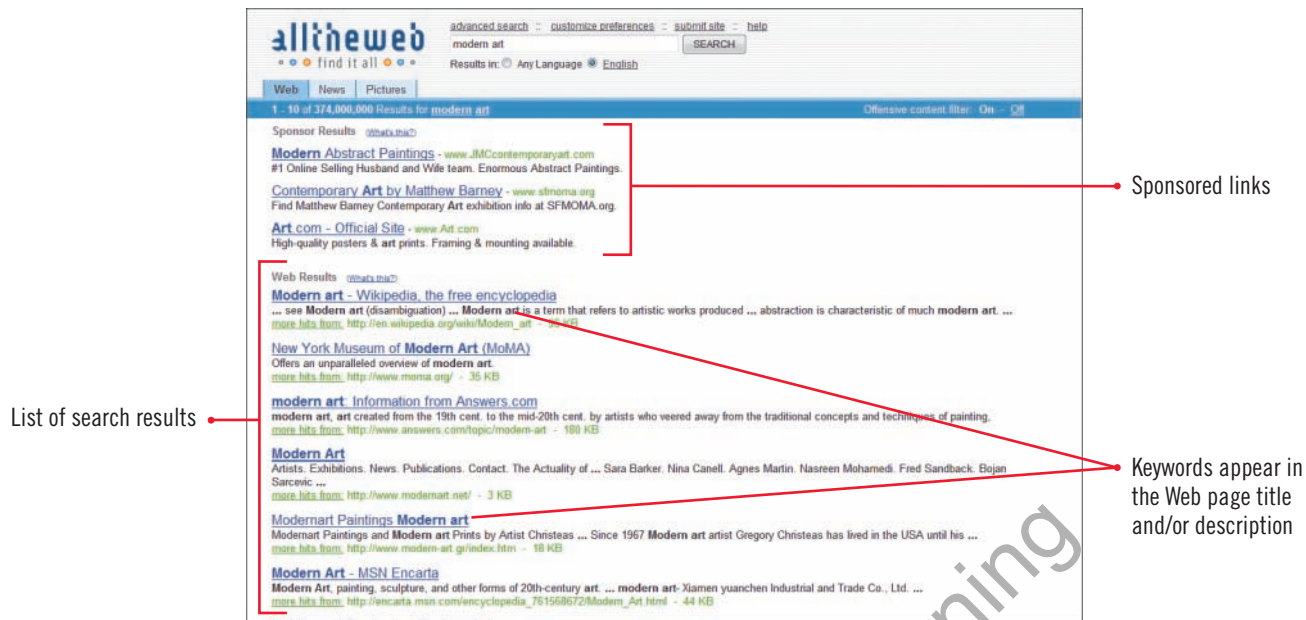
QUICK TIP

Many search engines include information about their search engines, robots, and databases on their Help or About pages.

QUICK TIP

Most major search engines search for variants of keywords automatically; for example, if you search using the keywords *Canada travel guide*, most search engines will return hits that include the keywords “Canadian” and “guides” as well.

FIGURE D-1: Results page for a search on keywords “modern” and “art” in the AlltheWeb search engine



Courtesy of Yahoo! Inc. © 2009 by Yahoo! Inc. YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc.

FIGURE D-2: Results page for a search on keywords “modern” and “art” in the Lycos search engine

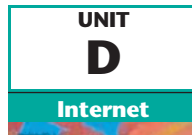


How search engines are financed


The organizations that operate search engines often sell advertising space on the search engine Web site and on the results pages. Many search engine operators also sell paid placement links on results pages. For example, a car manufacturer might want to purchase rights to the keyword *car*. When you enter a search expression that includes the keyword *car*, the search engine creates a results page that has a link to that company's Web site at or near the top of the results page. Most, but not all, search engines label these paid placement links as **sponsored links**. If the advertising appears in a box on

the page (usually at the top, but sometimes along the side or bottom of the page), it is called a **banner ad**. Search engine Web sites use the advertising revenue to cover the costs of maintaining the computer hardware and software required to search the Web and create and search the database, and to generate profit. The only price you pay for access to these tools is that you see advertising banners on many of the results pages, and you might have to scroll through some sponsored links at the top of results pages; otherwise, your usage is free.

Internet



Developing a Web Search Strategy

You can use the Web to quickly find answers to specific questions or as a resource to explore interesting concepts and ideas. Each of these question types, specific and exploratory, requires a different search strategy.  Before you start accepting research requests from Nancy and Ranjit, you decide to familiarize yourself with a searching strategy.

DETAILS

Consider the following as you develop a search strategy:

- **Specific Question**

A **specific question** is a question that you can phrase easily and has only one answer. Specific questions might require you to start with broad categories of information and then gradually narrow the search until you find the answer to your question. Figure D-4 shows this process of sequential, increasingly focused questioning. As you narrow your search, you might find results that do not lead you to the answer of your question. If that happens, you need to choose the result (or path) that will lead you to the correct answer, as shown in Figure D-4.

- **Exploratory Question**

An **exploratory question** starts with a general, open-ended question that leads to other, less general questions, which result in multiple answers. The answers to the questions at each level should lead you to more information about the topic you are researching. This information then leads you to more questions and answers. Figure D-5 shows how this questioning process broadens the scope of results as you gather information pertinent to the exploratory question.

- **Conducting a Search**

Once you determine the type of question you need to use in your search, you can begin the actual Web search process. After you carefully formulate and state your question, you select the appropriate tool or tools to use in your search, and then conduct the search.

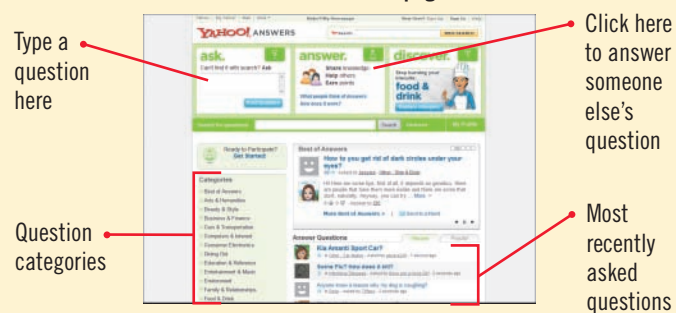
- **Evaluating the Results**

After obtaining your results from a Web search tool, you need to evaluate these results to determine if they answer your question. If they do not, you continue the search by refining or redefining your question and then selecting a different search tool to see if you get a different result. The determination of when your search process is completed is different for each type of question. For specific questions, you repeat this process as many times as necessary until you obtain the specific answer you seek; for exploratory questions, you repeat it until you have found a satisfactory range of information regarding your topic.

Asking for help from other people

Some search engines include a feature in which you can ask other people for the answer to a question. After you type your question and send it to the search engine, the question becomes available for anyone using the site to see. If people know the answer, they might reply and you will see the answer listed below your question. You need to remember, however, that not everyone who answers a question is an expert, and the answer might very well be wrong. On most of these sites, you need to sign up to be a registered user and then identify yourself as a user in order to ask or respond to a question. The Yahoo! Answers page is shown in Figure D-3.

FIGURE D-3: Yahoo! Answers page



Courtesy of Yahoo! Inc. © 2009 by Yahoo! Inc. YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc.

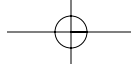


FIGURE D-4: Specific research question search process

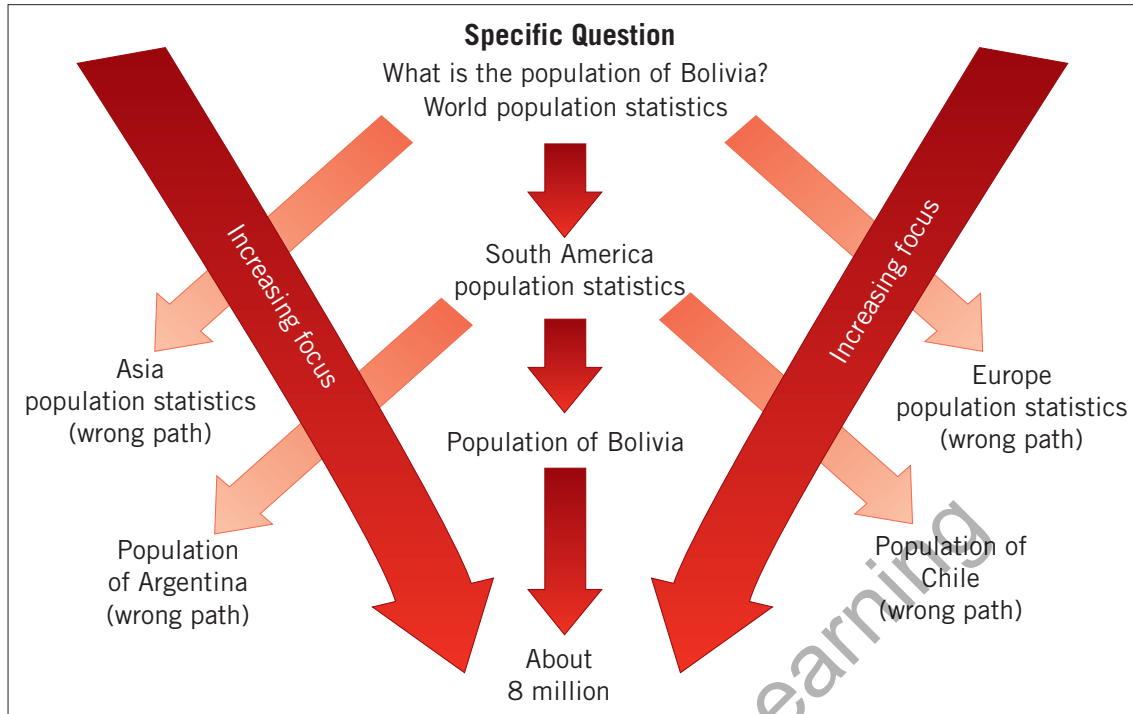
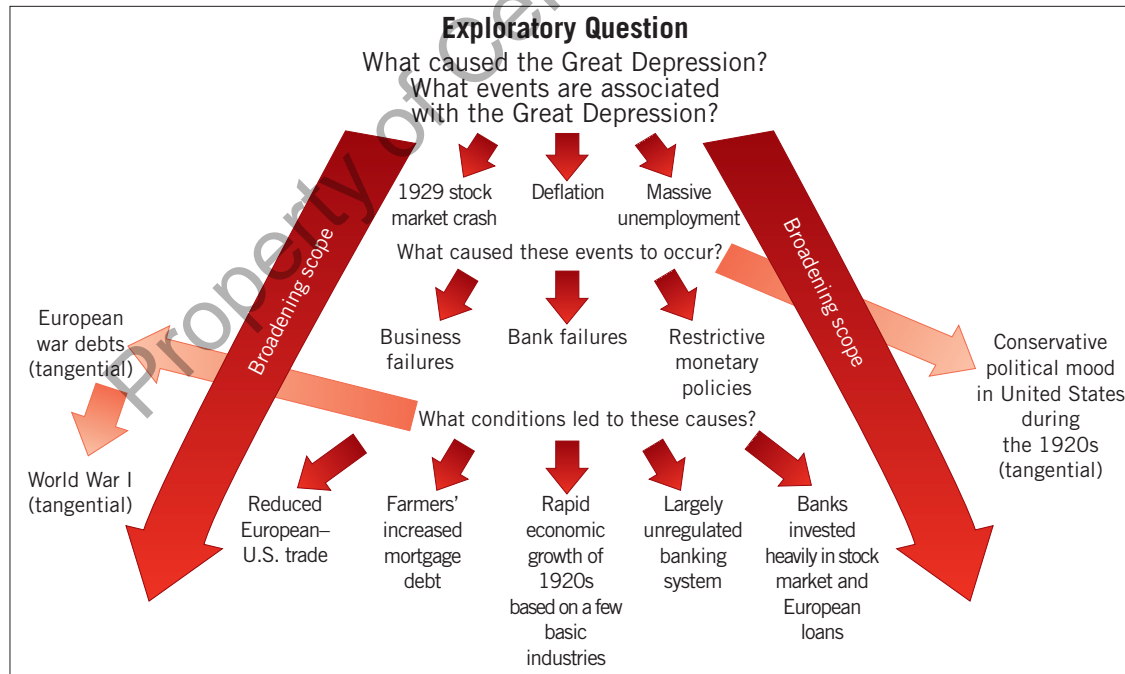
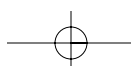


FIGURE D-5: Exploratory research question search process




Internet



**UNIT
D**
Internet

Using a Search Engine

No one knows how many Web pages exist on the Web, but the number is now in the billions. Each of these Web pages might contain thousands of words, images, or links to downloadable files. Unlike the content of a library, the content of the Web is not indexed in any standardized way. Fortunately, you can use search engines to help you find the information you need.  Nancy needs to know the average temperature on Mars for a story that she is writing. This search question is a specific question, not an exploratory question, because you are looking for one specific answer—the average temperature on Mars.

STEPS

1. **Start your Web browser, go to the Online Companion page for Unit D, then click one of the links to a search engine under Lesson 3**

Your browser opens the home page of the search engine you chose.

2. **Type *Mars* in the search expression text box**

Figure D-6 shows *Mars*, the search expression, entered in the Search text box on the home page of the Google search engine. Notice the drop-down list below the Search box. Several search engines provide suggestions in a drop-down list in this manner to help you formulate your search expression. If you see your search expression in the list, you can click it to accept that suggestion.

3. **Press the *Spacebar*, then type *average temperature***
4. **Click the appropriate button to start the search (this button is usually labeled “Search,” “Go,” or “Find”)**

The search results appear on a new results page. The results page indicates that there are well over a million Web pages that might contain the answer to your query. Figure D-7 shows the results page in the Bing search engine.

5. **Scroll down the results page, examine your search results, then click links until you find a Web page that provides the average temperature of Mars**

The average temperature on Mars is -63° Celsius, -81° Fahrenheit. (Some Web sites might report a slightly different number.)

6. **Return to the Online Companion page for Unit D, then click a link to another of the search engines listed under Lesson 3**

The home page of the search engine you chose opens.

7. **Type *Mars average temperature* in the search expression text box, then click the appropriate button to start the search**

The search results appear on a new results page.

8. **Scroll down the results page, examine your search results, then click links until you find the average temperature of Mars**

Once again, you should find that the average temperature on Mars is -63° Celsius, -81° Fahrenheit. Your specific search was successful.

TROUBLE

If the results page contains a message that the search engine could not find any results that matched your search expression, return to the Online Companion, choose a different search engine, then repeat Steps 2 through 4.

TROUBLE

If you do not find any useful links on the first page of search results, click the numbers at the bottom of the results page to open additional results pages.

QUICK TIP

Both Internet Explorer and Firefox allow you to access a search engine from a search box on the Address bar.

FIGURE D-6: Search expression entered in the Google search engine

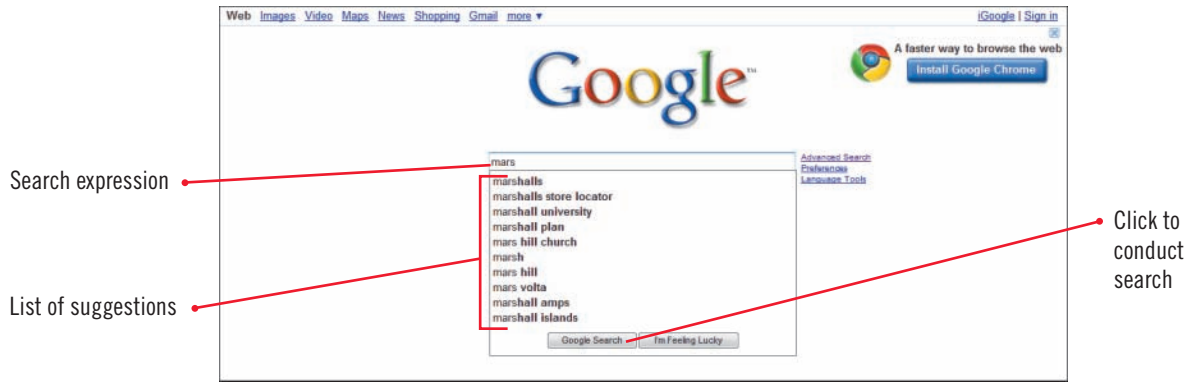
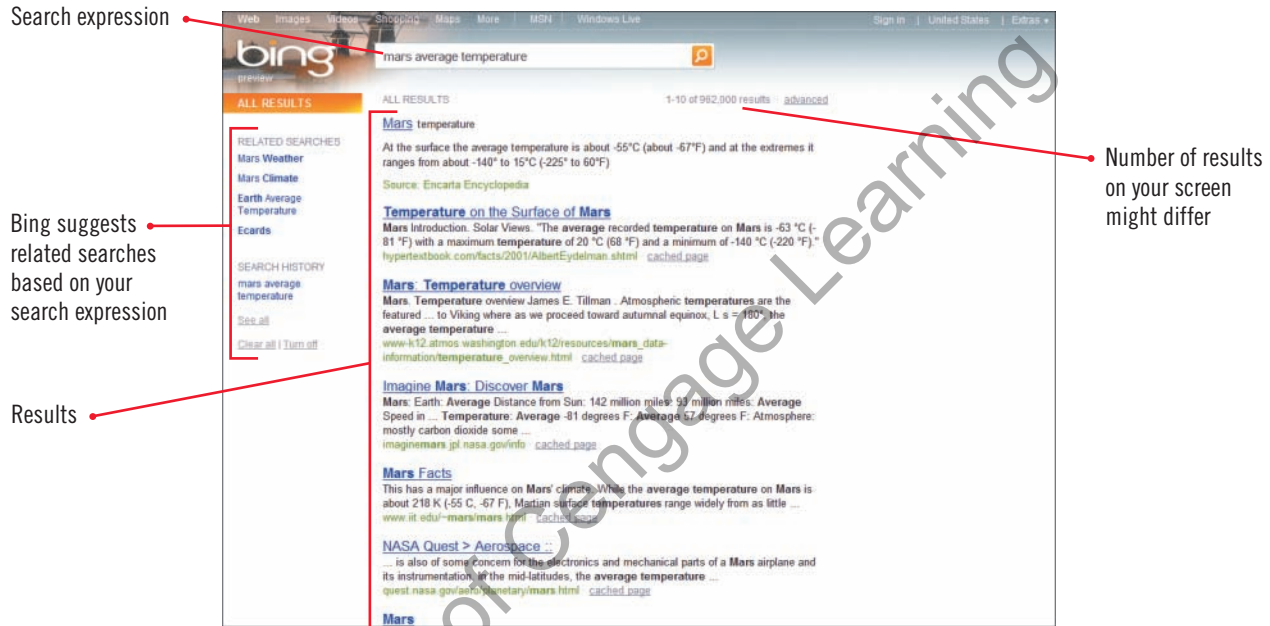


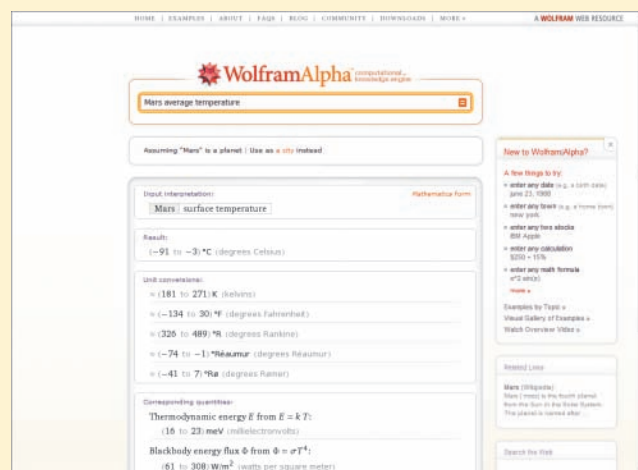
FIGURE D-7: Results page for “Mars average temperature” in the Bing search engine



Using natural language queries

Some search engines have attempted to include natural language querying. A **natural language query** allows users to enter a question exactly as they would ask a person that question. The search engine analyzes the question using knowledge it has been given about the grammatical structure of questions and then uses that knowledge to convert the natural language question into a search query. Although no major search engine has been able to make a natural language query interface that has worked consistently, mathematical software company Wolfram has a Web site that offers a natural language interface to a database of collected facts. This site, Wolfram Alpha, is a computational engine and performs calculations using the information it extracts from its database. It lets users ask questions in natural language that relate to the facts in its database. Figure D-8 shows the results of the search using *Mars average temperature* as the search expression. The results page includes an interpretation of the search expression and returns the temperature in various units, including degrees Celsius, Fahrenheit, and kelvin. It also includes several scientific calculations using the search results.

FIGURE D-8: Wolfram Alpha natural language interface results



Internet

**UNIT
D**
Internet

Using a Web Directory

A **Web directory** is a list of links to Web pages that is organized into hierarchical categories. Web directories and search engines both use a database of links to Web pages to enable users to search for information in different ways. Rather than using a database compiled by Web bots, however, a Web directory uses human editors or computers to decide which Web pages will be included in the directory and in which category to store the link to selected Web pages. Users can browse for information by general categories, rather than by using specific search terms. Many Web directories are one feature of larger search engines, and you need to click a link on the search engine's home page to view the directory. Ranjit wants to know the latest news and information about his profession. He asks you to provide him with a set of links to Web sites about the media industry. You use a directory to identify Web sites based on the category of information that Ranjit needs.

STEPS

1. **Go to the Online Companion page for Unit D, then click one of the links to a Web directory under Lesson 4**

The directory you selected opens. The home page contains a list of categories into which the directory links are organized. The Gigablast directory is shown in Figure D-9.

QUICK TIP

A category that contains the term "Media," "Professions," "Jobs," or "Reference" would be a good choice.

2. **Examine the categories on the directory's home page and click a link that is likely to contain information about the journalism profession**

The page you selected opens, showing links to lower levels in the hierarchy, links in the hierarchies of other categories, and, usually, some sponsored links.

3. **Examine the page that loads in your browser and search for links to subcategories that include words such as "industry news" or "industry updates"**

You might need to search several levels down in the directory to find the information you are looking for. You also might need to try a different directory.

QUICK TIP

In some Web directories, you can enter a search expression and search within a category.

4. **When you find a link to a subcategory that seems like it contains information about the media industry, click it.**

The Yahoo! Web directory, after following the News & Media/Industry Information links, is shown in Figure D-10.

5. **Examine the links on the page that opens in your browser, then, if the category does not seem narrow enough or the links are not appropriate, click additional subcategories until you see links that might contain the information you are searching for**

6. **Click one of the links listed in the subcategory and examine the Web page**

7. **Navigate back to the subcategory page, click a different subcategory link, then examine the Web page**

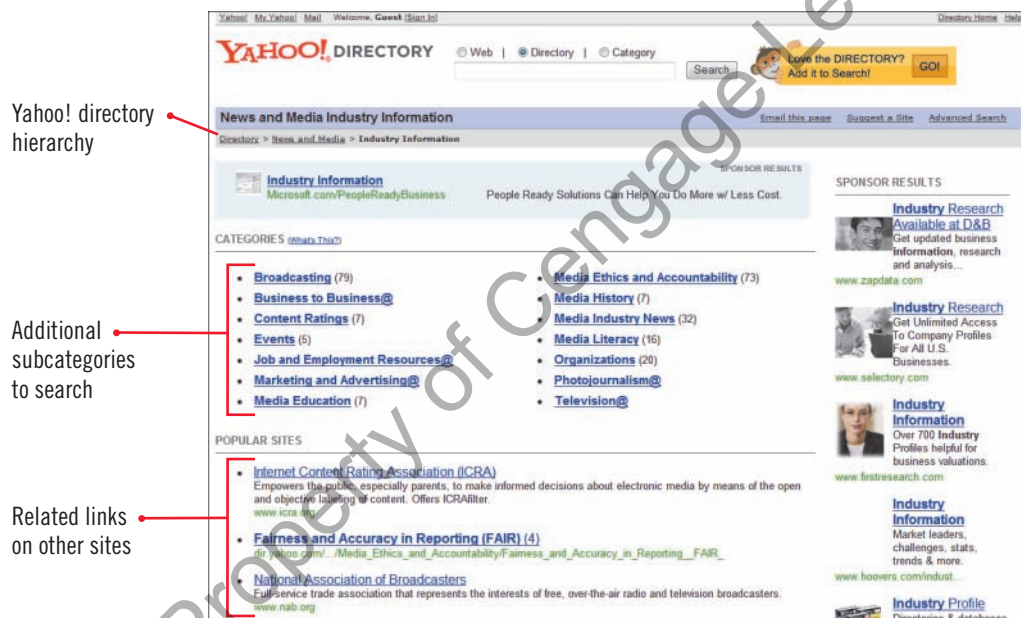
QUICK TIP

Web directories that are part of Web sites that also include search engines are called **hybrid search engines**.

FIGURE D-9: Gigablast directory categories



FIGURE D-10: Yahoo! News and Media Industry Information directory page



Courtesy of Yahoo! Inc. © 2009 by Yahoo! Inc. YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc.

Examining About.com and the Open Directory Project


About.com hires content experts to create and manage its Web directory entries, and also identifies its experts for you. Each of the About.com experts, called Guides, hosts a page with hyperlinks to related Web pages, moderates discussion areas, and provides an online newsletter. This creates a community of interested persons from around the world that can participate in maintaining the Web directory. The Open Directory Project uses the services of more than 40,000 volunteer editors who maintain listings in their individual areas of interest.

The Open Directory Project offers the information in its Web directory to other Web directories and search engines at no charge. Many of the major Web directory, search engine, and metasearch engine sites regularly download and store the Open Directory Project's information in their databases. For example, AlltheWeb, AltaVista, Dogpile, and Google all include Open Directory Project information in their databases.

Internet

**UNIT
D****Internet**

Using a Metasearch Engine

A **metasearch engine** is a tool that uses multiple search engines. Using a metasearch engine, you can search several search engines simultaneously, so you do not have to conduct the same search in different search engines. Most metasearch engines forward your queries to a number of major search engines and directories. After a metasearch engine sends your search expression to several search engines, the search engines compare the search expression against their databases of Web page information and return results to the results page of the metasearch engine for you to view. Some metasearch engines identify the search engine from which they retrieve the links; others do not.  You want to learn how to use metasearch engines so that you can access information more quickly. You decide to use Nancy's question about the average temperature on Mars to test a metasearch engine.

STEPS

1. **Go to the Online Companion page for Unit D, then click one of the links under Lesson 5**
The home page of the metasearch engine you chose opens.

QUICK TIP

Mamma.com was one of the first metasearch engines on the Web.

2. **Type *Mars average temperature* in the search expression text box**
3. **Click the appropriate button to start the search**
A results page appears showing the hits for each search engine that the metasearch engine searched. If you used the Dogpile metasearch engine, you might see results similar to the results page shown in Figure D-11.
4. **Examine your search results, then click appropriate links to find the average temperature on Mars**
As you scroll through the results pages in a metasearch engine, you might see a wide variation in the number and quality of the results provided by each search engine or directory. Although some of the Web pages returned by one search tool will not be returned by any other search tool, you also might notice duplicate hits.

FIGURE D-11: Dogpile metasearch engine results

The screenshot shows the Dogpile search results page for the query "Mars average temperature". The page includes a search bar at the top with the query entered and a "Go Fetch!" button. Below the search bar, there are navigation links for "Web", "Images", "Audio", "Video", "News", "Yellow Pages", and "White Pages". The main content area displays a list of search results from various engines like Google, Yahoo!, and Live Search. A sidebar on the right offers related search suggestions such as "Saturn Surface Area" and "Mars Atmosphere".

Annotations on the left side of the image:

- "Indicates number of results; yours might differ" points to the text "1 - 20 of 77 from All Search Engines (About Results)".
- "Identifies hit as a sponsored link" points to the second result, "Mars Average Temperature", which is marked as "Sponsored by: Live.com".

An annotation on the right side of the image:

- "Identifies search engines that returned this hit" points to the "Live.com" logo in the search engine attribution for the second result.

Examining the KartOO metasearch engine

Figure D-12 shows the results page from one of the more interesting metasearch engines, KartOO. KartOO presents results in a graphic format. (To see these graphical results, you need to click the button to the left of OPTIONS near the top of the KartOO window.) Each image is a link and the images are clustered around words that appear in the results pages. When you move the pointer over a word, the links appear as lines between the word and the images. The word you point to is added to the end of the search expression; the clustering of links based on that term are shown as orange lines. See Figure D-12. The list of links on the left side of the page also changes to reflect the addition of the term.

FIGURE D-12: KartOO metasearch graphic results

The screenshot shows the KartOO search results page. The search bar contains "Mars average temperature". The main area is a network graph where nodes represent search results and lines represent relationships between them. The graph is centered around the search term. On the left side, there is a list of links that updates based on the search term. A "Pointer" icon is shown hovering over a word in the graph, and a red line indicates the resulting search expression.

Click this icon to see graphic results


Pointer

Internet

**UNIT
D**
Internet

Using a Subject Guide

A **subject guide** (also called a **Web bibliography**, **clearinghouse**, or **virtual library**) organizes references into categories and subcategories. In a subject guide, each reference is a link to a Web page. Some subject guides include **annotations** (summaries or reviews) of Web pages. This information can help you identify Web pages that fit your level of knowledge or interest. You can often find useful subject guides by entering the search term along with the words “subject guide” into a regular search engine.

 Ranjit needs information about the business and economic effects of current trends in biotechnology and the potential effects of genetic engineering research. He asks you to find some Web sites that he can explore to learn more about biotechnology trends in general and genetic engineering research in particular. You know that biotechnology is a branch of the biological sciences, so you identify three keywords, *biotechnology*, *genetic engineering*, and *biology*, to use as you browse the subject guide’s categories. You decide to use a subject guide so that you can use the provided reviews to help you determine the best links to give to Ranjit.

STEPS

1. Go to the Online Companion page for **Unit D**, click one of the links to a **Web bibliography** under **Lesson 6**, then scroll down the Web page and read the category links

Figure D-13 shows the home page for the Awesome Library Web site.

QUICK TIP

As with ordinary search engines, you can also enter a search term in the search expression text box on the site.

2. Examine the page for links that might lead to information about **biotechnology, genetic engineering, or biology**, then click one of the links

A list of subcategories opens.

3. Click an appropriate subcategory link, then click links as necessary to find the information that Ranjit requested; again, look for categories that mention **biotechnology, genetic engineering, or biology**

A list of links appears, or you might need to click additional subcategories. The list of links that appears after following the Science/Technology/Biotechnology links in the Librarians’ Internet Index is shown in Figure D-14. Figure D-15 shows the links on the INFOMINE Web site after using the search function on that site with the keyword “biotechnology.”

QUICK TIP

If you regularly do research in a specific field, it can be helpful to ask other researchers who work in that field if they know of useful Web bibliographies that specialize in relevant subjects.

4. Open a new tab, go to the Online Companion page for **Unit D**, then click a link to another **Web bibliography** under **Lesson 6**

As you look at the links available on this Web site, note the similarities and differences to the first Web site you examined.

5. Click category and subcategory links to find a list of reference links for **biotechnology, genetic engineering, or biology**

Compare the list of links on this page with the list of links open in the first tab.

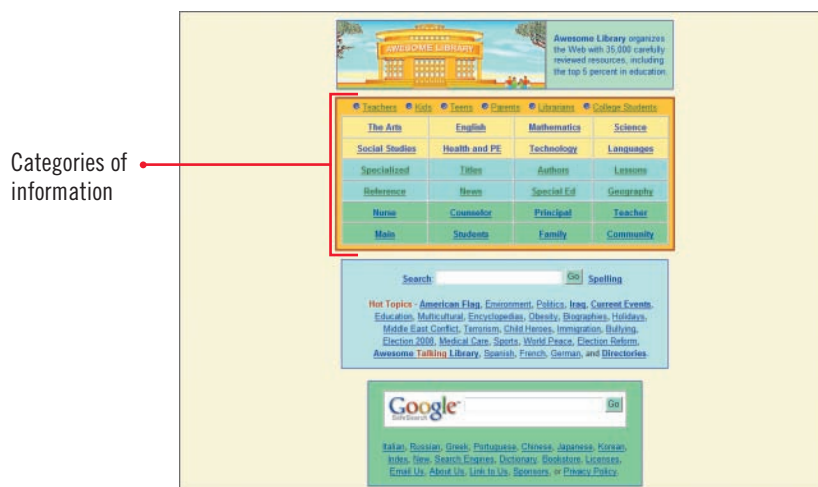
6. Close the second tab

Subject guides vs. search engines

Subject guides are compiled by reference experts or experts in the categorized fields, so they naturally do not list as many references as a search engine would. So why would you want to use a subject guide when you are given fewer results in response to a search? Subject guides are useful when you want to obtain a broad overview or a basic understanding of a complex subject area. For example, using a search engine or directory to find information about quantum physics

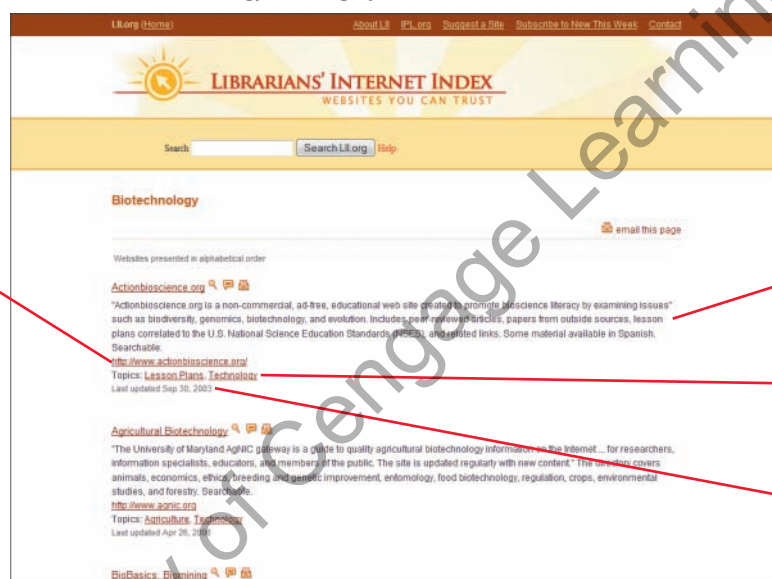
could give you millions of hits to technical papers and Web pages devoted to current research issues in quantum physics. In contrast, a subject guide page can offer hyperlinks to specific information about quantum physics at various levels so you can quickly find pages that give you a basic introduction to the subject or offer the latest news about the subject.

FIGURE D-13: Home page of the Awesome Library Web site



Categories of information

FIGURE D-14: Biotechnology subcategory on the Librarians' Internet Index Web site



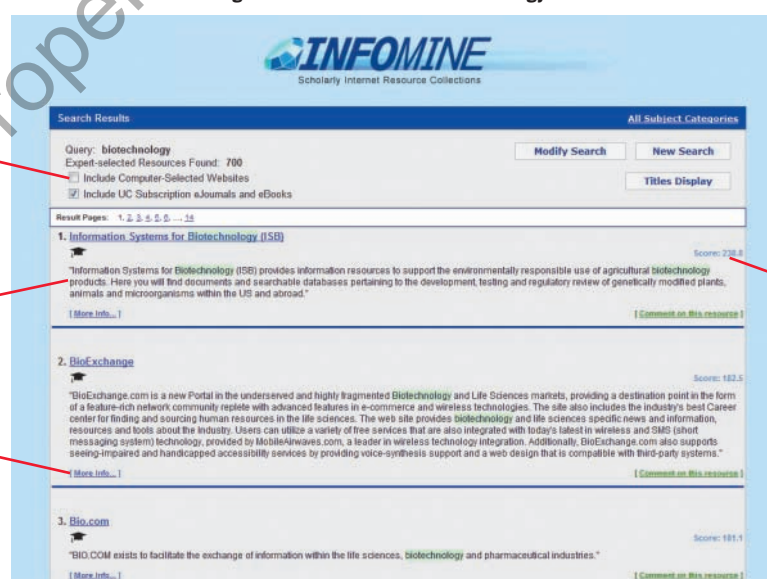
URL of site listed

Description of site listed

Topics included on the site

Identifies date the data was added to the site

FIGURE D-15: INFOMINE guide information for Biotechnology links



You can choose to include sites selected by a Web robot

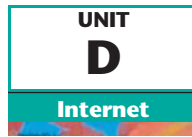
Description of site

Click to see more information about the site and this review


Higher score means a better match to search expression

Internet

Courtesy of ©1994-2009 INFOMINE, The Regents of the University of California. System developed and supported by the Library of the University of California, Riverside, IMLS and FIPSE.



Understanding Advanced Searching Techniques

To get reliable results from a search engine or a metasearch engine, you must select your keywords carefully. When the objective of your search is straightforward, one or two words often will work well. More complex search questions require more complex queries to broaden or narrow your search expression. Recall that you can restrict the search to pages that contain a specific phrase by enclosing your search expression in quotation marks. You can also use various additional techniques to perform advanced searches that will return results pages more relevant to your search question.  Some of the questions Ranjit and Nancy ask require you to find specific information. You decide to learn about techniques for formulating complex queries to reduce the number of irrelevant hits in your search results.

DETAILS

The following terms are associated with advanced Web searching:

- **Advanced Search Pages in Search Engines**
Most search engines use an Advanced Search page to provide users with a step-by-step process for conducting an advanced search. You use menus, option buttons, and text boxes to make selections that identify your search criteria. Figure D-16 shows the Advanced Search page on the Google Web site.
- **Search Filters**
A **search filter** eliminates Web pages from a search. You can access search filter options for a search engine on its Advanced Search page. You can use the filter to specify a language, date, domain, host, or page component (such as a URL, link, image tag, or title tag). For example, you could search for a term, such as *exports*, in Web page titles and ignore Web pages in which the term appears in other parts of the Web page. You can also use the keywords to filter your results by typing the keywords in a text box that indicates that the results pages will contain *all* of the keywords or in a text box that indicates that the results pages will contain *any* of the keywords. For example, if you wanted to find pages about holiday cards, you would type the keywords *holiday cards* in the text box that restricts the search to all of the keywords; but if you wanted to find information about the holidays and cards but not necessarily “holiday cards,” you would type the keywords *holiday cards* in the text box that opens up the search to include Web pages that contain only the keyword “holiday” or only the keyword “cards.”
- **Boolean Operators**
If you want, you can conduct an advanced search from the start page in a search engine by using Boolean operators. **Boolean operators**, also called **logical operators**, specify the logical relationship between the elements they join, just as the plus sign specifies the mathematical relationship between the two elements it joins. Most search engines recognize at least three basic Boolean operators: AND, OR, and NOT. You can use these operators in many search engines by simply including them in the search expression.
- **Precedence Operators**
When you join three or more search terms with Boolean operators, you can easily become confused by the expression’s complexity. To reduce the confusion, you can use precedence operators along with the Boolean operators. A **precedence operator**, also called an **inclusion operator** or a **grouping operator**, clarifies the grouping within a complex expression and is usually indicated by parentheses. Table D-1 shows several ways to use Boolean operators and precedence operators in more complex search expressions that contain the words *exports*, *France*, and *Japan*.
- **Location Operators**
A **location operator**, or **proximity operator**, lets you search for terms that appear close to each other in the text of a Web page. The most common location operator offered in search engines is the NEAR operator. For example, if you are interested in French exports, you might want to find only Web pages in which the terms *exports* and *France* are close to each other, so to perform this search you would type *exports NEAR France*.

FIGURE D-16: Google Advanced Search page

TABLE D-1: Use of Boolean and precedence operators in search expressions

search expression	search returns Web pages that include	use to find information about
exports AND France AND Japan	All of the three search terms	Exports from France to Japan or from Japan to France
exports OR France OR Japan	Any of the three search terms	Exports from anywhere, including France and Japan, and all kinds of information about France and Japan
exports AND France NOT Japan	Exports and France, but not Japan	Exports to and from France to anywhere else, except exports shipped to and from Japan
exports AND (France OR Japan)	Exports and either France or Japan	Exports from or to either France or Japan
exports AND (France NOT Japan)	Exports and France, but not if the Web page also includes Japan	Exports to and from France, except exports to and from Japan


Understanding search engine assumptions

When you enter a single word into a search engine, it searches for matches to that word. When you enter a search expression that includes more than one word, the search engine makes assumptions about the words that you enter. Most search engines assume that you want to match all of the keywords in your search expression (as if you had used the AND operator); however, a few search engines assume that you want to match any of the keywords (as if you had used the OR operator). These differing assumptions can make

dramatic differences in the number and quality of hits returned. You can always override the assumptions the search engine makes by using the advanced search page or Boolean operators. The best way to determine how a specific search engine interprets search expressions is to read the Help pages on the search engine Web site. Read these Help pages regularly because search engines change the way they interpret search expressions from time to time.

**UNIT
D**
Internet

Conducting a Search Using the Advanced Search Page

The Advanced Search page in a search engine provides users with a step-by-step process for conducting an advanced search. You use menus, option buttons, and text boxes to make selections that identify your search criteria.  Nancy is writing an article about Finland and would like to interview a professor she once met who taught graduate business students there. She does not remember the professor's name or the name of the university at which the professor teaches, but she does remember that the professor was part of the School of Economics at a university in Finland. She asks if you can search the Web to find the names of some Finnish universities that have a School of Economics. You decide to use the Advanced Search page of a search engine to conduct your search.

STEPS
QUICK TIP

If you choose the Bing or Yahoo! search engine, you must first conduct a search to see the Advanced link.

TROUBLE

If the search engine you chose does not offer a text box to restrict the search to searching to an exact phrase, type "School of Economics" (including the quotes) in the same text box you used in Step 5 to enter the word "Finland."

TROUBLE

If a message appears telling you that you did not enter any search terms, type **Finland** in the search expression text box at the top of the Advanced Search page, then repeat the search.

1. **Go to the Online Companion page for Unit D, then click one of the links under Lesson 8**
The home page of the search engine you chose opens.
2. **Click the **Advanced Search** or **Advanced link** on the Web page**
If you do not see a link to an Advanced Search page, look for an "Options" link, or try another search engine. First, you want to restrict the search to Web pages in English since you do not speak Finnish.
3. **Look for the section that restricts the search to Web pages in a specific language, then select the option to restrict the search to only Web pages written in English**
Next, you want to restrict your search to the domain .fi because schools in Finland use that domain.
4. **Look for the section that restricts the search to a specific domain, then select the appropriate option or type in the appropriate text box to restrict the search to Web pages in the Finland country domain (.fi)**
If the search engine you chose does not allow you to restrict the search to a domain, skip Step 4.
5. **In the section that provides options for identifying whether you want the search engine to search for all or any of the keywords you specify, look for options such as "Any of the words" or "All of the words," select the option for **All of the words**, then type **Finland** in the appropriate text box**
6. **Look for an option that restricts the search engine to searching for an exact phrase, then type **School of Economics** in the appropriate text box**
Figure D-17 shows the completed Advanced Search page in Google.
7. **Click the appropriate button to start the search**
The search results page opens. Figure D-18 shows the search results in Yahoo!.

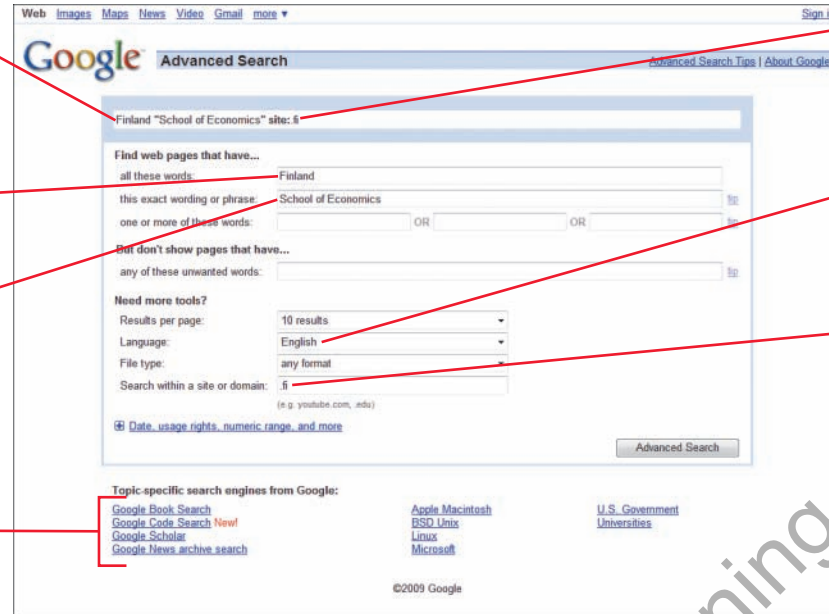
FIGURE D-17: Completed advanced search page in Google

Search expression appears here automatically as you fill in the boxes on the page

Results pages must include this term

Results pages must include this exact phrase

Click to restrict your search to specific topics



Some search engines will include "site:.fi" as part of the created search expression

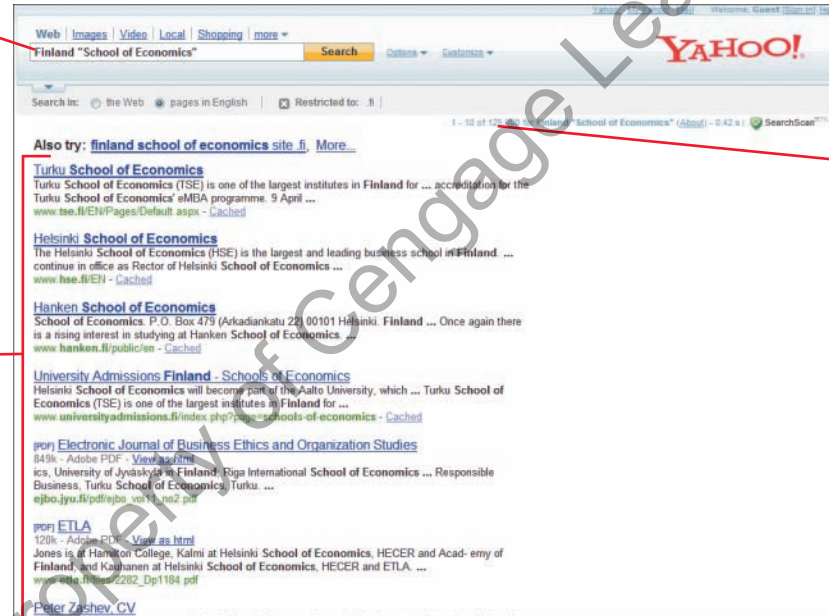
Results pages must be in English

Results pages are restricted to the .fi top-level domain

FIGURE D-18: Results page in Yahoo! after an advanced search

Search expression created automatically after using the advanced search page

Search results



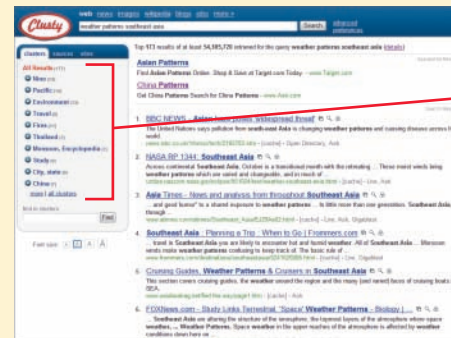
Number of hits

Courtesy of Yahoo! Inc. © 2009 by Yahoo! Inc. YAHOO! and the YAHOO! logo are trademarks of Yahoo! Inc.

Using search engines with clustering features

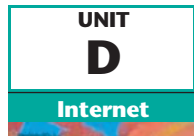
One problem with using search engines is that they often generate thousands (or even millions) of hits. Scrolling through hundreds of results pages looking for useful links is not very efficient. Some search engines use an advanced technology to group search results into clusters. The clustering of results provides a filtering effect. The filtering is done automatically by the search engine after it runs the search. Figure D-19 shows the search results for weather patterns in the past six months in Southeast Asia in Clusty, a search engine that clusters results.

FIGURE D-19: Search results page in Clusty clustered search




Search results collected into clusters of links

Internet



Conducting a Search Using Boolean Operators

Sometimes the Advanced Search page does not offer you enough options to structure your search expression the way that you want. In that case, you can try searching using Boolean operators.  Ranjit is writing about fast-food franchises in various developing countries around the world. He would like to feature this industry's experience in Indonesia in an upcoming story and asks you for help. You recognize this request as an exploratory question and decide to use Boolean operators in a search engine to conduct a complex search for Web pages that Ranjit can use for his research.

STEPS

- 1. Go to the Online Companion page for Unit D, then click one of the links under Lesson 9**
The home page of the search engine you chose opens.
- 2. Type “fast food” franchises Indonesia OR Thailand in the search expression text box**
This query instructs the search engine to look for Web pages containing the following characteristics: the exact phrase “fast food”; the word “franchises”; and either the word “Indonesia” or the word “Thailand.” You do not need to type the Boolean operator “AND” because all of the search engines listed under Lesson 9 use all the keywords in the search expression by default. Figure D-20 shows the search expression in the Bing search engine.
- 3. Click the appropriate button to start the search**
The search results appear in the window.
- 4. Note the number of hits at the top of the list of links, then examine some of the descriptions of the first 10 results**
You may need to go to several results pages to find the information you are looking for.
- 5. Exit your Web browser**

FIGURE D-20: Search expression with Boolean operators in Bing

Searching the Deep Web

One weakness of most current search engines and Web directories is that they only search static Web pages. A **static Web page** is an HTML file that exists on a Web server computer. The robots used by search engines to build their databases can find and examine these files. An increasing number of Web sites do not store information as HTML files. Instead, they store information in a database, and when a user submits a query, the Web server searches the database and generates a Web page on the fly that includes information from the database. These generated Web pages are called **dynamic Web pages**. For example, if you visit Amazon.com and search for books about birds, the Amazon.com Web server queries a database that contains information about books and generates a dynamic Web page that includes that information. This Web page is not stored permanently on the Web server and cannot be found or examined by search engine robots. Much of this information can only be accessed by users that have a login and password. In 2001, Michael Bergman of BrightPlanet published a paper that explored the difficulties that search engines face when trying to include this information in their search databases. He called this information the **deep Web**; other researchers use the terms **hidden Web** and **invisible Web**. Researchers working at the University of Utah have created an experimental Web site that allows visitors to search the deep Web. The home page of this site, called DeepPeep, is shown in Figure D-21.

FIGURE D-21: DeepPeep home page