Module 3
Introduction

The World Wide Web is a subset of the Internet, linking the information world with hypertext. The Web is currently the service that most people use to access Internet resources and services. Because the Web is not indexed in any standard way, finding relevant information often seems an impossible task. There are several basic types of search tools that may be used to locate web resources: search engines, meta-search engines, metasites, and directories. The following chart details the differences between these search tools and provides examples of when to use each.

<table>
<thead>
<tr>
<th>Internet Search Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search Engines</strong></td>
</tr>
<tr>
<td>Database generated by computer program</td>
</tr>
<tr>
<td>Index a large percentage of web resources</td>
</tr>
<tr>
<td>Use keywords for precise searches</td>
</tr>
<tr>
<td>Use for specific, focused searches, narrow topics</td>
</tr>
</tbody>
</table>

Since the ever-changing nature of the Web provides access to vast numbers of information resources, web sites and documents appear, are deleted, or are moved to a different location each day. In this dynamic environment, search engines can be the most efficient way of locating information on a specific topic since they provide access to immense, continuously updated databases of Internet resources. There are hundreds of search engines designed to help you find information, whether you are looking for a topic of personal interest, or material for a scholarly research project.
Using search engines effectively may seem intimidating since new search engines appear frequently and existing engines often change their search interface and format. Though there is at present no consistent standard that governs search engines, they do share many basic features that allow the searcher to retrieve relevant information.

**Free Resources Available Via the Web**

The number and type of resources available through the Internet increases daily. The following types of information are usually free to any Internet user:

- Current events from newspapers, current issues of magazines, and news wire feeds
- Corporate information, including annual reports, product information, and stock quotes
- Government information such as current laws, regulations, court decisions, and information from local, state, and federal government departments and agencies
- Ready reference material, including dictionaries, some encyclopedias, statistical sources and other quick answer sources, such as:
  - *Encyclopedia Britannica*
  - *Merriam-Webster's Dictionary*
  - *Statistical Abstract of the United States*
  - *Occupational Outlook Handbook*
- Bibliographic information from library OPACs (Online Public Access Catalogs). Books and other materials located in remote catalogs can often be borrowed from a local library via interlibrary loan.
- Bibliographic information from various disciplines, including:
  - *PubMed*, which offers bibliographic references and abstracts to articles from over 4,800 biomedical periodicals.
- Texts of books in the public domain (generally books published more than 75 years ago, which are not protected by copyright laws) from sites such as:
  - *Project Gutenberg*, the oldest producer of free electronic books, currently offering more than 18,000 texts.
  - *The Camelot Project*, which offers public domain literature relating to the Arthurian legends.
- Material on popular culture, such as cinema, television, and sports such as:
  - *The Internet Movie Database*, which provides information on movies, producers, and actors.
- An increasing number of websites from colleges, universities, and associations, which post information ranging from student research papers to scholarly works by professors and others who are experts in their subject fields
- Postings to discussion groups, asking or answering specific questions on a particular topic

Articles from some current issues of popular and scholarly journals may be found through searchable databases such as *Google Scholar* and *FindArticles.com*. In addition, there are many electronic journals freely available via the Web. However, most
academic research will require access to journal articles that are only available through library subscription databases.

**How Do Search Engines Work?**

Most search engines use a computer program called a "spider" to collect information and index web resources. Sometimes called "webcrawlers" or "robots", these computer programs crawl through websites on the Internet, gathering information from all the pages of a website. The spider returns the information to a central database and then indexes the information it has gathered. When you use a search engine, you are searching the database compiled and indexed by the spider.

While all search engines rely on spiders to collect and index information, each performs its tasks in a slightly different way. Each search engine has its own search interface and uses different criteria for matching searches with documents. Each may also differ in terms of search speed and how it ranks results in order of relevance.

Searching would be easier if the search engines used a common standard. However, each search engine operates a little differently, and each search engine database contains a large number of unique documents, with limited overlap. Therefore, it is a good idea to search using more than one search engine to be sure you have retrieved most of the relevant information available on your topic.

**Relevancy and Search Terms**

A search is performed by submitting keywords in the search box. The search engine compiles a list of websites that contain these terms. The order of these sites is often determined by relevancy (i.e., how closely the site matches the query). Search engines look at the location and frequency of occurrence of search term to help determine relevancy. The higher up on a website a search term appears, the higher the ranking of that website. A website that contains a search term in the title or in the first few paragraphs of text will be determined to be more relevant than one in which the search term appears toward the end of the document. Search engines also look at the number of times search terms appear in the text of the website. Sites with a higher frequency of a search term are determined to be more relevant. Google even looks at font size and boldness to help determine relevancy.

**Ranking and Popularity**

In addition to text-matching techniques, an increasing number of search engines are using popularity and link analysis as a means of ranking search results.

Google uses PageRank technology to rank the usefulness of a website. Google interprets a link from website A to website B as a vote by site A for site B. The more votes or links a site receives, the more relevant the site is. In addition to looking at the number of links a site receives, Google also analyzes the sites casting the votes. Votes cast by sites which are themselves major sites, like CNN for example, are weighed.
more heavily than votes from other less popular sites. Link analysis is comparable to the time-honored tradition of researchers rating the importance of a study or article by the number of times it is cited elsewhere.

**Sponsored Links**
Most major search engines accept paid listings. Some search engines sell commercial spots on the results list so that the buyer’s page is near the top as if it was one of the best results according to a link analysis. In the best search engines, sponsored links or paid listings are clearly labeled, kept separate from search results, but are relevant to the search.

**Size**
When search engine producers refer to their size, they are usually counting unique URLs as opposed to unique sites, which may contain a number of URLs. The search engine with the largest collection of sites is not necessarily the best search engine, but, potentially, the larger the search engine the greater the chance that you will find something.

**General Search Features**
Most of the major search engines support the following search techniques, although each search engine operates a little differently. To find out which features are supported by a search engine, read the help page. There is usually a link to a help page near the search box or near the top of the search engine’s home page. If it is not in one of these places, try selecting the search engine’s Advanced Search option.

Just like other Internet resources, search engines often change their appearance and features with little or no notice.

**Bottom Line:** If you are not certain which techniques the search engine uses or if your search statement does not work, reread the help page.

**Case Sensitivity**
Some search engines are case sensitive, requiring that proper names and place names be capitalized. In general, when a search statement is entered in all lower case, both lowercase and uppercase will be retrieved. The reverse is not true. When uppercase is used, the search engine will only retrieve the exact match. For example, AIDS will not retrieve the common word, aids.

**Boolean Operators**
Most search engines support Boolean searching, allowing **AND**, **OR**, and **NOT** searches. Some search engines require that the Boolean operator be capitalized; others do not, although those not requiring capitalization accept it. Therefore, it is a good idea to capitalize any Boolean operator. See also Lesson 2E on Boolean searching.

Many search engines use a simplified form of Boolean operator, replacing the operator with a symbol:
the + sign for an AND search
Example: +drinking +driving searches for the words drinking AND driving, in no specific order in the text of the web page.

the - sign for a NOT search
Example: +dolphins -football will search for documents which contain the word dolphins but NOT the word football

Google defaults to an AND search (automatically placing an AND between terms), and uses a - sign to indicate NOT. This means that you do not have to type AND in your Google search statements. However, for explanatory purposes, in this course the AND operator will be included in search examples, and for class exercises you should include this operator in your search statements where applicable.

**Nesting**

Search statements combining more than one type of Boolean operator must also use parentheses around synonymous terms. This technique is called nesting. The parentheses tell the search engine to perform that search first. For example, suicide AND (teen OR youth OR adolescent) will search for documents containing any or all of the terms within the parentheses before combining that result with the word suicide.

**Phrase Searching and Truncation**

Most search engines support the use of quotation marks around words, terms or names you want searched as a phrase, i.e., appearing in exactly the order you enter them. For example, "ozone layer depletion" searches for this exact phrase with the words in the order given.

When devising a phrase search, be sure to evaluate the likelihood of your phrase being used by others. For instance, if you were doing a search on the benefits of reading to children, "reading children" would not return results as well as "reading to children." Phrase searching is the one time you may use minor words like of, in, to, etc.

Some search engines automatically look for singular and plural forms of terms as well as -ing or -ed endings. Others use the asterisk (*) to specify that all endings of the root term be searched. This technique is called truncation.

**Field Searching**

Some search engines allow you to limit your search to specified fields, such as the title of the document, a word from the URL, the domain name, the type of file, and the availability of such features as images, sound, and video. In the following table, four types of field searching are demonstrated (title, URL, domain, and file type) in addition to phrase searching and truncation. All of these syntaxes will work in Google except for the truncation symbol (Google now uses stemming technology to automatically truncate for you).
### Creating Effective Search Statements

The next table demonstrates how these techniques can be combined to create effective search statements.

<table>
<thead>
<tr>
<th>Search Query</th>
<th>Search Techniques</th>
<th>Search Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>You want government sites that discuss bioterrorism</td>
<td>domain searching</td>
<td>bioterrorism site:gov</td>
</tr>
<tr>
<td>A friend told you about a great site on elephants that had wildlife in the</td>
<td>URL searching, title</td>
<td>elephants inurl:wildlife intitle:africa</td>
</tr>
<tr>
<td></td>
<td>searching</td>
<td></td>
</tr>
</tbody>
</table>
URL and Africa in the Web page title

You need an Excel document with statistics on international adoption

| Boolean operator, phrase searching, file type searching | statistics AND "international adoption" filetype:xls |

You are looking for sites that relate to children who have ADHD

| nesting, Boolean operators, phrase searching, truncation | (ADHD OR "attention deficit hyperactivity disorder") AND child* |

These are just a select sample of search techniques commonly available for search engines. For additional search features, read the help page of the search engine you are using.

**Advanced Search**

Many search engines offer an advanced search mode. In advanced search, you are able to perform many search techniques by utilizing designated pull-down menus instead of correct syntax to limit your search. Since syntax will vary between search engines, using advanced search often saves time and frustration. However, keep in mind that not all search techniques will be available in advanced search. The following graphic shows Google's Advanced Search Screen:
Notice that the form allows you to use Boolean Operators as follows:

**AND** = all these words
**OR** = one or more of these words
**NOT** = any of these unwanted terms

Also, "this exact word or phrase" is equivalent to using quotation marks to designate a phrase.

For more information on Google's Advanced Search features, check out the Advanced Search Help Guide.

Be sure to check out the advanced search options of your favorite search engines.

## Comparison of Major Search Engines

Some of the most popular search engines are listed below, along with links to their help files. Some search engines, such as Ask.com and Wolfram Alpha, work to provide you the answer to your search without having to go to a third party site. Google is also incorporating this feature into their search engine by providing the answer to common questions at the top of their search results.

<table>
<thead>
<tr>
<th>Major Search Engines</th>
<th>Help Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask.com</td>
<td>Ask.com Help</td>
</tr>
<tr>
<td>Bing</td>
<td>Bing Help</td>
</tr>
<tr>
<td>Google</td>
<td>Google Help</td>
</tr>
<tr>
<td>Yahoo! Search</td>
<td>Yahoo! Search Help</td>
</tr>
</tbody>
</table>

## Spotlight on Google

According to comScore, a "digital marketing source" in November of 2015, Google conducted 63.9% of all online searches in the U.S. (Microsoft was the next busiest search engine with 20.9% of searches). Google’s size, uncluttered interface, and fast searching have made it easily the most popular search engine. The following are examples of additional features that make Google stand apart:

- **Google Images** provides access to more than billions of images
- **Google News** provides access to nearly 10,000 worldwide news sources
- **Google Translate** provides translation services for text selections or entire Web pages
- **Google Scholar** indexes scholarly information including peer-reviewed articles, theses, books, preprints, abstracts, and technical reports

*Note: Some full-text results retrieved by Google Scholar will only be available for a fee*
For the full selection of Google features, click the **box icon** link at the top of the general screen, or above your results on your search screen:

![Google Search Interface](image)

**Meta-Search Engines**

A special kind of search engine, called a **meta-search engine**, allows you to query several search engines at once. Instead of doing a search itself, a meta-search engine sends your request to other search engines, compiles the results, and displays them for you. This process can be much faster than querying several search engines separately.

Meta-search engines do not own a database of web pages—they use and deliver results from the databases and search programs of each of the individual search engines they query. Meta-search engines act as an intelligent middleman to pass your search through, gathering the responses and then giving you a report from several engines at once. As well as saving time, this kind of search engine can give you an overview of the kind of document you may find using your search terms and may even result in giving you exactly what you need if you are searching for a unique term or phrase.

There are some disadvantages in relying exclusively on meta-search engines. None of the meta-search engines query all of the largest search engines. If a search connection takes too long, one or more of the search engines may time out and produce no results. If you submit a complicated search to a meta-search engine that one of the queried tools does not "understand," you may get no hits at all from that engine. However, you will usually get results from another tool that supports your search strategy.

Meta-search engines retrieve only the first 10-50 hits from each search engine; the total number of hits may be less than you would retrieve with a direct search on a single search engine. Thus, meta-search engines do not eliminate the need to learn how to intelligently search at least one or more general web search engines.

Each meta-search engine has its own interface and method for letting you choose engines to search, so it is important to consult the "Help" pages for each meta-search engine.
**Links to Major Meta-Search Engines**

Some of the most popular meta-search engines are listed below, along with links to their help pages:

<table>
<thead>
<tr>
<th>Meta-Search Engines</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dogpile</td>
<td>Dogpile FAQs</td>
</tr>
<tr>
<td>Info.com</td>
<td>Info.com’s How to Get a Better Search Result</td>
</tr>
<tr>
<td>Yippy</td>
<td>Yippy FAQs</td>
</tr>
</tbody>
</table>

Licensed under the [Creative Commons Attribution Share Alike 3.0 License](http://creativecommons.org/licenses/by-sa/3.0/)