Using the Web for research

In this tutorial:

1. How the Web is different from a library
2. Background information: Web directories and search engines
3. A simple strategy for searching the Web
4. How to use a search engine (like Google or Yahoo!)
5. The results page
6. Judging the quality of information
7. Using Wikipedia

Whether you are shopping, checking your symptoms to help decide if you are seriously ill, or researching for school or your job, being able to find reliable information is a handy skill. Two of the best places to find information are libraries and the Web. The library and Internet systems are organized very differently, and each requires its own skills.

The problem with the Web:

How is it possible to find the information you want in a system where people can put anything they like anywhere they want and then change or move it without notice? It sounds impossible! Luckily, it's not. There are a few ways that you can go about finding what you need on the Web.

How the Web is different from a library

	Libraries	World Wide Web
How is information organized?	Books and other resources are kept on shelves and organized according to categories determined by the Dewey system or another system like it.	There are no categories on the Web. Information is just posted on websites and linked together in whatever way the website creators decide is appropriate.
Who does the organizing?	Each piece of information is analyzed and placed in the appropriate category by a librarian.	After a website is discovered by a search engine, automated computers repeatedly scan it for updates and changes, and record the results in giant databases.
Who decides what information is available?	People at publishing companies decide which information to publish. Librarians are responsible for choosing which books, magazines, and other resources are purchased in order to best meet the needs of the community.	Anyone can publish a website and submit it to the search engines to be indexed. Search-engine companies spend a tremendous amount of time trying to determine which information people are most likely to want and making it easy for them to find it. Advertisers and scam artists try to get their sites listed at the top of search results through any means necessary.
How do you find what you need?	Learn the library system and ask a librarian for help when required.	Learn how a search engine works and keep up with improvements and new features as they are added.

Background information: Web directories and search engines

Web directories

A Web directory is a list of links that has been organized into useful categories. Directories only contain links to a tiny portion of the Web, but each link has been added to the list after someone has personally checked it out. Web directories are a good starting place for doing research, but they shouldn't be your only strategy.

Yahoo! directory

Yahoo! (http://ca.search.yahoo.com/dir) created the first really successful Web directory. They organized lists of useful links, every one of which was reviewed by an employee. They still have a directory, but Yahoo! is more famous for free email and other services.

Open directory project

The largest human-edited directory is the Open Directory Project (http://dmoz.org). ODP is a volunteer effort to organize information on the Web. Links are listed alphabetically under each category.

Google directory

Google directory (http://directory.google.com/) builds on the hard work of the ODP volunteers and combines it with the clever technology that runs Google's search engine to provide a more organized directory.

Search engines

Most people rely on search engines to help them find what they need on the Web. All you have to do is type a few words into a search engine and all the information in the world is at your command. Understanding what goes on behind the scenes at a search engine helps you to use search engines effectively.

How it all works

Web crawlers

Search-engine companies use automated computers to visit every web page they can find. The visits may be as frequent as several times a day or as infrequent as every two weeks, depending on how often the site is changed or updated. (News sites can be changed every few hours, for instance.) Every word, every image, and every audio and video clip is recorded in a giant database so that when someone is searching for it, it can be found.

Ranking information

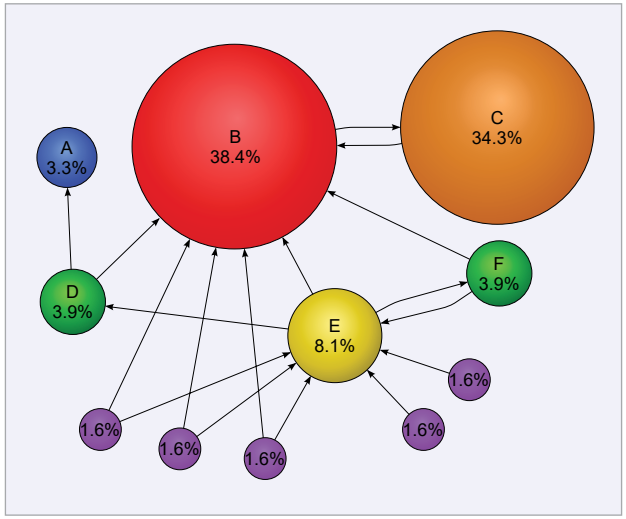
Unlike a Web directory or a library, no one decides which links are good and which are a waste of time. The web crawler just records them all. The main job of a search engine is to present the links in some kind of useful order, so that you don't get overwhelmed by advertisements.

Popular search engines

Google (http://www.google.ca)
Yahoo! (http://www.yahoo.ca)
Bing! (http://www.bing.com)

Google uses a system called PageRank to determine which search results get listed at the top of a results page. The actual details of how PageRank works are top secret, but the basic idea is not too hard to understand. People create web pages that link to other pages. Google treats each one of those links as a vote for the page it points to. The page with the most links leading to it is considered to be the most important, so it is usually displayed at the top of the results list when you search for a term.

It is still important that you think for yourself and judge the quality of the information you find, though. Search engines don't always place the best information at the very top.



An example of how PageRank decides the most important page on a topic. Each circle is a web page, and larger circles with more links to them are considered by Google to be more reliable.

A simple strategy for searching the Web

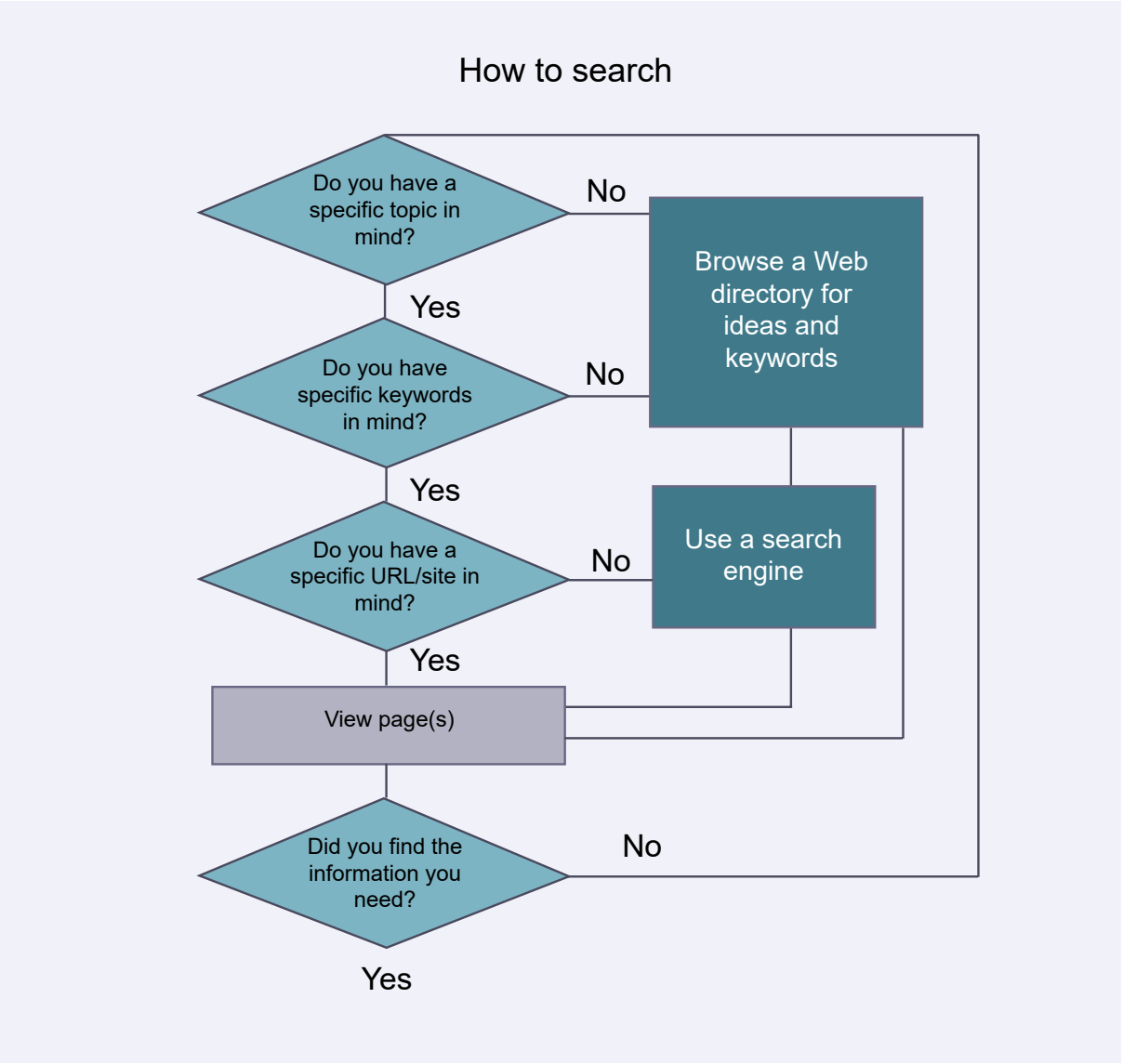
The information on the Web changes every day. Businesses and advertisers find ways to get their pages to the top of search engine results in the hope that you will click on them. From month to month, searching for the same words can give you very different results. You need a strategy for finding what you are looking for!

- ❑ It helps to be familiar with the words that are being used. Browsing a Web directory, a Wikipedia article, or a few pages related to your topic can help you pick out search terms.
- ❑ If at first you don't succeed, try again. Look at the results you get when you search and see if there are any hints as to what you want to focus on or avoid.
- ❑ Be patient, you will find what you need.

Try this!

Find some information about the oldest form of martial arts that is still practised. Is there somewhere near you that teaches this practice?

The flow chart on the right will help to guide your search.



How to use a search engine (like Google or Yahoo!)

Basic tips

1. Use words likely to appear on the pages you want to find. Use a web directory or Wikipedia article to help identify keywords.
2. Less *is* more. The fewer words you enter in the search engine, the greater number of results you will get. Start with general terms, look over the results you get, and then add words to get fewer, but more specific results.
3. Don't worry about spelling. Most search engines will catch your spelling mistakes and offer to correct them.

Try this!

1. Compare searching for the quotation "Above everything, we are Canadian" both with and without the quotation marks.
2. Determine whether some kinds of cat litter cause health problems for cats.
3. Find the process used to make the flu vaccine.
4. Find out where the names Yahoo! and Google came from.
5. Find instructions to help you refinish a maple coffee table.
6. Find a definition for the word *tarantism*.

Search operators

Along with the keywords you enter, there are a few more **search operators** that can be used to make your search more specific. Example searches have been enclosed in square brackets [ ].

- (**AND**) Most search engines assume that you want to find all of the words you enter in the search box. This means that search engines assume you have included "AND" in between search terms. Searching for [dog training] is the same as searching for [dog AND training], so the search engine will return links to pages that have both words on them.
- (**OR**) Using this operator lets you search for two different things at the same time. Searching [dogs cats] will get you pages with both words. Searching for [dogs OR cats] will give you a list of pages containing the word *dogs* and the word *cats* and both *dogs* and *cats*.
- (**""**) Sometimes you want to search for an exact phrase made up of several words, such as a direct quotation. If you enclose your phrase in quotation marks, the search engine will give you results that have the exact phrase. Without the quotes, you will get any page that happens to have all of the words on it.
- **(-)** It can be helpful to exclude certain words from your results. If you search for [basketball], you will get a lot of results about the NBA. To eliminate those results from your search, try [basketball -NBA]. *Don't leave any spaces between the - and the word.*
- (~) Sometimes you know what you want, but can't find the best word to search for. This operator (called a *tilde*) tells the search engine to look for pages with the word you chose and any synonyms (words that mean the same thing) of that word. Compare [desert] and [~desert].
- (**site:**) Not every site in the Web is well organized. If you know who has the information you want, you can search their site specifically. Use the **site:** operator to narrow down your results to just one website. If you want a news story you remember from the *Toronto Star* about teen drivers, try [teen drivers site:thestar.com] Compare that to results from [teen drivers] alone.
- (**define:**) One of the most convenient things you can use a search engine for is to find definitions for words and phrases. To define a word like *egotism*, try [define:egotism]

The results page

Search engines return a list of links to web pages in the order they think is most useful to you. Usually each link includes the following.

Title: The web page's title, if the page has one. Click on the page title to visit the page.

Snippets: Each search result usually includes one or more short excerpts of the text from the page that matches your query, with your search terms in boldface type.

URL: The web address of the search result

Size: A rough indication of how much information is on the page

Date: The date on which the site was last visited to index the content. Some links are very current, while others can be weeks out of date.

It's usually easy to spot paid advertisements somewhere on the results page. This is how search engines make money.

The search results page may also include links to audio and video information, images, and news stories. Often, these are presented in groups or as categories that you can switch to from your main search results.

The results page shown on the right may not be exactly how Google appears when you use it, as they constantly improve and update their services. But it gives you an idea of all the different kinds of links they provide.

Judging the quality of information

Don't believe everything you read

And, it doesn't hurt to think twice about what you watch and listen to either. All kinds of things can be easily published, especially on the Web. That means you have to be careful about what information you choose when you want to learn about something or make a decision.

Some people post information that is related to a favourite cause, while others may post things as a joke. Some people post information that is of little value, since they are just not very well informed themselves. When reading information on the Web, look at the following criteria.

Relevance: Does the source contain the information you need? Is the information written at an appropriate level and does it thoroughly answer your questions? If you are reading something that is too complicated, or something written for young children, it may not be worthwhile.

Credibility: Can you trust the author? What are his/her credentials for writing this particular work? The author should display some expertise in the field and belong to a credible organization. Judging an organization and the person's connection to it are difficult. Unfortunately, people make claims about themselves that are untrue.

Accuracy: Is there evidence of good research and documentation? Look for appropriate references, both to justify claims and to demonstrate that copyright material is not being misused. Facts and arguments should be easy to double-check and confirm, using other credible sources.

Objectivity: Is the author's purpose and intended audience clearly stated? Does the work provide factual information, opinion, or propaganda? Does the author stereotype people or have a biased point of view? If you realize that a source is biased, that doesn't mean you can't use it. But you should find other sources that express a different point of view as well, so that your own work is not biased.

Currency: Is the publication date given? Is the information current enough for your topic? Sometimes today's date is automatically displayed on websites, so look for clues as to how old the information is. Do the provided links still work?

Try this!

Public health is one of the most debated topics in the world. In North America, some of the topics that are often argued about include:

- Adding fluoride to drinking water to promote dental health
- The vaccination of children, which is mandatory, and which some people believe can cause health problems
- The effects of second-hand smoke

Use the Web search skills you have learned so far to locate some information about one or more of these topics. Apply the criteria above to the information you find and determine which side of the argument does the best job of presenting their information.

Using Wikipedia

Wikipedia is the largest encyclopedia in the world. It has thousands of contributors and articles on just about any topic. Wikipedia sounds like the perfect first stop when you want to learn about something. But wait! Wikipedia allows anyone who visits their site to change almost anything, at any time, no matter who they are. In fact, they don't even ask who they are. Now that sounds like a disaster waiting to happen! Whether it is perfect, a disaster, or somewhere in between, is up to you to decide.

In fact, Wikipedia encourages you to be critical. There are guidelines posted on the site to help you judge the articles:

http://en.wikipedia.org/wiki/Wikipedia:Researching_with_Wikipedia.

Try this!

Compare the Wikipedia articles about former Canadian Prime Ministers Pierre Trudeau and Lester B. Pearson with the current Canadian prime minister or leader of the Opposition. Visit the "history" and "discussion" pages for each article. What kinds of issues are being discussed? Are recent changes to the article minor edits or major additions? Are the articles tagged with any warnings? Which article would you trust the most?

How to evaluate a Wikipedia article

Quality:

- The article should contain references and links to resources to help you verify what you are reading. Always double-check the information you find on Wikipedia.
- The article should be easy to read and well organized in small sections. If it's long and confusing, it wasn't written with care.

History:

The "**history**" tab at the top of the article displays a record of all changes that have ever been made to the article. Look for the following:

- When was the article created? Older articles have likely been checked more often.
- Have many people contributed to and edited the article? Many, or just a few?
- Is there a dispute over the article? Have the same back-and-forth changes happened over and over again?
- Is the page continually being vandalized and then fixed, over and over? If so, check to see that you have a good version.

Discussion:

The "**discussion**" tab at the top of the article shows what other people think about the article. If there are questions about the validity or objectivity of the article, you may want to avoid it.

Warnings:

Is the article tagged? People who read and write articles on Wikipedia help you to spot unreliable information by adding tags at the top of an article. You might see, for example:

- "*To meet Wikipedia's quality standards, this article or section may require cleanup.*"
- "*This article or section does not cite its references or sources.*"
- "*Some information in this article or section has not been verified and may be unreliable.*"
- "*This article or section may contain original research or unverified claims.*"

References

Google guide
http://www.googleguide.com/crafting_queries.html

Wikipedia
http://en.wikipedia.org/wiki/Main_Page

http://en.wikipedia.org/wiki/PageRank

http://en.wikipedia.org/wiki/Wikipedia:Researching_with_Wikipedia.

Google
http://www.google.com

http://www.google.com/corporate/tech.html

http://www.google.com/support/websearch/bin/answer.py?hl=en&answer=136861

Try this! hints

Page 3

A simple search for [martial arts] or [martial arts history] should locate a Wikipedia article that might be useful. Use the guidelines for judging the quality of a Wikipedia article on page 7 of this tutorial to decide if it's a good resource. A search for [oldest martial arts] might lead to a discussion forum containing some information.

Page 4

1. Including the quotation marks leads to a lot of results about the direct quote from Sir George Etienne Cartier. Without the quotation marks, you will still find a few references to the quotation, but your search results will have lots of unrelated links as well.
2. A simple search for [cat litter health problems] should identify concerns about clumping cat litter that contains sodium bentonite, and cats' health. Use the guidelines in this tutorial to assess the quality of that information.
3. Searching for [flu vaccine] will lead to some general articles that discuss how the vaccine is made. Searching [how to make flu vaccine] should give more specific results.
4. Searching for just Google or Yahoo! brings up a long list of the services each company provides. [Google history] or [Google name] and [Yahoo! history] should lead to more specific results.
5. This is a good example of how less is more. A search for [refinish maple] should give you excellent results, while [refinish maple furniture] and [refinish maple coffee table] don't really do any better.
6. A search for [define:tarantism] should give you a definition.

Page 6

All three of the topics mentioned generate a lot of passion in people. Often, the information posted relates to personal experience, lacks references to credible scientific studies, and should not be trusted for academic use. However, a few websites run by organizations like Health Canada or the World Health Organization should provide useful facts and information.

Page 7

The articles about Pierre Trudeau and Lester B. Pearson have been around for a long time and the people they are about have died. These articles tend to have less recent changes and limited discussion. Articles about the current prime minister are updated frequently and can generate controversy and disputes. It is best to double-check anything you learn from these articles.