# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Defining web spaces</td>
<td>2</td>
</tr>
<tr>
<td>Google Tools Overview</td>
<td>3</td>
</tr>
<tr>
<td>Non-Google Tools Overview</td>
<td>4</td>
</tr>
<tr>
<td>General things to know about Google</td>
<td>5</td>
</tr>
<tr>
<td>A Deeper Look at Google Search</td>
<td>7</td>
</tr>
<tr>
<td>A Deeper Look at Google Scholar</td>
<td>8</td>
</tr>
<tr>
<td>A Deeper Look at Google Books</td>
<td>11</td>
</tr>
</tbody>
</table>
Introduction

The Internet began as a means of scholarly communication, and continues to be a primary medium of scholarly activity. There are many online-only journals, and even established print-based scholarly journals usually have an online version of at least some of their content. A wealth of information lies at your fingertips on the web, but figuring out how to find things, and really understanding what you're finding makes academic research on the Internet more challenging than many people think.

This handout discusses options for doing academic searching on the web, and particularly explores tools offered by Google, and their applicability for academic research. Because no search tool is appropriate in all situations, we’ll also explore a lot of non-Google search tools, and discuss their applicability as well.

Defining web spaces

Public web:

The public web includes anything that exists on the web and is accessible to anyone with an internet connection with no additional charges involved. This may include web pages, publicly-available online journals or databases, images, pdf files, and more.

There's something to the old saw that “If something seems too good to be true, it probably is.” A lot of the free information on the web has little to no quality control. It's up to you as a consumer of information to make sure that what you're getting is actually of value.

Fee-for-service web:

The fee-for-service web includes online resources that require a subscription for access, or through which you can access information at a per-piece, per-search, or per-time-unit fee. For scholarly purposes, this may include full-text databases, indexing and abstracting services, online full-text journals (in HTML or PDF format), archives, and more. Libraries subscribe to many such resources and often share access with patrons even outside library or campus buildings. Others may require that individual users purchase their own access.

Usually fee-for-service resources have much better quality control (at the very least, they usually explain what is included, and how included resources are selected). Unfortunately, they're often slightly harder to access and use than the free resources, so users often settle for what’s available on the public web.
General Google search

Searches in a database of billions of web sites on the public web. No sites are intentionally excluded, so results may include sites from commercial enterprises, non-profit organizations, government entities, entertainment industries, random individuals, etc. Results are ranked according to the “PageRank” algorithm, which brings more popular sites to the top of the list of results.

+ Can usually answer brief questions well, and quickly.
+ May find large numbers of results
- No quality control – search results could be from anyone, anywhere.
- More popular things rise to top of list, so academic resources (especially on obscure topics) may not be easily found

Google Scholar – http://scholar.google.com

Searches in a database of millions of web sites and online information sources, including mostly resources on the public web, but some fee-for-service resources. The database is targeted to include largely scholarly resources, including sites from professional and industrial organizations, academic presses, public online scholarly resources (such as public online journals and databases), and sites within .edu domains. Result rankings are determined by “relevance”, which according to Google “considers the full text of each article, the author, the publication in which the article appeared, and how often the piece has been cited in other scholarly literature.” It’s possible that the “PageRank” algorithm used in general Google searching somehow impacts ranking, too.

+ Targets your search to scholarly resources
+ Targets publicly-available resources
+ Includes information often left out of subscription scholarly databases
- Not at all clear what is included, and what is excluded – may include resources of dubious scholarly value, and exclude others of better quality.
- Not clear how results are ranked – may bias search results to more popular resources?
Google Books – http://books.google.com

Searches in a database containing the text of many books. Right now, books are either submitted by publishers under agreement with Google, or scanned in from one of the participating libraries. Books that are out-of-copyright can be viewed in their entirety through this site (for the most part.) For books that are still in-copyright, you can view a few pages, then you'll be locked out of viewing more.

+ Allows you to search inside of books!
+ Often links to library catalogs, or to places you can purchase the book online
- Small catalog as yet, biased by what publishers submit/what has been scanned
- Limited ability to browse contents of books

Non-Google Tools Overview

Other web search engines

Like general Google, these tools search the public web. All rank their results in different ways, and most search different databases of web pages, so you may find very different results from one search engine to another.

Examples: Yahoo, Metacrawler, etc.

Fee-for-service databases

The University of Michigan library subscribes to thousands of online informational databases, and has access to thousands of journals in online formats. Many of these databases and online journals provide the full-text of articles online. Others may simply provide a citation, leaving it up to you to track down the physical object. Most of these online resources are targeted to specific academic subject areas, and largely contain information that is NOT available on the public web (and thus not found by searching Google and other public online search engines.)

Many online fee-for-service databases and electronic journals allow users without institutional affiliations to purchase access on an individual basis.

Examples: Proquest, Medline, Web of Knowledge, subject-specific databases, online journals
Library catalogs

Almost all public and academic libraries allow anyone to search their catalog on the web. This allows you to discover what books they may own, or journals they may subscribe to. It is rare for a library catalog to include any way to search for individual journal articles.

There are also a couple of “union” library catalogs (catalogs that include the contents of many libraries) online – you may have to be at a library to use them, though.

Examples: Mirlyn (U-M Library Catalog), AADL online catalog, WorldCat (OCLC union catalog)

General things to know about Google

PageRank – what sets Google apart

Just like all other search engines, Google indexes the content of a webpage to figure out what the page is about. Most other search engines rank their results based on a page’s content: the top item on a results list is the item whose text seems to most closely resemble your search. Google, however, sorts its search results using its PageRank technology. The rank for “page X” is based on the number of other pages Google knows about that link to page X. The more pages that link to page X, the higher its PageRank will be. Links from pages with higher ranks count for more than links from pages with lower ranks. It’s all a popularity contest, really.

In the context of the Web, a popularity contest turns out to be a good way to find things. It makes use of the human knowledge that is inherent in the linked structure of the Web; people link to your website when they find something on your site interesting or useful. Each outside link to your page is counted as a “vote” for your page. Of course, Google counts many layers of “votes” in their page rank technology – it’s not just who voted for you, but who voted for them, and who voted for them, and so on.

PageRank and Academic Searches

The “popularity contest” aspect of PageRank can be a concern when doing academic searching. In a general Google search, results related to an academic concept may be far less popular than results with a commercial or entertainment focus, so you may have trouble tracking them down. It is clear that PageRank technology is applied to the results of standard Google searches. It’s much less clear whether and/or how this idea is being applied to results from Google Scholar or Google Books searches.
Issues to consider…

**Personally identifiable information**

Any for-profit company providing free services does usually want something from you in return. You should be aware that like most commercial websites, Google does track your search queries, which pages you visit, and so on. This is most obvious on search results pages, where Google generates advertising related to your search. It is wise to be aware of this when searching, but Google does not seem to associate the information about your behavior with personally identifiable information (though they do seem to track your IP address.)

Some other services that Google provides, like Gmail, or viewing some pages in GoogleBooks - any that require registration, or that require that you install software on your computer - may well associate information about your behavior with information about who you are personally. Google’s stated corporate philosophy is “Don’t Be Evil”, and many people feel they more than live up to this philosophy. However, as Google continues to extend their services, other individuals have raised concerns about their business practices and content ownership. Both pro- and anti-Google discussions are available on various websites and blogs.

**Advertising**

Google has stated several commitments about advertising. They do not accept money for placement in search results. They offer advertising only when it is relevant to the interests of users. Any advertisements they do show are text-only, not distracting pictures. They have made good on these commitments, but it is good to be aware that some of Google’s business model depends on being able to sell advertising space.

Some Google results contain out-and-out ads, in the form of **Sponsored Links** – these appear only when Google thinks your interests might match up with some of their ads, and are always marked as such. Other results will contain **Product Results**, which are not exactly advertising. Product Results appear only when Google thinks you might have been looking for a product for sale. They link to results from Google’s online shopping tool, Froogle, or from websites selling products. You may do many searches that result in no advertising at all, but other searches may bring up quite a bit.
A Deeper Look at the general Google Search

Search tips

*Did you mean?*

If you enter a search query that is misspelled, Google may offer some alternatives to your spelling. While the “Did you mean:” feature is very useful, it does *not* catch all spelling mistakes, especially those which are common enough to appear repeatedly on the web!

*Phrase Searching*

When you input a series of words, Google always treats them as separate words, but it will guess at common phrases. You can force Google to search for words together by putting quotes around them.

*Stop words*

Google has many *stop words* and *stop phrases* - words that are so common that it prefers not to search for them. Examples include “the”, “an”, “of”, and “how do you”. *When Google throws words or phrases out of your search, it may not tell you that it has done so.* Most likely, you will simply see that some of your search words are *not* highlighted below your search box – these words were dropped out of the search.

*Requiring or excluding words*

You can force Google to search a word or phrase (overriding the decision to throw out your stop words) by placing a *plus sign* before the word or phrase. You can also use the *minus sign* to exclude words or phrases from your search. The search `disney biography` returns many biographies of Walt Disney, but few from other folks. `disney biography –walt –walter` brings a much wider variety of results at the top of a shorter list.

Advanced Searching

The Advanced Search Page can be reached by clicking “Advanced Search” from the Google home page. Use the pulldown menus to build complex searches, to search for specific languages, dates, file formats, and much more. Similar functionality is available from the standard Google search page, using Advanced Search Operators. These are simple codes and symbols that allow you to build more complex searches (the plus and minus signs, for example). See [http://www.google.com/help/operators.html](http://www.google.com/help/operators.html) for a complete list.

Additional Google Databases

Google has lots of specialized databases, including an image search, a shopping tool (Froogle), and a maps and local business search. All of these, and much more detail about basic Google searching, are discussed in the “Google Unleashed!” class (register online at [http://www.lib.umich.edu/exploratory](http://www.lib.umich.edu/exploratory)) and handout (available online at [http://www.lib.umich.edu/knc/howto/ibasics.html#google](http://www.lib.umich.edu/knc/howto/ibasics.html#google)).
Google Scholar aggregates a variety of web-based academic information, and largely focuses on information available on the public web. Their own description of included sources lists: “peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations.”

Search Tips

Starting a search
Most of the simple search operators outlined above for a general Google search, also work for Google Scholar. To search for an author name or article title, simply enter it in quotes.

Stop Words/Requiring and Excluding Words
In Google Scholar, any words from your search that were ignored by Google as too common are highlighted below the search box on the results page:

You can require and exclude words and phrases using the plus (+) and minus (-) signs just as in general Google search. Putting a plus sign in front of a word or phrase will override Google’s decision to treat it as a stop word.

Library Links
Google Scholar uses OpenURL technology to provide links to fee-for-service online resources owned by libraries. If you are on the University of Michigan campus, Google will automatically detect this from your computer’s internet address, and provide links to online resources from our libraries.

If you are working from outside the University of Michigan campus, the availability links will not appear automatically. You must authenticate as a U-M user to see them. This is best done by starting your search through the Library website. Visit the Electronic Resources page at http://www.lib.umich.edu/eresources and put Google into the search box to get the link to Google Scholar.
Advanced Searching

Google Scholar has an Advanced Search page just as the general Google search does. You can reach it by clicking “Advanced Scholar Search” next to any search box. This page lets you specify particular combinations of words and phrases, date ranges, publication titles, and even subject areas.

Additionally, many of the Advanced Search Operators available for a general Google search work in the Google Scholar search box.

Preferences

Clicking on “Scholar Preferences” brings you to a page where you can set long-term preferences (they are stored in cookies, and so will last only on this browser and computer.) You can choose how many items per page of results, or set up from which library to see linked results.

Citation Linking

Google Scholar provides counts of citations for articles. If article X has a “cited by” link, click it to see other articles that cited article X.

It is entirely unclear exactly what is included in Google’s count of citations. They do track citations to articles that they cannot find on the public web (as evidenced by the non-clickable “Citation” search results listings.) But where exactly they’re drawing their boundaries is not
known. Citation counts in Google Scholar should be regarded as estimates of popularity and importance; if a page has a lot of cites in Google Scholar, it’s probably at least somewhat influential.

For serious citation linking and tracking, try the ISI Web of Knowledge resource offered by the library. It’s the online version of the old print citation indexes, and it’s well-tested and reliable. It gives more accurate counts of how many times something has been published within peer-reviewed scholarly literature, and clearly delineates what is being counted, what is excluded, and doesn’t duplicate any citations.

### Alternatives to Google Scholar

The fee-for-service databases provided by the library provide richer scholarly content than Google Scholar. They include a great deal of content that Google cannot index, and many have the online full-text of articles. A good place to start is the Library’s Networked Electronic Resources page, at [http://www.lib.umich.edu/eresources](http://www.lib.umich.edu/eresources).

For general searches, several all-purpose databases exist. **Proquest** is a great place to “get your feet wet” on a topic, and includes scholarly, industry, and popular literature; usually, it has full-text access and page image views of articles. Proquest also provides full-text and page image access to **Historical Newspapers** (New York Times, Washington Post, and others) as far back as the 1800s.

**Web of Knowledge** (mentioned above under Citation Linking) also has broad subject-area coverage, and full-text for many articles. It contains only scholarly literature, and in many disciplines, indexes back quite a few years.

For higher-level research, you may wish to target your searching in **subject-specific databases**. Examples would include **Medline** (medical and biological sciences research), **PsycInfo** (psychology and related fields), and **Inspec** (engineering). Enter discipline related words (or parts thereof) into the search box on the Networked Electronic Resources page to find subject-specific databases. You can also try **SearchTools**, a new service of the University Library, to find databases in your area, and to search across multiple databases simultaneously.

For more help selecting subject-specific tools, contact the librarian responsible for your subject area. A full list of University Library subject specialists is available at [http://www.lib.umich.edu/collections/specialists.html](http://www.lib.umich.edu/collections/specialists.html)
Google Books contains full-text searchable page images from books that have been submitted by their publishers, and from books that have been scanned from library holdings (including many from the University of Michigan Libraries.) You can search for a title, author, or just on a topic. Your search results will tell you what book the search terms were found in, and show you snippets of relevant text.

Click the title of a book to go to a page all about that book. You may be able to view some or all pages immediately, or you may be asked to sign in to view pages. To see all relevant search results within that book in “snippet” form (your search terms highlighted in context), click the “More results from this book” link. “About this book” gives you a synopsis or abstract of the book, and links to online book reviews, and more information about the book, publisher, etc.

For books provided through corporate partnerships between Google and publishers, there are usually links to various sources where you can buy the book. For books provided through partnerships with libraries, Google also includes a link to “Find this book in a library”, which takes you out of Google to a public interface for WorldCat (one of the union catalogs mentioned earlier.)

**Advanced Book Search**

Many of the advanced search operators from the general Google Search work in the Google Books search box. You can also click on the “Advanced Book Search” link next to all search boxes to visit the Advanced Search Page.
From this page, you can restrict by date, publisher, search ISBNs, and so forth. This is very helpful when searching for a particular edition of a book.

**Alternatives to Google Books**

Library catalogs provide a great way to search for books, and you have a little better knowledge about what books might be included. In a public library catalog, for example, you would expect to be able to find fiction and non-fiction books aimed at a wide variety of readers, and newspapers and magazines of general interest. In an academic library catalog, you would expect to find literature and academic non-fiction, as well as newspapers from a variety of sources, and scholarly journals. Union catalogs allow you to search the holdings of many libraries at once. From the U-M Networked Electronic Resources page, you can access multiple union catalogs. Two major union catalogs are: **WorldCat** (through OCLC, represents the holdings of most libraries in North America and some in other parts of the world), and **RLIN** (through RLG, materials held by major research and academic libraries in North America).

There are also other places to search and browse freely within the text of books.

**Amazon.com** allows you to search and browse inside many books (wherever the publisher has agreed to such searches and provided the necessary information; it’s very similar to the publisher-provided books on Google Books).

**Project Gutenberg** ([http://www.gutenberg.org/](http://www.gutenberg.org/)) has been around since the earliest days of the Internet, and currently provides free full-text access to 17,000 books (in many languages) that are out-of-copyright in the United States. Anyone with Internet access may use Project Gutenberg. Books are provided as plain text files, or sometimes as PDF page-image files.