Searching the Internet

An OLLI Course in 4 weeks

by **Lewis Overton**
Revised for Fall Session 1, 2016

```
University of Alaska Fairbanks
OSHER Lifelong Learning Institute
UAF Summer Sessions
```

Day 1 Introduction

Scope of the course

This course will explore searching of the internet. Please be aware that the art and science of searching extremely large data sources is one of the oldest, most studied, and complex topics in all of information technology. Basic searching techniques have become simple to perform. Anyone who has used a computer connected to the internet has performed searches, perhaps without knowing they are doing so. I will introduce extensions beyond basic searching, but in four classes of well under two hours each, I can only point out some places where you can explore in depth. That exploration is up to you, depending on your interest and your perseverance.

The instructor (hereafter known as "I", "me", or "Lewy") will use his (I'm male) personal MacBook Pro running Mac OS X *El Capitan*, version 10.11.5 for examples and demonstrations. However, the principles of internet searching apply to any operating system including Windows. A Windows computer is available if people want to see how things are done using that OS. I have not used Windows in years, so do not expect amazing, wizardly tips.

This course draws heavily on work by Wendy Boswell, published at **http://goo.gl/apdQTS** (from **https://about.com**). Exploring this link will lead you down a long trail of breadcrumbs about many aspects of internet searching.

We will be looking at three levels of the internet: the Surface Web, the Deep Web, and the Dark Web.The surface web is the part that is searchable from sites like Google and similar search engines, which accumulate web addresses by crawling through thousands of web sites and building an index of words found at those sites. The Deep Web refers to online databases and other dynamic pages that cannot be indexed by common search engines like Google. These sites contain data which are accessed through sites that provide access to details. The Dark Web contains data which people are trying to hide. Those people may be drug dealers, gun dealers, and others who may have nefarious reasons for keeping you out. We will discuss the dark web, but approach it with caution or, even better, stay away. Here be dragons.

Limitations:

Access to the internet is available from our classroom, but it is unlikely that everyone can connect at the same time. I will use a separate connection so that I don't get locked out by network overload.

Most of the capability to search the internet is dependent on software running on a remote computer. The search capabilities are only marginally related to to the browser you use. Your choice of browser will be touched on only lightly and is mostly decided by personal taste and ease of use. My primary browsers are Chrome and Firefox. I occasionally use Safari or Tor, either of which may be chosen for capabilities unique to what they do in specific circumstances.

Conventions

I have provided a **Notes File** which contains text from the lecture part of the class, links to sites with relevant information, and suggestions for search terms on specific topics. The Notes File is available from the OLLI Web server (see below).

The Notes File is provided in PDF (Portable Document Format), which should be readable on any computer. This file is intended to relieve you of the need to write down links and comments from class.
I **strongly encourage** you to download the Notes File and save it on your computer. The text is provided in PDF format (Portable Document Format). All of the searches, downloads, and supporting materials can be used from any location where you have access to the internet. I suggest you work with the Notes File from home. Most browsers will take you to the destination website if you click on the link. Work with the PDF file, not a print out.

Using links to additional information

Links to web sites should appear in blue text or in bold text. Which depends on which editor I use to convert to PDF. I presently have less control of that than I like. We take what we are served. There are two formats for links: the full link or a shortened link. For example, **access to the OLLI server** is available in two formats. Both forms will take you to the same place. Most browsers will allow you to click on a link to go there. Links to other sites may take one of two forms: short form or long form.

I usually use the short form for convenience which looks like this:

> **https://goo.gl/Y4G4lI**

The long form looks like this:

> **https://uaf.edu/olli/classes-and-lectures/reference-materials-for-s/**

Either form of these links will take you to the same internet site. The short form is provided to avoid what are often very long internet addresses. I will often use the short form

Search for information

You may also encounter a set of search terms which will help you to search the internet for multiple sites related to our discussions. The prefix **Search for:** will be followed by a few keywords. Copy the keywords listed, paste them into the search field of your browser, and execute the search. This should return a list of web sites on our current topic. Browse through some of them for more information on the topic.

Home Tasks

1. Download the Notes File from the OLLI server and save it on your computer. When you are working with the file, use the saved PDF version, not a printed copy. The Notes File is small and should not take long to download. The OLLI site, by the way, is part of the deep web which is deliberately made available to us.
2. Visit the page at **http://goo.gl/ZTnBNl** . This page is from About.com/tech on the topic of searching. Read or at least look through the different pages on searching.

DAY 2 The Surface Web

Introduction

Boswell describes three basic types of search tools that people use to find what they are looking for on the Web. (Boswell, **http://goo.gl/ZTnBNl**, updated 20 Feb 2016)

There are three basic types of search tools that most people use to find what they are looking for on the Web (there's more than this, but these are the basics that everyone should start with):

Search Engines - index of selected pages
Subject Directories - index by topic
MetaSearch Tools - index drawn from multiple search engines

None of these search tools allow you to search the entire Web. That would be an almost impossible task. However, you can use these Web search tools to scour different parts of the Web, obtain different types of information, and broaden your Web search horizons.

Search the Web with Search Engines

Search engines are proprietary tools which build indexes toa large number of web sites and provide a public link to access the index. They discover sites to index using web crawlers (spiders), entering those sites, and collecting key words which identify the themes of the sites. Google is a well know search index, You type in a keyword or phrase and the search engine retrieves pages that correspond to your search query.

> Search results gathered from these search engines are not always relevant to the keywords entered, since these engines are not intuitive and cannot infer dynamically what it is you might be searching for (although results are getting better all the time). Although the spiders run 24/7, it takes a long time to cover the surface web. Information from such search indexes is not always current.

The simplest form of searching is to type a plain language question in the search bar of your browser and press return. An alternative is to go to a website such as Google which offers a web interface to the Google Search Engine. Note that Google is but one such web interface to proprietary search engines.

Some browsers allow you to specify which search engine you want to use for any default searches. This makes it easy to use a favorite search engine without going through a web page to use it. Both Google and Firefox browser offer this option. Other browsers may offer the same kind of choices. I haven't tried all the browsers, of course. There are far too many.

You can visit **http://goo.gl/CxQmfj** for a list of 10 different search engines. You can also continue to scroll down the page for related information and topics. I will visit this site in class to show you what it is like.

Add Filters to your searches

Search Engines have been adding Artificial Intelligence (AI) or Machine Intelligence (MI) to the software involved in the process. A simple query is often sufficient, but may produce hundreds of responses to the query. Fortunately, it is possible to add refinements to reduce the responses. This practice is commonly known as "filtering" and is dependent on which search tool you are using. Filtering with Google will be different from filtering with Firefox or with DuckDuckGo.

To see a list of filters **for Google searches**, go to **https://goo.gl/6aP09x**. Remember that these filters apply only to Google. In addition, Google has discontinued use of the + symbol to require that a search item is required to be present in the results. Other search engines often use a similar set of codes, including + symbol. Look for filter definitions for search engines you may use.

Day 3 Subject Directories

Use Subject directories to focus your search

There are many directories covering limited, specialized searches. For example, you may find directories which index sites that focus on books, or free books, directories about gardening, or … well, pretty much any topic or subject in which you are interested. Some of these directories are accessible in the Surface Web. Others can only be found through the Deep Web.

Examples

Public records

Government in Alaska is good about making public information available by internet.

Actions within the courts, which includes both history and pending cases, is online at
http://www.courtrecords.alaska.gov/eservices/home.page.2

The Sex Offender and Child Kidnapper registry lists people who have been convicted or sex related crimes at
http://www.dps.state.ak.us/cdvsa/Resources-Registry.html

The Property Database lists information about real property including owner, location, business name, property tax data and other information about a particular address. Find this at
http://old.fnsb.us/assessing/propsearch.aspx

An overall listing of free public records including search by town of zip code is available at
http://publicrecords.onlinesearches.com/Alaska.htm

The Noel Wien Public Library offers many online services, including paid access databases. Look for more information at the library. Their home page explains and provides access to many services:
http://fnsblibrary.org/

Federal Databases

There is a metadata collection of publicly available web sites by topic. It is extensive and often useful. Find it at
http://publicrecords.onlinesearches.com/Alaska.htm

The CIA World Factbook offers basic information about the countries of the world including population, economy, and demographic information. The contents are public information.
https://www.cia.gov/library/publications/the-world-factbook/

General Access

Most people are probably familiar with Wikipedia. There are a number of Wiki-whatever or whatever-pedia sites. Be careful (as always) with the credibility of sources referenced and the reliability of the data.
https://www.wikipedia.org/

The Online Tech Dictionary from Webopedi provides a lookup for technical terms about computing at
https://goo.gl/ClInBA

There are many other dictionaries online, including, of course:
https://www.oxforddictionaries.com/us

This is but a sample of what is available from widely available databases. Most can be found via the Surface Web because they are registered with search engines. There are many more than I can count, much less demonstrate.

The rest of the story

Only a portion of the Internet is accessible from typical search engines like **Google**, **Duck Duck Go**, and other forms of search engines which crawl the web and attempt to index what they find. This method returns a large amount of information which is enough for casual usage, but for serious research purposes, the approximately 500 times more web sites are available through the Deep Web than through the Surface Web.

The Deep Web

The Deep Web comprises web sites which are not found by search engines and are not publicized over the internet. In general, this means that the usual process of collecting and indexing these web sites is difficult or impossible by web crawling (spider) technology such as Google and other easily available indexing search engines. Web developers may also attract attention from web crawlers either by registering their sites or by including key words in the home page of their sites. Search Engine Optimization (SEO) has become a booming business in Information Technology. SEO analysis helps web developers attract business to a web

site by mathematical analysis of terms which people may use to bring others to a site. Their objective is to make money from selling their service. Buyers hire their services to increase the number of people who visit a site, bringing revenue to their clients through advertising. This is especially important for sites that are not easily discovered by indexing web crawling technology such as online databases, libraries and library catalogs, specialized sites, or proprietary internet sites. These may not be interesting or even accessible to the general public, but in many cases they will be a rich source of information not so easily located through the surface web.

The Dark Web

There is also a portion of the web dedicated to criminal activities. It is sometimes called the Dark Web, which may contain sites related to criminal activity. I urge **EXTREME CAUTION** before you venture into the Dark Web. **Even better,** avoid this area of the internet completely unless you know **exactly** what you are doing. In particular, be prepared with high-tech security and anonymity tools to protect yourself and your identity. For more information on the Dark Web, I recommend exploring the history of the Silk Road. Start with the wikipedia link **https://en.wikipedia.org/wiki/Silk_Road_/(marketplace/)**. From there you will find ways to learn more about the Dark Web. Procede with care. At least consider using Tor browser to help hide the link back to your computer if you attempt to access the Dark Web. The Wikipedia site mentioned above is safe. Entering the Dark Web is not.

Evaluations

As usual in OLLI classes, we ask that you fill out an evaluation of this course and the instructor. I use these to adapt future courses that I may teach and to build on the experiences of my students and myself. I pay attention, including suggestions for future courses. Anonymous evaluations are fine, but, if you want, you are welcome to include contact information and I will contact you about the course and your opinions.