Basic Google Hacking

Prerequisite: This document assumes that you are familiar with the use of a computer keyboard and mouse, have a working knowledge of Microsoft Windows and that you are familiar with using the World Wide Web. It also assumes some familiarity with using a web browser such as Netscape or Internet Explorer.

Contents

Purpose and Objectives

This document aims to provide a new experience in using different search techniques, including advanced search features, in major web search engines.

When you have completed these exercises, you should be able to:

• Understand how to use various search strategies on search engines, such as phrase searching and truncation, as well as use of Boolean operators, including but not limited to, Fields and/or Advanced operators.

• Make effective use of web search engine features and functions.

Copyright

Disclaimer

This document: "Basic Google Hacking" provides readers with information about various hacking tricks that would help them from being hacked. Since this document provides such controversial information on hacking, it is a must to implement privacy policy. We do not promote nor sponsor Hacking in any form. Rather, this document will help you gain entry into the minds of seasoned computer criminals, so that you would be able to forestall their attempts and pre-empt any harmful attacks. Hence, you will be well equipped to detect ways in which hackers can infiltrate your system.

IMPORTANT: This information is for research and academic purposes only! This info is not to be abused! We are not responsible for any damage that you may create!

Contact Information

Email: founder@security-science.com

Website: www.security-science.com

I. Including & Excluding Terms

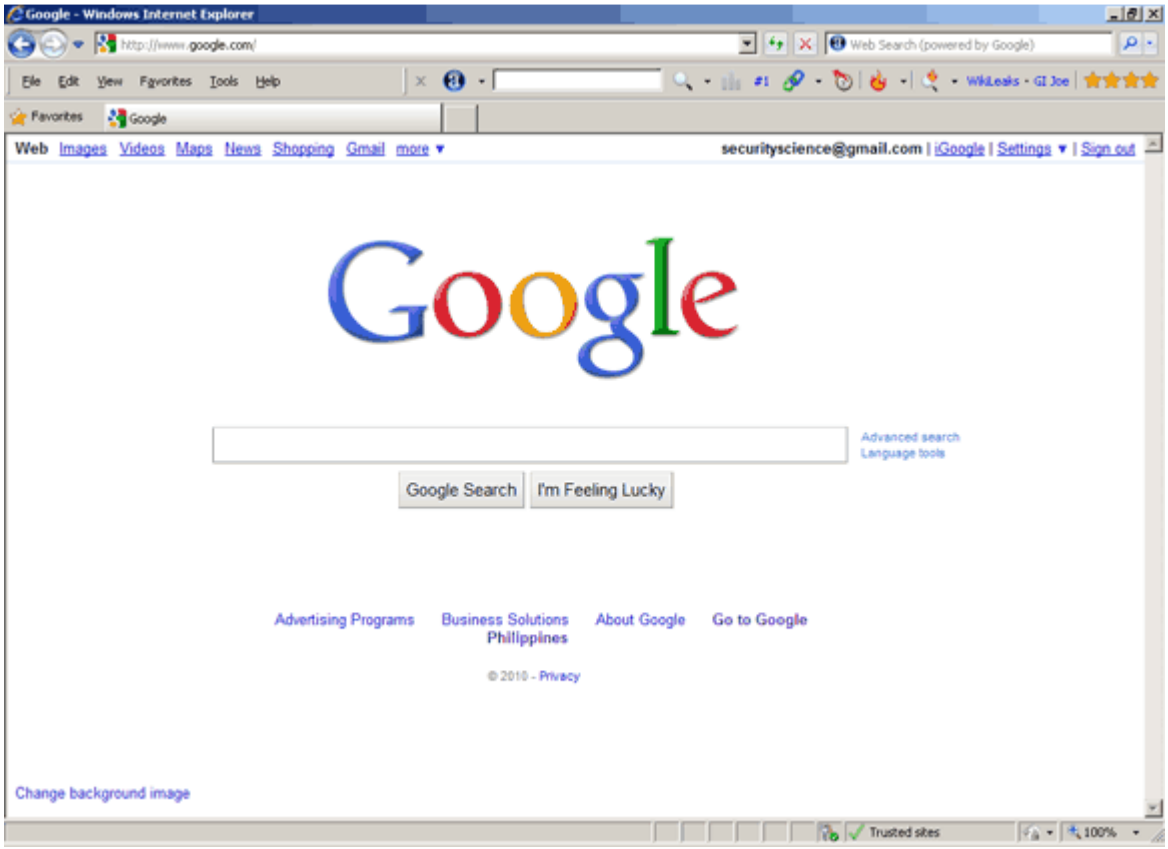
Objective: To learn how to include or exclude terms from your search engine results.

Method: You will access a search engine and use the plus and minus signs to control your search.

Comment: Often, when you enter keywords into a search, it is difficult to know what exactly the search is doing with those keywords – is it searching for results containing all the terms, or just one? Simple search methods can ensure you can control your search.

Go to the Google home page at the address:

http://www.google.com



In the search box, enter the following terms:

`internet security science` or `+internet +security +science`

Click on `Google Search` and make a note of the number of results returned:

• Notice that this search has retrieved both upper and lower case versions of these search terms – Google is ***not*** case-sensitive.

• You don't need to use a plus (+) sign in front of your terms, as Google automatically searches for both terms.

To exclude a word from a search the minus sign can be used. Try the following search:

```
internet security –science
```

This should reduce the number of results and make them more relevant to the `internet security`.

Note: Be careful when excluding words – because search engines index so much text, you may unwittingly exclude results that are relevant, but which just happen to mention your exclusion term in passing.

Now try the phrase search:

```
"internet security science"
```

The phrase search retrieves the words placed between the quotes in the order you have specified, exactly next to each other.

Using a phrase search should have once again reduced the number of results retrieved quite dramatically and improved the relevance. In this example, `internet security science` is a phrase, and a phrase search would be the best technique to try initially.

Note: Phrases can also be included or excluded by prefixing them with the plus or minus signs.

II. Phrase Searching

Objective: To target a search more effectively using phrase searching and wildcards.

Method: You will use the Google search engine to carry out a phrase search.

Comment: Phrase searching is a type of proximity search. It requires that the words you enter as a phrase are contained within your search results *in the exact order you have specified*.

`internet security science` vs `"internet security" science`

Using Google, enter the following search:

```
internet security science
```

Press the `Google Search` button and review the results – how relevant do you feel they are? Are there any unexpected results?

See if you can work out why you might be retrieving irrelevant results.

Note: Notice that this search does not use the plus signs to require words. This is because in the previous task you established that Google's default is to require all words be contained in results, so the plus signs are not necessary.

Now enter the following search:

```
"internet security" science
```

Examine your results (make a mental note of the number of results retrieved).

You should find that the results are more relevant – this is because Google is now searching for "`internet security`" as a phrase, i.e. only returning results where the words appear next to each other in that order.

Note: You should also find that the number of results has decreased. This is because fewer pages are likely to contain "`internet security`" as a phrase.

Try the following search:

```
to be or not to be
```

Examine your search results – do they look at all relevant to Shakespeare? Several of the words in the above quote are in fact **stop words** – words that are ignored in the search.

Look at the words that have been highlighted in your results – you should find that only the word ***not*** is highlighted, and this is the only word that has been searched for.

Scroll further down the page – you should see Google has actually run a second search for you automatically: a search on "`to be or not to be`".

These results should be a lot more relevant to Shakespeare

Note: You can force Google to search for stop words, either by phrase searching, or by using the plus sign in front of a stop word.

III. Field Searching

Objective: To understand how to use field searching options to search more effectively.

Method: You will use the title search at Google.

Comment: Field searching can be useful in searching the web as it enables you to narrow a search down to specific parts of a web document, such as a title.

Try the following search:

```
internet security science
```

Make a mental note of the number of results returned. Review the listing and you will find that most sites are already relevant to search engine tutorials, but you can increase relevance by restricting the search to the title of the web page.

Now try the search:

```
allintitle:internet security science
```

This time you are searching for all of the words in the title of a page (the **title** of a web page is the text that appears in the bar at the top of the browser window). The title is also listed as the main heading to each site in the search results list. You can see the search terms highlighted in the title in the figure below.

You should find that you retrieve fewer results and the results should be more targeted to your search term.

Notes: It is likely that a document which contains your term in the **title** will be more relevant for you than one which just mentions your keyword somewhere in the body of the document.

Try out some of the following search options:

`link`: Restricts the search to hypertext links in documents. It can be useful if you want to find out how many sites link to a particular page. For example, `link:www.security-science.com`

`intitle`: This will search for any of the words in the title of the web page. This differs to `allintitle`: used above where *all* words must be in the title.

`inurl` or **`allinurl`**: Searches for a particular URL (URL = Uniform Resource Locator = web address) or part of a URL on a web page. For example, `inurl: internet-security-science` will retrieve web addresses containing the term `internet-security-science`.

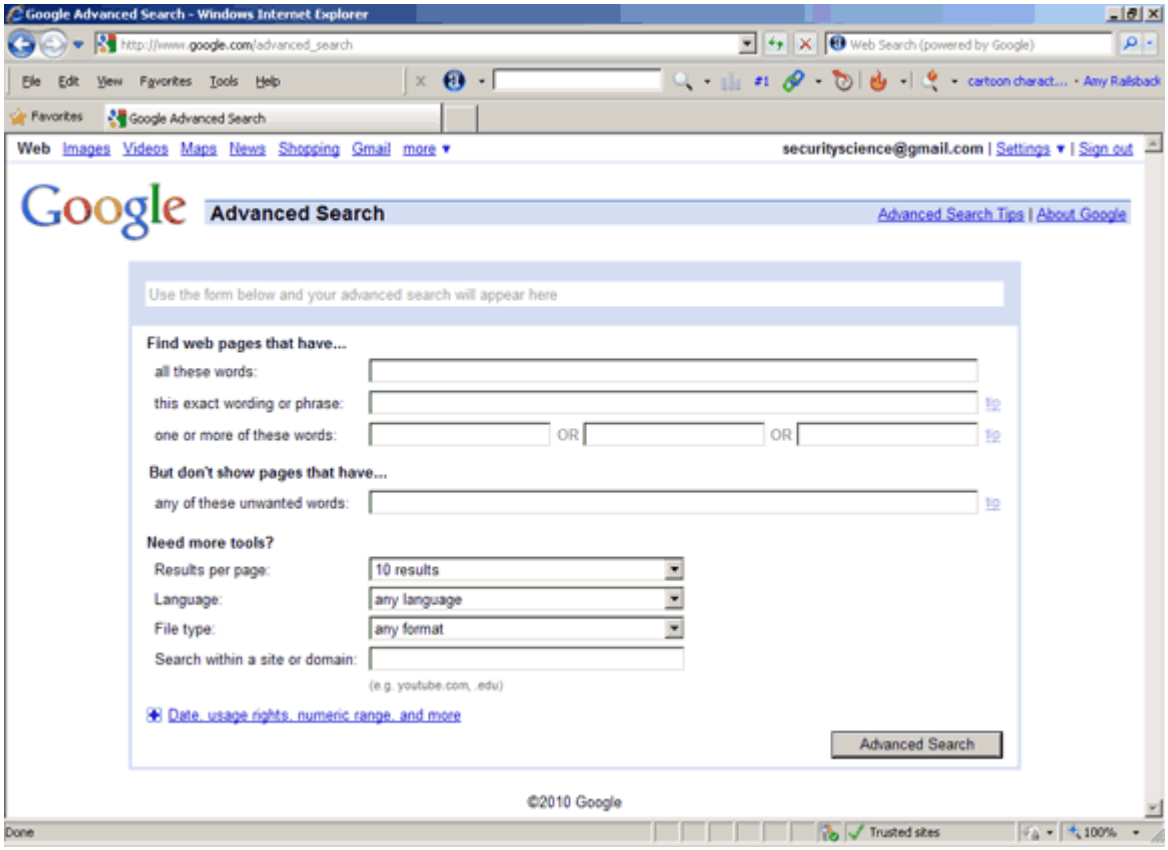
IV. Google Advanced Search

Objective: To explore the advanced search features at the Google search engine.

Method: You will use the advanced search page at Google.

Comment: Google generally works well using simple search queries, but it does offer several advanced search options to help you construct and restrict your search.

From the main Google home page, select the link to `Advanced Search`.



Google's advanced search features are relatively new to the site and are continually being developed and expanded.

Under the `Advanced Web Search`, examine the different search options available. This search interface allows you to do a limited type of Boolean search operation using a form instead of search syntax.

Notice also that there are options to restrict by `Language`, `File Format`, `Date` and `Domain`. The `Occurrences` option allows you to restrict your search to certain parts of the web page, for example `title` or `URL`.

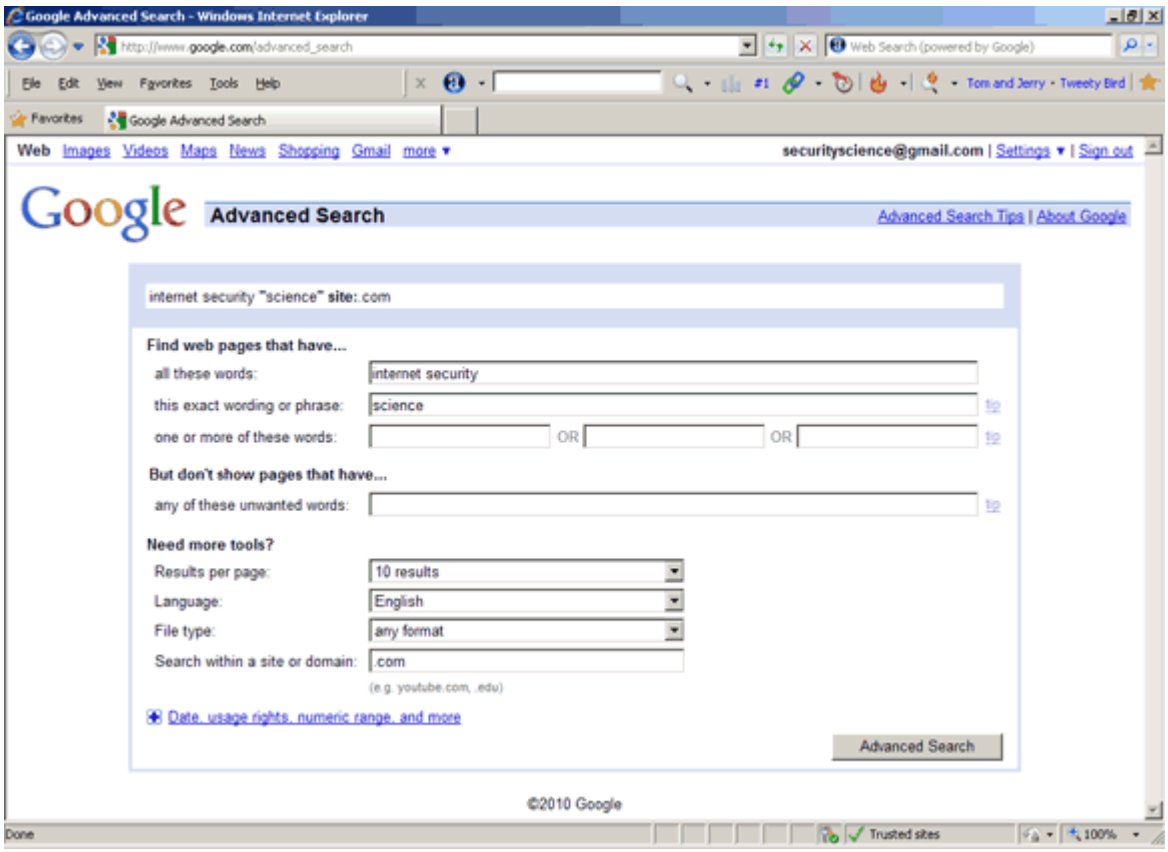
Use the form to enter the following search:

Find results `with all of the words`: **`internet security`**

Find results `this exact wording or phrase`: **`science`**

Need more tools? `Language`: **`English`**

Need more tools? `Search within a site or domain`: **`.com`**



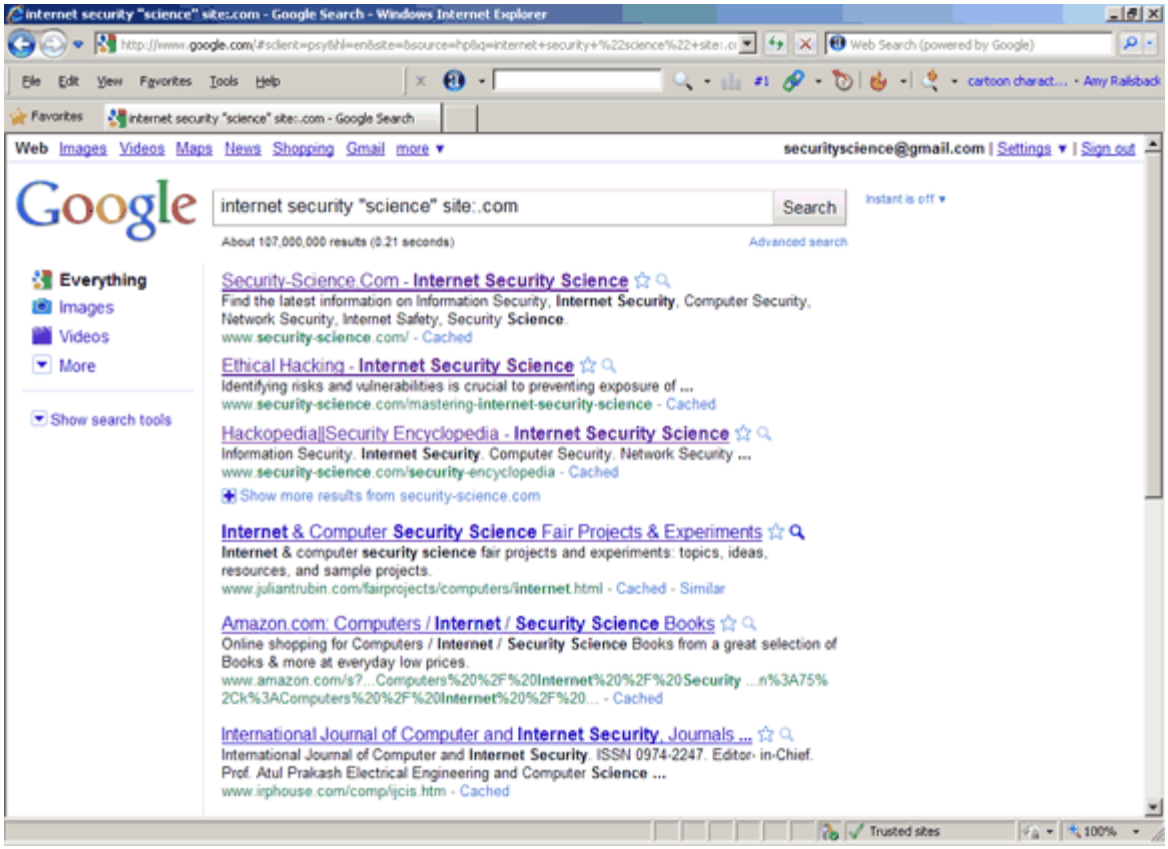
When you have entered the search, press the `Google Search` button.

Examine your search results.

You should find that your search retrieves `.com` sites which have titles containing the word `internet security` and with exact word `science`.

Note: On your search results page, examine the search box. Google has automatically entered the search query it has constructed from the information you entered in the forms.

```
internet security "science" site:.com
```



Notice that this query doesn't need the `AND` operator – this is because by default, Google automatically combines search terms using `AND`.

Go back to the `Advanced Search` page and experiment with some of the other search options.

V. Boolean Searching

Objective: To learn how to structure a Boolean search query.

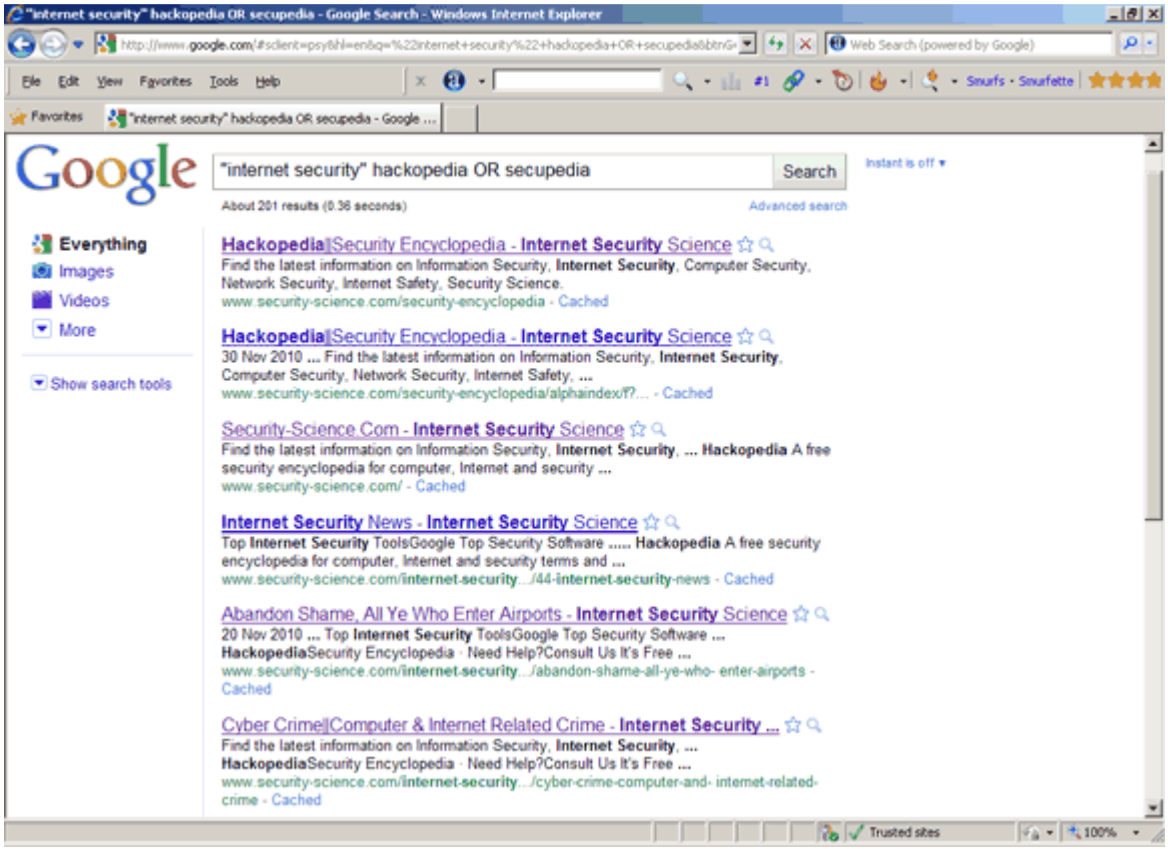
Method: You will use Boolean search features at Google.

Comment: Boolean operators and brackets can be used to create complex search queries. Not all search engines support full Boolean searching.

Google does not support full Boolean searching using the `AND`, `OR` and `NOT` operators and brackets. It does offer the `OR` operator which can be used with the plus (+) and minus (-) signs.

In the Google search box, enter the following search:

```
"internet security" hackopedia OR secupedia
```



The search should retrieve `internet security` for either `hackopedia` or `secupedia`.

Note: It is very important that the `OR` operator is entered in capital letters – otherwise it will be treated as the normal word *or* which is a stop word and as such will be ignored.

Reference: The Search Engine Showdown site contains a good overview of how Boolean can be used at Google:

http://www.searchengineshowdown.com/features/google

Complex Boolean searches cannot be carried out on Google as it does not support the full range of features. AltaVista supports full Boolean.

Go to AltaVista UK at:

http://www.altavista.com

From the options to the right of the search box select `Advanced Search`.

VI. Search Engines: What's Under the Bonnet?

Objective: To understand which search engine databases are unique, and which are owned by the same company.

Method: You will go to the Searchenginewatch.com site and learn more about search engines.

Comment: Understanding which database is underlying each search engine can help you quickly identify which search engines are the best alternatives to the one you regularly use.

In your web browser go to:

http://searchenginewatch.com/2156401

This will take you to the SearchEngineWatch "Who Powers Whom? Search Providers Chart". The chart explains the (rather complicated!) nature of commercial search engines – read through the introductory text that explains the chart.

Chart Key

Search Providers: These are listed at the top of each column. Read down to see what they power at major search engines. Click on their names to learn more about them.

Search Engines: These are listed at the beginning of each row, in order of share of searches shown on the comScore Media Metrix Search Engine Ratings page. Here's a guide to the color coding:

- **Dark Orange**: search engines with 25 percent or greater share.
- **Light Orange**: search engines with 10 percent or greater share.
- **Light Blue**: search engines with 1 percent share or greater share.
- **Gray**: search engines with less than 1 percent share. They are shown only because of the name recognition they may still have among some long-time searchers or marketers.

Main: Indicates that a search provider provides the "main" editorial results to a particular search engine, the most dominant listings that will be seen.

Paid: Indicates that a search provider provides paid placement listings to a particular search engine. Also see the Buying Your Way In page for detailed information about paid listing partnerships.

Backup: Indicates that a search provider provides the "backup" results that appear in cases where a search engine's main results fail to find good matches. See the Search Engine Results Page for more about "backup" or "fallthrough" results.

Option: If shown in the notes section, Indicates that information from this source is made available either on results pages or in other ways, though the prominence of the information may not be high.

Dates: Where shown, dates indicate when a particular partnership is due for renewal. Dates are shown in MM/DD/YY or similar format.

Search Engine (Read Down)	Provider: Google	Provider: Yahoo/Overture	Notes
Google	Main & Paid		Open Directory an option
Yahoo		Main & Paid	Yahoo Directory an option
MSN		Main & Paid (12/05 & 6/05)	
AOL	Main & Paid (est. 10/05+)		AOL-owned Open Directory an option
Excite Network	Main & Paid (at iWon, MyWay, My Web Search)		Excite.com is InfoSpace-powered
Ask Jeeves	Paid (until 2007)		Main from Ask-owned Teoma.
InfoSpace	Runs several meta search engines. Dogpile is most popular, representative of others. Google (2006), Yahoo (3/06) & many small providers have distribution deals.		
AltaVista		Main & Paid	Open Directory an option; owned by Yahoo

Examine the chart - key points are that:

• At the time this chart was last updated, there were only 2 main search engine database providers – Google and Yahoo! Before February 2004, Yahoo! used to use Google's database, but after buying Inktomi, AltaVista and AlltheWeb, it launched its own database.

• Although Google and Yahoo! are the major providers, there are some others, including Teoma (owned by Ask). See the Notes next to each search engine.

• As well as providing normal search engine results, Google also provides paid listings to other search engines (its Sponsored Links). search engines, for example Teoma, may use Google's paid listings while using their own database for their main web listings.

• Yahoo's paid listings are provided by Overture, a company which it also owns.

Note: Although the same search engine database may underly 2 different search engines, there may offer different search features. For example, Yahoo! owned AltaVista offers full Boolean searching, which is not offered at Yahoo!.

VII. Google Bomb and PageRank

Objective: To understand Googlewashing and how Google determines page relevance and rating.

Method: You will go to the wikipedia.com site and learn more about Google Bomb and PageRank.

Comment: Understanding Google bomb or Google wash and PageRank can help you quickly formulate which technique or strategy is best to implement to your site.

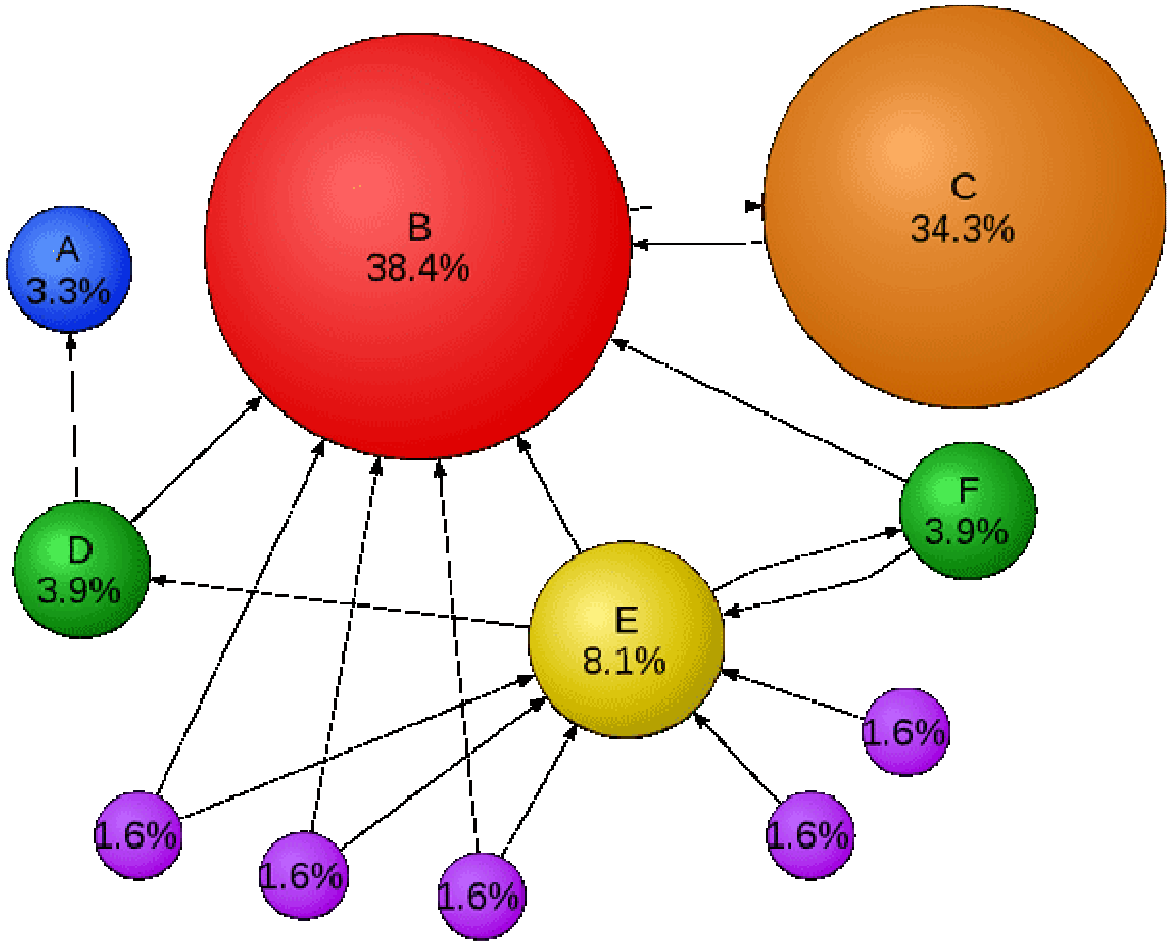
A **Google bomb** or **Google wash** is an attempt to influence the ranking of a given site or particular pages in results returned by the Google search engine in order to increase the likelihood of people finding and clicking on selections in which the individual or other entity engaging in this practice is interested. It is done for either business, political, or comedic purposes (or a combination of the latter two). Due to the way that Google's Page Rank algorithm works, a website will be ranked higher if the sites that link to that page all use consistent anchor text.

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

Google describes PageRank:

"**PageRank** reflects our view of the importance of web pages by considering more than 500 million variables and 2 billion terms. Pages that we believe are important pages receive a higher PageRank and are more likely to appear at the top of the search results.

PageRank also considers the importance of each page that casts a vote, as votes from some pages are considered to have greater value, thus giving the linked page greater value. We have always taken a pragmatic approach to help improve search quality and create useful products, and our technology uses the collective intelligence of the web to determine a page's importance."



Mathematical PageRanks (out of 100) for a simple network (PageRanks reported by Google are rescaled logarithmically). Page C has a higher PageRank than Page E, even though it has fewer links to it; *the link* it has is of a much higher value. A web surfer who chooses a random link on every page (but with 15% likelihood jumps to a random page on the whole web) is going to be on Page E for 8.1% of the time. (The 15% likelihood of jumping to an arbitrary page corresponds to a damping factor of 85%.) Without damping, all web surfers would eventually end up on Pages A, B, or C, and all other pages would have PageRank zero. Page A is assumed to link to all pages in the web, because it has no outgoing links.

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

Search Engine Optimization (SEO) is probably the most important aspect of website promotion. He refers to different methods by which your site receives higher rankings in search engine to use. There are many techniques for optimizing your website for search engines, while some of them are from search engines and promoted as a white hat techniques are known, there are other techniques, the black hat on your site the search engine can be blocked. *White Hat SEO* to ensure the correct use of keywords and metadata, good link building, proper use of *ALT* tags refer, in other words, a site that integrated keep people in mind and not the search engine robots is.

Content is king for Website Promotion: There is no doubt that good content will automatically invite people to link to your website and help with website promotion. You have to keep it interesting, informative and fresh, or else lose the visitor interest in your site and not come back. Remember also to avoid non-keyword spamming. He used several times refers to the use of keywords in a page only for the purpose of *SEO*. This technique is often difficult to understand the content. Search engines can tell when this is happening and not enter the website importance. The accepted keyword density is up to 4-5%.

Link Building for Website Promotion: Link Building is an important step in website promotion, offering a high return on investment that you in the form of money, time and effort. The first thing to look into is Directory Submissions. There are four different types of categories that you can be listed on your site – free directories, niche directories, reciprocal links and paid directories. Apart from these directories, you can also participate in link exchange programs, but do not forget to watch out link farms. Another good way to get one way links to your website is through Article Writing. You can write good articles that are appropriate for your industry and remember your company's information and link at the end of the article to add. Please send your articles to various article directories to get good connections a way back to your website.

Internet Marketing: You can also opt for internet marketing as part of your website promotion campaign. Apart from the above-mentioned search engine optimization, Internet marketing also means search engine marketing, e-mail marketing, affiliate programs and banner advertising.

http://www.webcrutch.com

So What Determines Page Relevance and Rating?

• **Exact Phrase**: are your keywords found as an exact phrase in any pages?

• **Adjacency**: how close are your keywords to each other?

• **Weighting**: how many times do the keywords appear in the page?

• **PageRank/Links**: How many links point to the page? How many links are actually in the page?

Equation:

```
PR = (Exact Phrase Hit)+(AdjacencyFactor)+(Weight) *
(PageRank/Links)
```

VIII. Use Search Engine as Hacking Tool

Objective: To understand how search engines work.

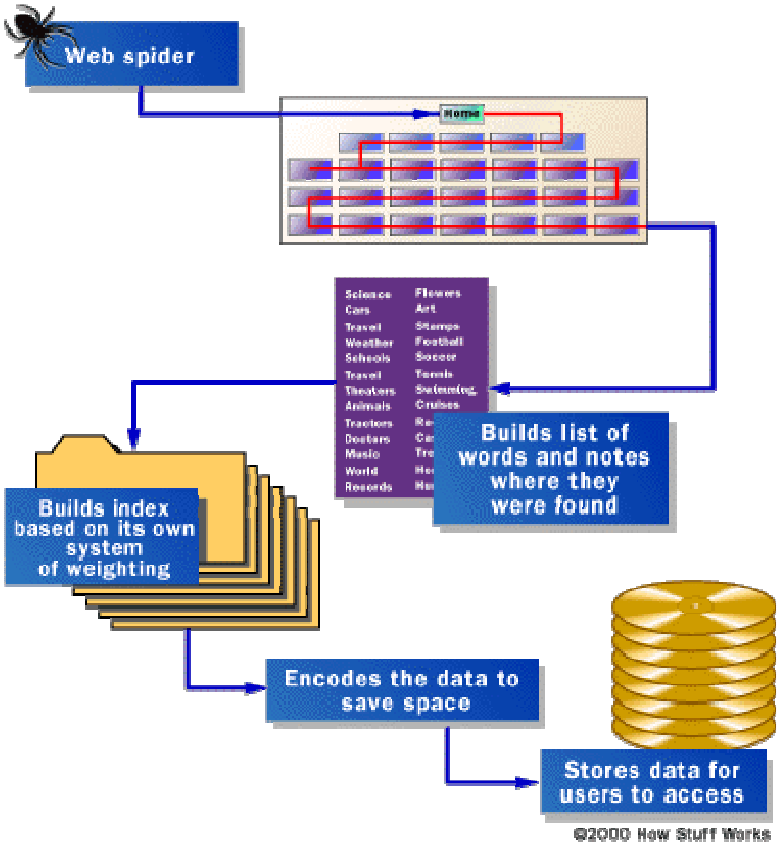
Method: You will use Google to practice basic hacking or conduct reconnaissance phase.

Comment: Understanding how search engines work will help you get the information you want to gain.

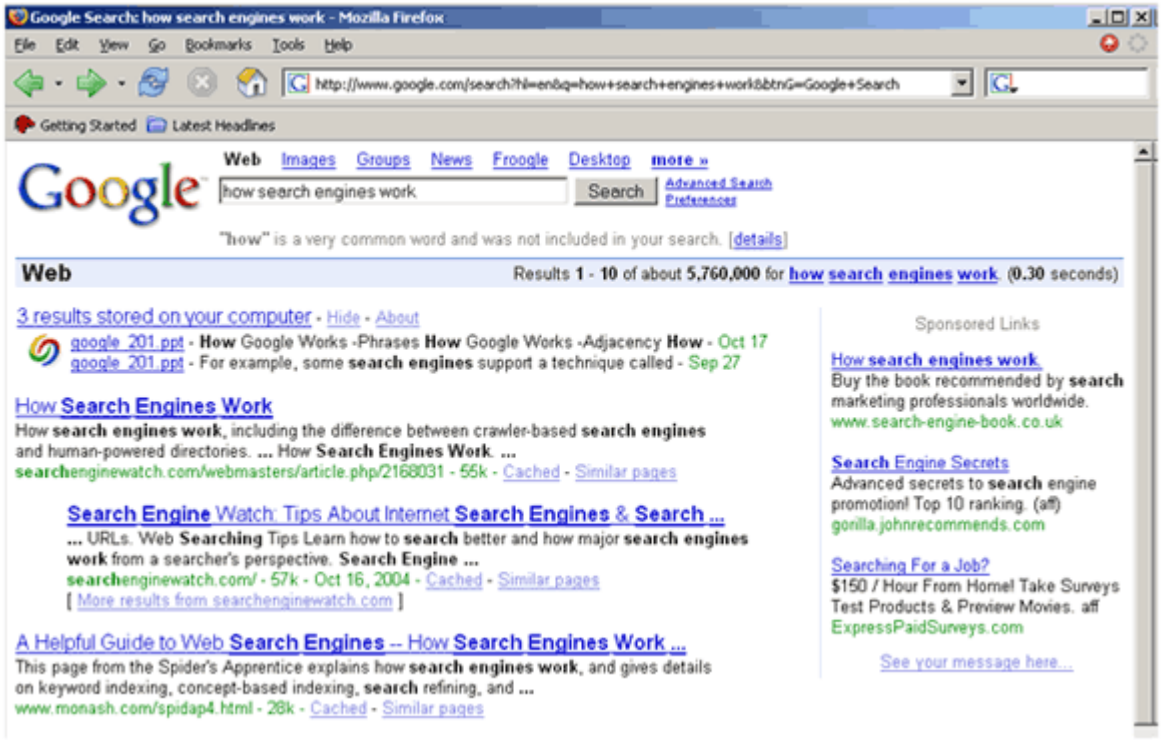
Anatomy of a Search

"Spiders" take a Web page's content and create key search words that enable online users to find pages they're looking for.

Server Side



Client Side



http://computer.howstuffworks.com/internet/basics/search-engine1.htm

Does Google finds pages that are only connected web pages indexed?

• NO! – Opera and Chrome submits every URL viewed to Google for later indexing….

How Do I Get Results?

• Pick your keywords carefully & be specific

• Do NOT exceed 10 keywords

• Google ignores some words*:

a, about, an, and, are, as, at, be, by, from, how, i, in, is, it, of, on, or, that, the, this, to, we, what, when, where, which, with

Advanced Search Operators

Search Service	Search Operators
Web Search	allinanchor:, allintext:, allintitle:, allinurl:, cache:, define:, filetype:, id:, inanchor:, info:, intext:, intitle:, inurl:, link:, phonebook:, related:, site:
Image Search	allintitle:, allinurl:, filetype:, inurl:, intitle:, site:
Groups	allintext:, allintitle:, author:, group:, insubject:, intext:, intitle:
Directory	allintext:, allintitle:, allinurl:, ext:, filetype:, intext:, intitle:, inurl:
News	allintext:, allintitle:, allinurl:, intext:, intitle:, inurl:, location:, source:
Product Search	allintext:, allintitle:

Advanced Operators

• Google advanced operators help refine searches.

Advanced operators use a syntax such as the following:

`operator:search_term`

– Notice that there's no space between the operator, the colon, and the search term.

• The **`site`**: operator instructs Google to restrict a search to a specific web site or domain. The web site to search must be supplied after the colon.

• The **`link`**: operator instructs Google to search within hyperlinks for a search term.

• The **`cache`**: operator displays the version of a web page as it appeared when Google crawled the site. The URL of the site must be supplied after the colon.

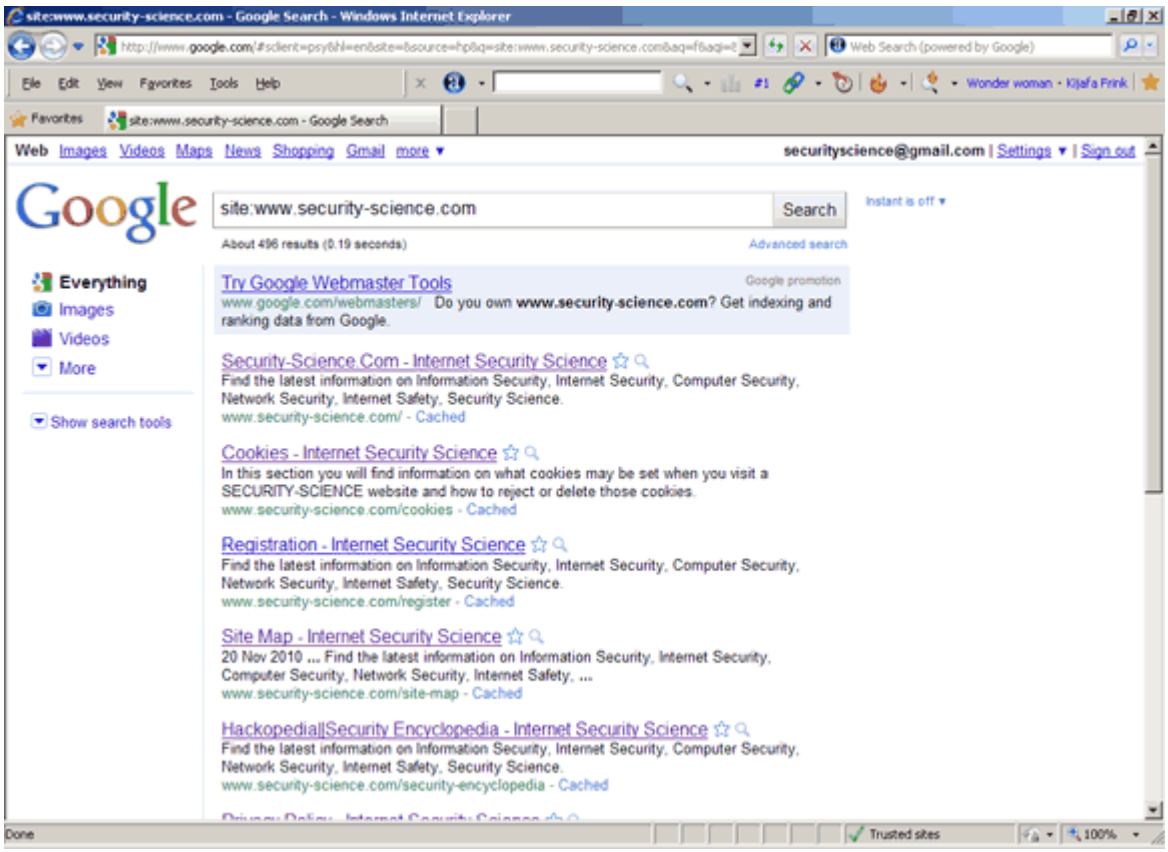
– Turn off images and you can look at pages without being logged on the server! Google as a mirror.

• Google searches not only the content of a page, but the title and URL as well.

• The **`intitle`**: operator instructs Google to search for a term within the title of a document.

• The **`inurl`**: operator instructs Google to search only within the URL (web address) of a document. The search term must follow the colon.

• To find every web page Google has crawled for a specific site, use the **`site`**: operator.



IX. What Can Google Search?

Objective: To learn more advance Google search and how to protect your site from Google hackers.

Method: You will use Google to practice basic hacking or conduct reconnaissance phase.

Comment: Understanding how search engines work will help you get the information you want to gain and how to countermeasure Google hacking.

Google Can Search:

• The `filetype:` operator instructs Google to search only within the text of a particular type of file. The file type to search must be supplied after the colon. Don't include a period before the file extension.

– Everything listed at http://filext.com. In addition, e.g., `filetype:phps` to only search `.phps` files.

• `filetype:phps mysql_connect`

• Adobe Portable Document Format (pdf)

• Adobe PostScript (ps)

• Lotus 1-2-3 (wk1, wk2, wk3, wk4, wk5, wki, wks, wku)

• MacWrite (mw)

• Microsoft Excel (xls)

• Microsoft PowerPoint (ppt)

• Microsoft Word (doc)

• Microsoft Works (wks, wps, wdb)

• Microsoft Write (wri)

• Rich Text Format (rtf)

• Shockwave Flash (swf)

• Text (ans, txt)

• And many more….

Directory Listings

• Useful for an attacker

– `intitle:index.of server.at`

– `intitle:index.of server.at site:aol.com`

• Finding Directory Listings

– `intitle:index.of "parent directory"`

– `intitle:index.of name size`

• Displaying variables

– "HTTP_USER_AGENT=Googlebot "

– Frequently an avenue for remote code execution

```
http://sub.somedomain.com/~user/demo.cgi?cmd=`cat /etc/passwd`
```

Protecting Yourself from Google Hackers

• ***Keep your sensitive data off the web!*** Even if you think you're only putting your data on a web site temporarily, there's a good chance that you'll either forget about it, or that a web crawler might find it. Consider more secure ways of sharing sensitive data, such as SSH/SCP or encrypted email.

• Consider removing your site from Google's index.

http://www.google.com/remove.html

Robots.txt

Use a robots.txt file. Web crawlers are supposed to follow the robots exclusion standard. This standard outlines the procedure for "politely requesting" that web crawlers ignore all or part of your web site. This file is only a suggestion. The major search engine's crawlers honor this file and its contents. For examples and suggestions for using a robots.txt file, see http://www.robotstxt.org or http://www.robotstxt.org/orig.html.

Interesting Google Searches

- intitle:"Index of" passwords modified
- allinurl:auth_user_file.txt
- "access denied for user" "using password"
- "A syntax error has occurred" filetype:ihtml
- allinurl: admin mdb
- "ORA-00921: unexpected end of SQL command"
- inurl:passlist.txt
- "Index of /backup"
- "Chatologica MetaSearch" "stack tracking:"

Number Ranges to find Credit Card Numbers

– Amex Numbers:

300000000000000..399999999999999

– MC Numbers:

5178000000000000..5178999999999999

– visa

4356000000000000..4356999999999999

Listings of what you want

• change the word after the parent directory to what you want

- "parent directory " DVDRip -xxx -html -htm -php –shtml -opendivx -md5 -md5sums
- "parent directory "Xvid -xxx -html -htm -php –shtml -opendivx -md5 -md5sums
- "parent directory " Gamez -xxx -html -htm -php –shtml -opendivx -md5 -md5sums

• "parent directory " MP3 -xxx -html -htm -php –shtml -opendivx -md5 -md5sums

• "parent directory " Name of Singer or album" -xxx –html -htm -php -shtml -opendivx -md5 -md5sums

Music

• You only need add the name of the song/artist/singer.

```
intitle:index.of mp3 jackson
```

CD Images

• You can change the string to whatever you want, ex. Microsoft to Adobe, .iso to .zip etc…

```
inurl:microsoft filetype:iso
```

Passwords

• FrontPage passwords… very nice clean search results listing!

```
"# -FrontPage-" inurl:service.pwd
```

Passwords in the URL

```
"http://*:*@www" domainname
```

This is a query to get inline passwords from search engines (not just Google), you must type in the query followed with the domain name without the .com or .net

`"http://*:*@www" gamespy` or `http://*:*@www"gamespy`

Another way is by just typing

```
"http://bob:bob@www"
```

IRC Passwords

- `"sets mode: +k"`

This search reveals channel keys (passwords) on IRC as revealed from IRC chat logs.

- `eggdrop filetype:user user`

These are `eggdrop config` files. Avoiding a fullblown discussion about eggdrops and IRC bots, suffice it to say that this file contains usernames and passwords for IRC users.

Access Database Passwords

- `allinurl: admin mdb`

Not all of these pages are administrator's access databases containing usernames, passwords and other sensitive information, but many are!

DCForum Passwords

- `allinurl:auth_user_file.txt`

DCForum's password file. This file gives a list of (crackable) passwords, usernames and email addresses for DCForum and for DCShop (a shopping cart program(!!!). Some lists are bigger than others, all are fun, and all belong to googledorks.

MySQL Passwords

- `intitle:"Index of" config.php`

This search brings up sites with "config.php" files. To skip the technical discussion, this configuration file contains both a username and a password for an SQL database. Most sites with forums run a PHP message base. This file gives you the keys to that forum, including FULL ADMIN access to the database.

The ETC Directory

- `intitle:index.of.etc`

This search gets you access to the etc directory, where many, many, many types of password files can be found. This link is not as reliable, but crawling etc directories can be really fun!

Passwords in backup files

- `filetype:bak inurl:"htaccess|passwd|shadow|htusers"`

This will search for backup files (*.bak) created by some editors or even by the administrator himself (before activating a new version). Every attacker knows that changing the extension of a file on a web server can have ugly consequences.

Serial Numbers

- Let's pretend you need a serial number for Windows XP Pro.
- In the Google search bar type in just like this

```
"Windows XP Professional" 94FBR
```

- The key is the 94FBR code.. it was included with many MS Office registration codes so this will help you dramatically reduce the amount of 'fake' sites (usually pornography) that trick you.
- If you want to find the serial for WinZip 8.1

```
"WinZip 8.1" 94FBR
```

X. Further References

Objective: To be aware of useful sites on the internet to learn about web searching.

Method: You will explore a list of references.

How Google Works

http://www.googleguide.com/power_works.html

Google Extras…

http://www.google.com/help/features.html

Advanced Operators

http://www.googleguide.com/advanced_operators.html

Searching and Evaluation, from University of Cal at Berkeley

http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html

Search Engine Showdown (review of search features), run by Greg Notess

http://www.searchengineshowdown.com

Search Engine Watch

http://www.searchenginewatch.com

Including "Search Features Chart" from the following Web Search Tips:

http://www.searchenginewatch.com/facts/

Sherman, Chris. Why Search Engines Fail

http://searchenginewatch.com/2160661

Best Search Tools Chart

http://www.infopeople.org/search/chart.html

7 Stupid search mistakes

http://www.searchenginewatch.com/searchday/article.php/2159561