MOREnet
Community • Technology • Results

Google Hacking

Beth Young
Kris Trower
MOREnet Security
security@more.net

Agenda

- Google Hacking introduction
- Automating searches
- Building queries
- What to expect from your results
- Controlling how your content is indexed
- References

What is Google Hacking?

Google hacking is an advanced search technique that could allow someone to find sensitive data or vulnerabilities on any site indexed by a search engine.

What benefit does it provide?

An attacker

Anonymous profiling of your network/organization through publically accessible information

- Documents and data on your website
- Configuration and network information that you may have posted on a mail list or forum
- Device and software information that you may have posted in job openings
- Information about your employees, students and patrons that could be used in a social engineering attack
- Misconfigurations and vulnerabilities on your server

What benefit does it provide?

You

The ability to find potential exposures and vulnerabilities and correct them!

Is Google Hacking unique to Google?

No!

Each search engine will allow you to perform advanced searches by using their own unique search operators.

- Every search engine has the ability to index the same pages and documents that Google indexes. Most of them do.
- Any problems you find with Google probably exist in the indexes of the other search engines.
- Undesirable results will need to be addressed with each search engine independently.
- Look for the ‘advanced search’ or ‘custom search’ help pages.

MOREnet
Community • Technology • Results

Building Google Queries

General Rules

- Google queries are not case sensitive
- Every word in the query will be used (with a few exceptions)
- Punctuation is ignored (with a few exceptions)
- Google limits query length to 32 words
- Use as few search terms as possible
- Use search terms that are likely to appear in the results you desire.

General Rules

Be as descriptive as you can, but avoid using unnecessary words.

Example – You would like to find information about cross site scripting issues in Apache 2.2x

With unnecessary words: what cross site scripting vulnerabilities exist in apache version 2.2

More concise with better results: apache 2.2 cross site scripting vulnerability

Exceptions: Stop Words and Punctuation

- Commonly used words, called ‘Stop Words’ (for, a, the, on, where, how), may be ignored depending on how Google interprets the query.

- Google ignores some punctuation and special characters, including @ ! , . [ ] / < > ? ; #

- You can force Google to use stop words by using them in a phrase surrounded by quotes, or by preceding the stop word with a + sign.

Wildcards

- The asterisk (*) is used as a wildcard in Google queries.
- The asterisk can replace a single word in a search phrase.
- It can not be used as the representation of a single letter or part of a word.

Example

Proper use: senator * voted * on the * bill (returns stories on various senators and how they voted on various bills)

Improper use: vulnerabilit* (vulnerabilities; vulnerability)

Wildcards

- Wildcards do not count as words in the 32 word query limit.

Example (Quote from Mitchell Kapor)

14 word query: “Getting information off the Internet is like taking a drink from a fire hydrant”

10 word query: “Getting information off * Internet * like taking * drink from * fire hydrant”

Advanced Search Operators

- Exact phrase – Enclose the search terms in quotes

Example: “gone with the wind”

- Ensure a term/phrase is included – Precede the search term with a plus sign

Example: “gone with the wind” +book +”Margaret Mitchell”

+book will make sure all of the results include book information.

+”Margaret Mitchell” would ensure that all of the results include the authors name.

Advanced Search Operators

Exclude a word or phrase – Precede the search term with a minus sign

Example: “gone with the wind” +book +”Margaret Mitchell” -movie -"free book”

-movie will prevent Google from displaying results that mention the word movie. This would be helpful if you were looking for information on the book specifically and didn’t want information about the movie version.

-”free book” would narrow the search down further if you were not interested in links to download the free book.

Advanced Search Operators

The OR (|) operator

- Use to allow either one of several words
- OR must be in all caps
- OR can be substituted with the pipe (|) symbol

Example: “kansas city chiefs” stram OR schottenheimer OR haley
or “kansas city chiefs” stram | schottenheimer | haley

Either search query will return results containing information about either Coach Stram, Coach Schottenheimer, or Coach Haley.

Advanced Search Operators

The site: operator

- Allows you to limit your query to a specific site
- When Google hacking your site, you will use this operator in almost every query!

Example: site:more.net “security awareness”

This query will display all indexed mentions of security awareness on the MOREnet website.

Advanced Search Operators

We will be defining additional search operators throughout the remainder of this presentation. For a comprehensive list of search operators, take a look at the following links!

Advanced Operators
http://www.google.com/intl/en/help/operators.html

Google Guide to Searching (not an official Google publication)
http://www.google.com/notebook/public/10994812642363749232/BDQtXSwoQ4pbI678h

Book preview: Google Hacking for Penetration Testers, Volume 2 (By Johnny Long)
http://books.google.com/books?id=bvB1-MmhEjQC&lpg=PP1&pg=PA50#v=onepage&q=&f=false

MOREnet
Community • Technology • Results

What are you exposing on the World Wide Web?

What is important to you?

Before you can use Google hacking techniques effectively, you need to know what you are looking for. What information should not be visible to the public?

- Student names, grades and vital information
- Employee HR data/Social Security Numbers
- Financial/credit card information
- Usernames and passwords
- Email stores
- Device configuration information
- Database information
- Vulnerabilities on your server
- Spammed forums and message boards
- Unauthorized web sites on your network
- Inappropriate language or content
- Links to your site from inappropriate sites

What would you put on your list?

Not seeing what you expected to see?

- Building successful queries takes time and practice
- If the query doesn’t produce favorable results, try something different
- The examples we give may need to be tweaked to work in your environment

Not getting a result may be a good thing!

- Some of the examples given will be used to find problems that may (or may not) exist
- Many of them may not give you any results
- Not getting a result may mean you don't have a problem… That's good!

Needle in a Haystack

- Be prepared to trudge through a lot of invaluable data to find the ‘good’ stuff.
- If a search returns too much data, be creative in using excludes to filter out the excess.
- Review your query to determine if you can use more descriptive keywords and remove unnecessary words.

Flash vs. Function

- There are hundreds of flashy (but impractical) queries we could show you, but we don't want to waste your time!
- We spent MANY hours looking at various queries and if they didn't produce results on member sites, we are not going to talk about them.
- We will be discussing queries that produce results frequently and consistently.

In the following queries…

- In the query examples that follow, the information that you need to supply will be enclosed in brackets. When you perform the query, always remove the brackets!
- Unless you are wanting to search on a specific subdomain, use your parent domain.

Example of what you will see:

site:[your_domain] filetype:pdf

What the query looks like using our domain:

site:more.net filetype:pdf

What is out there besides web pages?

Performing a query that excludes web page extensions may point out some interesting things!

- Other file types that you may want to search for later (pdf, doc, xls, etc.)
- Subdomains of your web domain
- The directory structure of your web site

Why perform the search?

- To gather information for later searches
- To look for unauthorized directories
- To look for subdomains or directories that should be deleted (such as test pages)
- To look for content that should not be indexed by search engines
- To get a feel for what is out there without having to wade through the web pages

Query example (tailor it to exclude the type of code you use on your site)

```
site:[your_domain] -filetype:html -filetype:htm -filetype:php -filetype:asp -filetype:pl
```

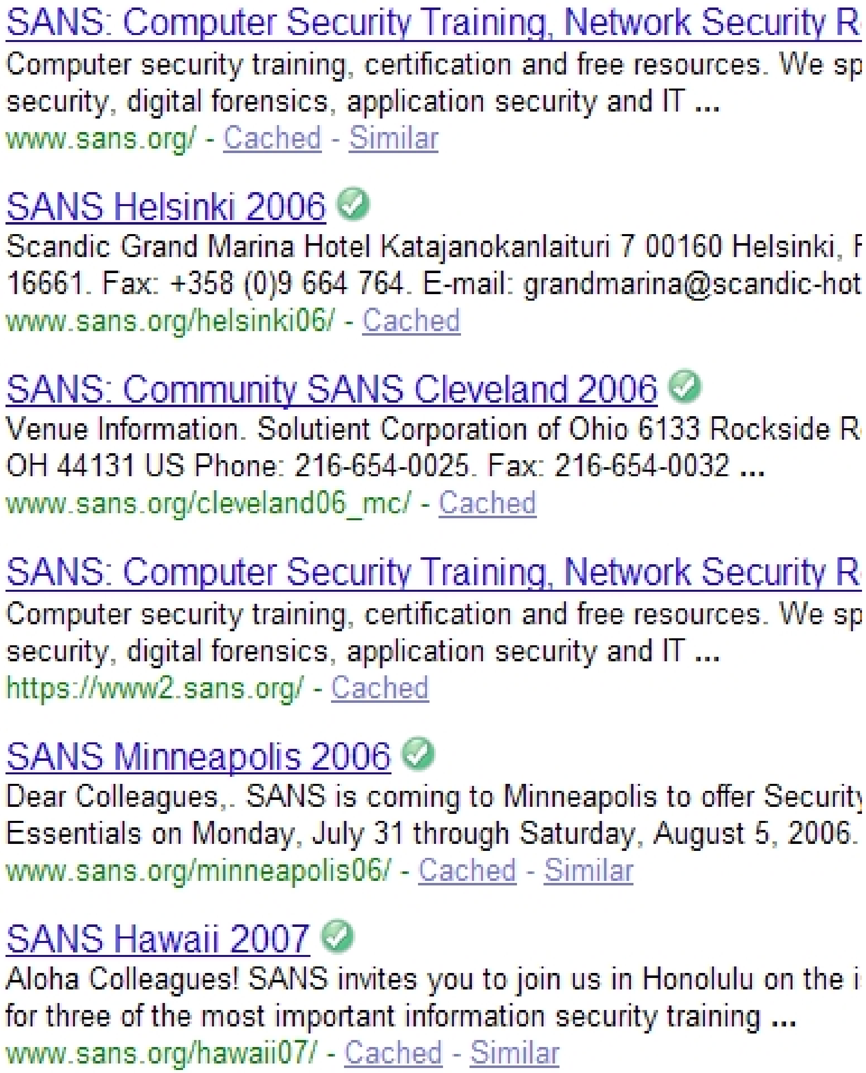
Directory Walking

In this example, the query is helpful in identifying some of the many directories that exist for sans.org.

You will also notice an additional subdomain.

Example

site:sans.org -filetype:html -filetype:htm -filetype:php -filetype:asp -filetype:pl



SANS: Computer Security Training, Network Security R
Computer security training, certification and free resources. We sp
security, digital forensics, application security and IT ...
www.sans.org/ - Cached - Similar

SANS Helsinki 2006 ✓
Scandic Grand Marina Hotel Katajanokanlaituri 7 00160 Helsinki, F
16661. Fax: +358 (0)9 664 764. E-mail: grandmarina@scandic-hot
www.sans.org/helsinki06/ - Cached

SANS: Community SANS Cleveland 2006 ✓
Venue Information. Solutient Corporation of Ohio 6133 Rockside R
OH 44131 US Phone: 216-654-0025. Fax: 216-654-0032 ...
www.sans.org/cleveland06_mc/ - Cached

SANS: Computer Security Training, Network Security R
Computer security training, certification and free resources. We sp
security, digital forensics, application security and IT ...
https://www2.sans.org/ - Cached

SANS Minneapolis 2006 ✓
Dear Colleagues,. SANS is coming to Minneapolis to offer Security
Essentials on Monday, July 31 through Saturday, August 5, 2006.
www.sans.org/minneapolis06/ - Cached - Similar

SANS Hawaii 2007 ✓
Aloha Colleagues! SANS invites you to join us in Honolulu on the i
for three of the most important information security training ...
www.sans.org/hawaii07/ - Cached - Similar

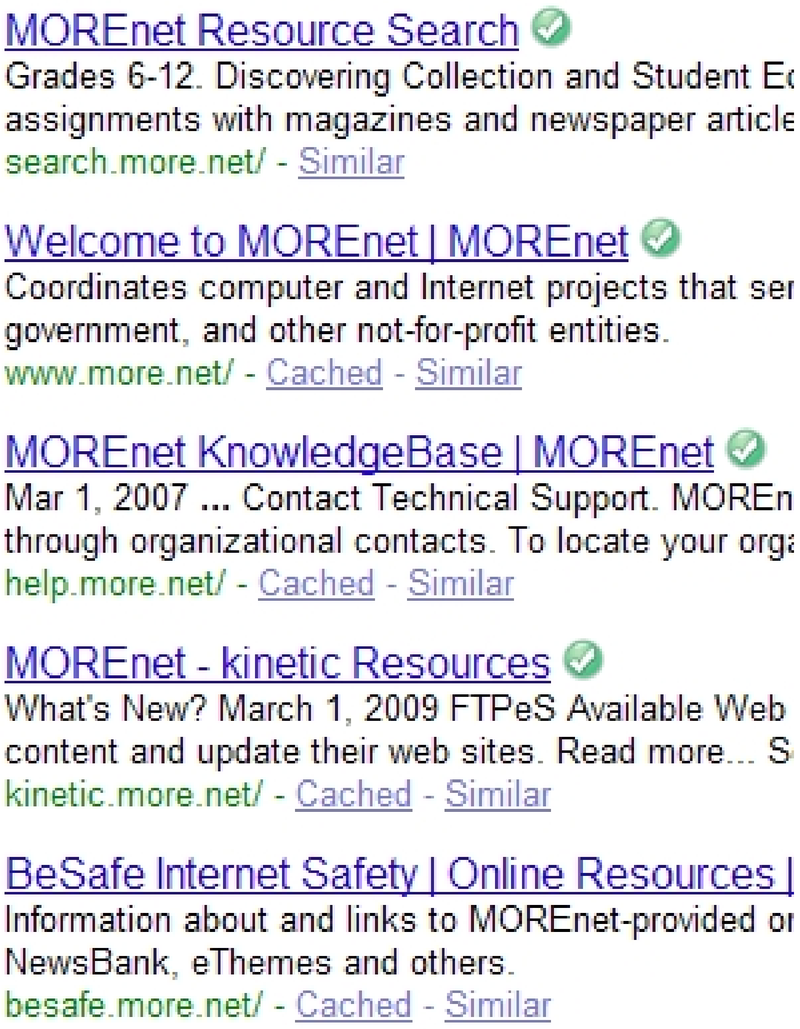
MOREnet
Community • Technology • Results

Subdomains

In this example, the query is helpful in identifying the many subdomains that exist for more.net.

Example

site:more.net -filetype:html -filetype:htm -filetype:php -filetype:asp -filetype:pl



MOREnet Resource Search ✓
Grades 6-12. Discovering Collection and Student E
assignments with magazines and newspaper article
search.more.net/ - Similar

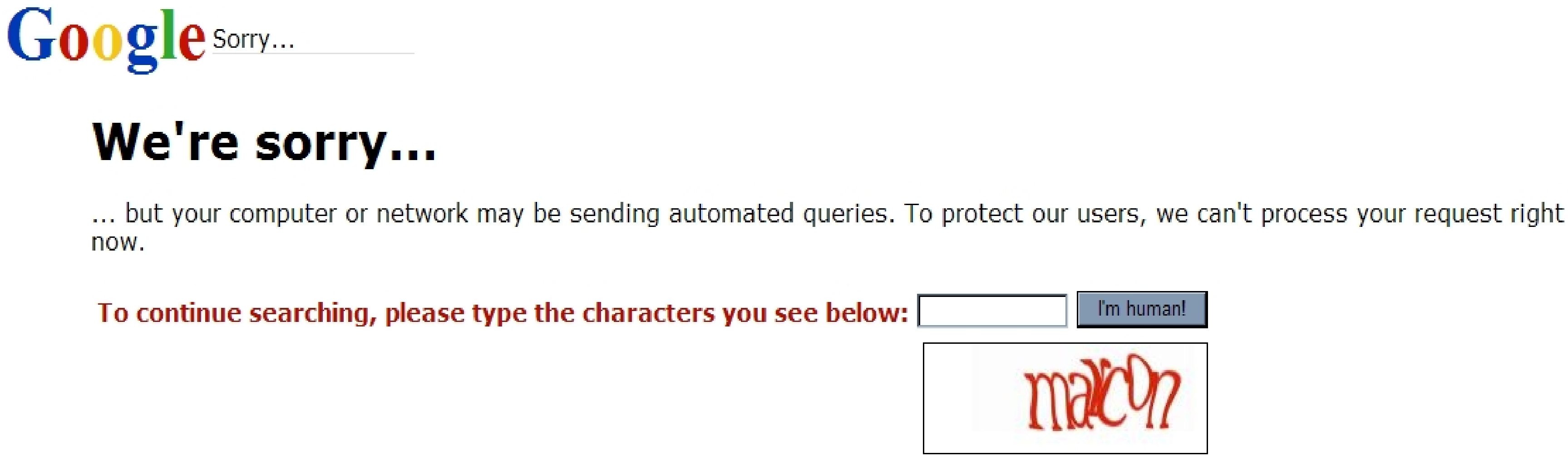
Welcome to MOREnet | MOREnet ✓
Coordinates computer and Internet projects that se
government, and other not-for-profit entities.
www.more.net/ - Cached - Similar

MOREnet KnowledgeBase | MOREnet ✓
Mar 1, 2007 ... Contact Technical Support. MOREn
through organizational contacts. To locate your org
help.more.net/ - Cached - Similar

MOREnet - kinetic Resources ✓
What's New? March 1, 2009 FTPeS Available Web
content and update their web sites. Read more... S
kinetic.more.net/ - Cached - Similar

BeSafe Internet Safety | Online Resources |
Information about and links to MOREnet-provided o
NewsBank, eThemes and others.
besafe.more.net/ - Cached - Similar

oops!



Google Sorry...

We're sorry...

... but your computer or network may be sending automated queries. To protect our users, we can't process your request right now.

To continue searching, please type the characters you see below:

After performing about 10 queries looking for examples, Google presented us with a captcha. It looks like they are becoming more aggressive about looking for automated Google hacking type searches!

MOREnet
Community • Technology • Results

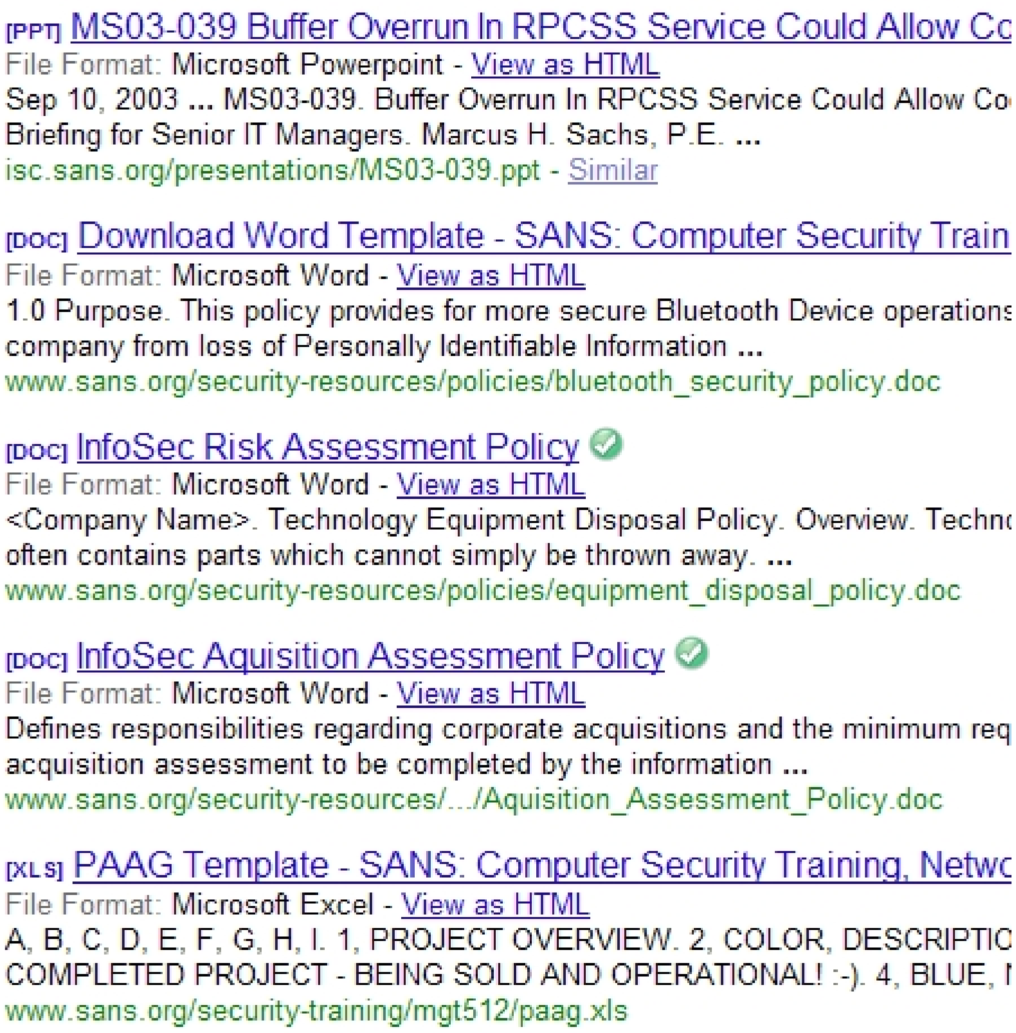
Office Documents and PDFs

Combined Query

site:[your_domain] filetype:pdf OR filetype:doc OR filetype:xls OR filetype:ppt OR filetype:docx OR filetype:xlsx OR filetype:pptx

Example

site:sans.org filetype:doc OR filetype:xls OR filetype:ppt



[PPT] MS03-039 Buffer Overrun In RPCSS Service Could Allow Cc
File Format: Microsoft Powerpoint - View as HTML
Sep 10, 2003 ... MS03-039. Buffer Overrun In RPCSS Service Could Allow Co
Briefing for Senior IT Managers. Marcus H. Sachs, P.E. ...
isc.sans.org/presentations/MS03-039.ppt - Similar

[DOC] Download Word Template - SANS: Computer Security Train
File Format: Microsoft Word - View as HTML
1.0 Purpose. This policy provides for more secure Bluetooth Device operations
company from loss of Personally Identifiable Information ...
www.sans.org/security-resources/policies/bluetooth_security_policy.doc

[DOC] InfoSec Risk Assessment Policy ✓
File Format: Microsoft Word - View as HTML
<Company Name>. Technology Equipment Disposal Policy. Overview. Techno
often contains parts which cannot simply be thrown away. ...
www.sans.org/security-resources/policies/equipment_disposal_policy.doc

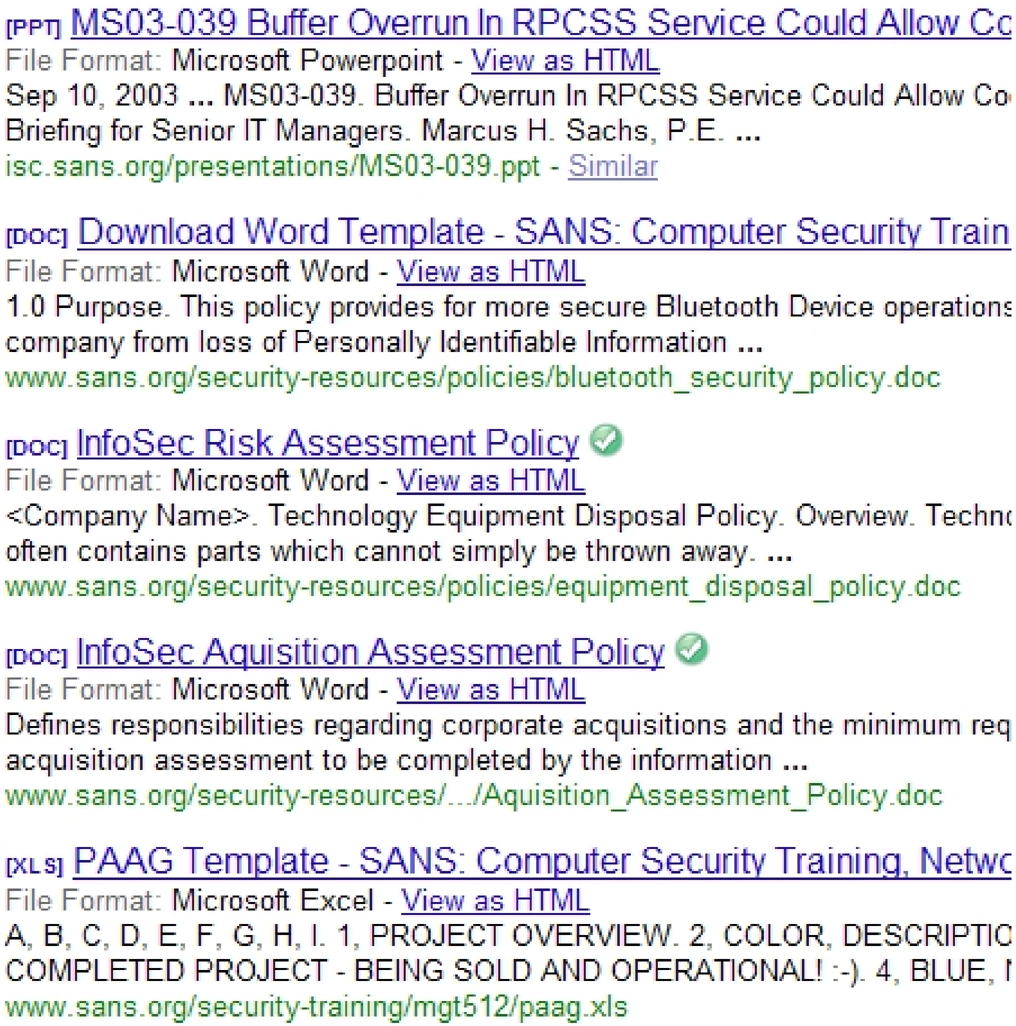
[DOC] InfoSec Aquisition Assessment Policy ✓
File Format: Microsoft Word - View as HTML
Defines responsibilities regarding corporate acquisitions and the minimum req
acquisition assessment to be completed by the information ...
www.sans.org/security-resources/.../Aquisition_Assessment_Policy.doc

[XLS] PAAG Template - SANS: Computer Security Training, Netwo
File Format: Microsoft Excel - View as HTML
A, B, C, D, E, F, G, H, I. 1, PROJECT OVERVIEW. 2, COLOR, DESCRIPTIO
COMPLETED PROJECT - BEING SOLD AND OPERATIONAL! :-). 4, BLUE, I
www.sans.org/security-training/mgt512/paag.xls

We have found (elementary) student vital statistics and grades in xls documents!

Office Documents and PDFs

- These file types are very common and can be found on most web sites
- Sometimes they contain information that shouldn't be publically available
- They can be combined into one query, or broken out individually



[PPT] MS03-039 Buffer Overrun In RPCSS Service Could Allow Co
File Format: Microsoft Powerpoint - View as HTML
Sep 10, 2003 ... MS03-039. Buffer Overrun In RPCSS Service Could Allow Co
Briefing for Senior IT Managers. Marcus H. Sachs, P.E. ...
isc.sans.org/presentations/MS03-039.ppt - Similar

[DOC] Download Word Template - SANS: Computer Security Train
File Format: Microsoft Word - View as HTML
1.0 Purpose. This policy provides for more secure Bluetooth Device operations
company from loss of Personally Identifiable Information ...
www.sans.org/security-resources/policies/bluetooth_security_policy.doc

[DOC] InfoSec Risk Assessment Policy ✓
File Format: Microsoft Word - View as HTML
<Company Name>. Technology Equipment Disposal Policy. Overview. Techn
often contains parts which cannot simply be thrown away. ...
www.sans.org/security-resources/policies/equipment_disposal_policy.doc

[DOC] InfoSec Aquisition Assessment Policy ✓
File Format: Microsoft Word - View as HTML
Defines responsibilities regarding corporate acquisitions and the minimum req
acquisition assessment to be completed by the information ...
www.sans.org/security-resources/.../Aquisition_Assessment_Policy.doc

[XLS] PAAG Template - SANS: Computer Security Training, Netwo
File Format: Microsoft Excel - View as HTML
A, B, C, D, E, F, G, H, I. 1, PROJECT OVERVIEW. 2, COLOR, DESCRIPTIO
COMPLETED PROJECT - BEING SOLD AND OPERATIONAL! :-). 4, BLUE, I
www.sans.org/security-training/mgt512/paag.xls

We have found (elementary) student vital statistics and grades in xls documents!

Numrange Operator

numrange is expressed by placing .. between the two numbers you want to use as the range

Examples

mustang 1966..1969

This query will give you pages that mention Mustangs between the years 1966 & 1969

computer "windows 7" $800..$1500

This query will help you find computers running Windows 7 that are in the $800-$1500 price range

Social Security Numbers

- The numrange operator can be useful in finding Social Security Numbers
- Using numrange to find social security numbers will take some tweaking and some practice
- The ranges that you use will have to be broken down into several smaller queries
- The first three digits designate the state
- The second group of digits indicate the group number
- The third group of numbers is the serial number

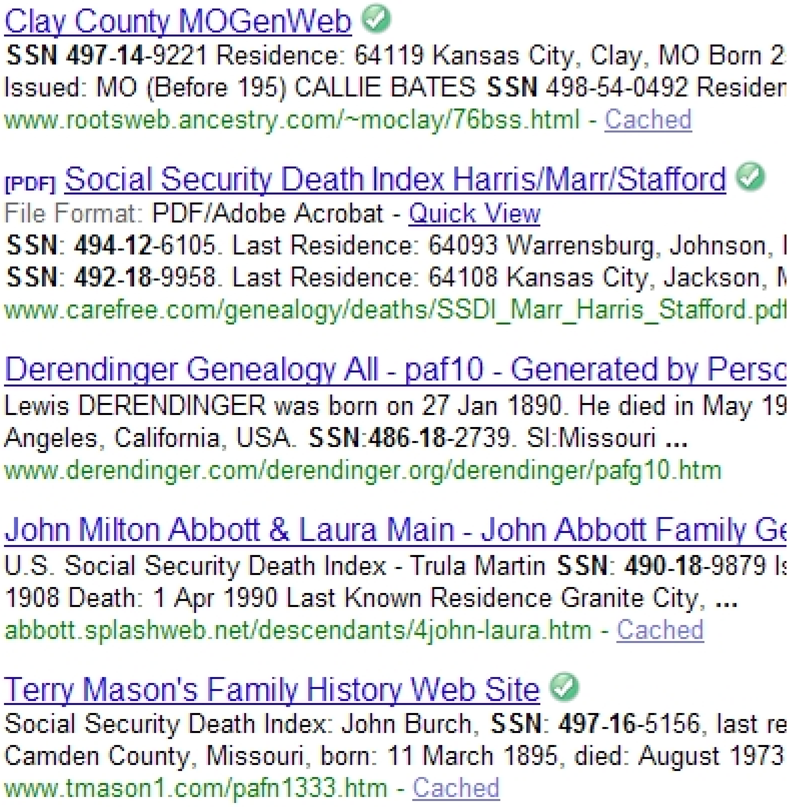
MOREnet
Community • Technology • Results

Social Security Numbers

- Missouri's first three digits are in the range 486-500

- The group number (second set of digits) can range from 01-99

- The serial number can range from 0001-9999



Clay County MOGenWeb
SSN 497-14-9221 Residence: 64119 Kansas City, Clay, MO Born 2
Issued: MO (Before 195) CALLIE BATES SSN 498-54-0492 Resider
www.rootsweb.ancestry.com/~moclay/76bss.html - Cached

[PDF] Social Security Death Index Harris/Marr/Stafford
File Format: PDF/Adobe Acrobat - Quick View
SSN: 494-12-6105. Last Residence: 64093 Warrensburg, Johnson, I
SSN: 492-18-9958. Last Residence: 64108 Kansas City, Jackson, M
www.carefree.com/genealogy/deaths/SSDI_Marr_Harris_Stafford.pdf

Derendinger Genealogy All - paf10 - Generated by Persc
Lewis DERENDINGER was born on 27 Jan 1890. He died in May 19
Angeles, California, USA. SSN:486-18-2739. SI:Missouri ...
www.derendinger.com/derendinger.org/derendinger/pafg10.htm

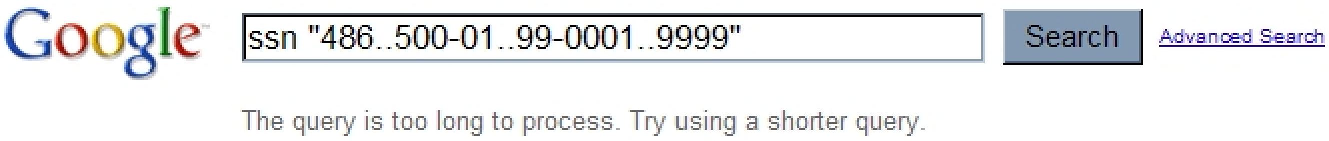
John Milton Abbott & Laura Main - John Abbott Family Ge
U.S. Social Security Death Index - Trula Martin SSN: 490-18-9879 Is
1908 Death: 1 Apr 1990 Last Known Residence Granite City, ...
abbott.splashweb.net/descendants/4john-laura.htm - Cached

Terry Mason's Family History Web Site
Social Security Death Index: John Burch, SSN: 497-16-5156, last re
Camden County, Missouri, born: 11 March 1895, died: August 1973
www.tmason1.com/pafn1333.htm - Cached

Example above:
ssn "486..500-12..23"

Social Security Numbers

As you can see by our attempt to search for all Missouri numbers, the query was too long and it didn't provide good results.



ssn "486..500-01..99-0001..9999"

The query is too long to process. Try using a shorter query.

MOREnet
Community • Technology • Results

Social Security Numbers

Missouri Query (may need to be narrowed down more to be effective)

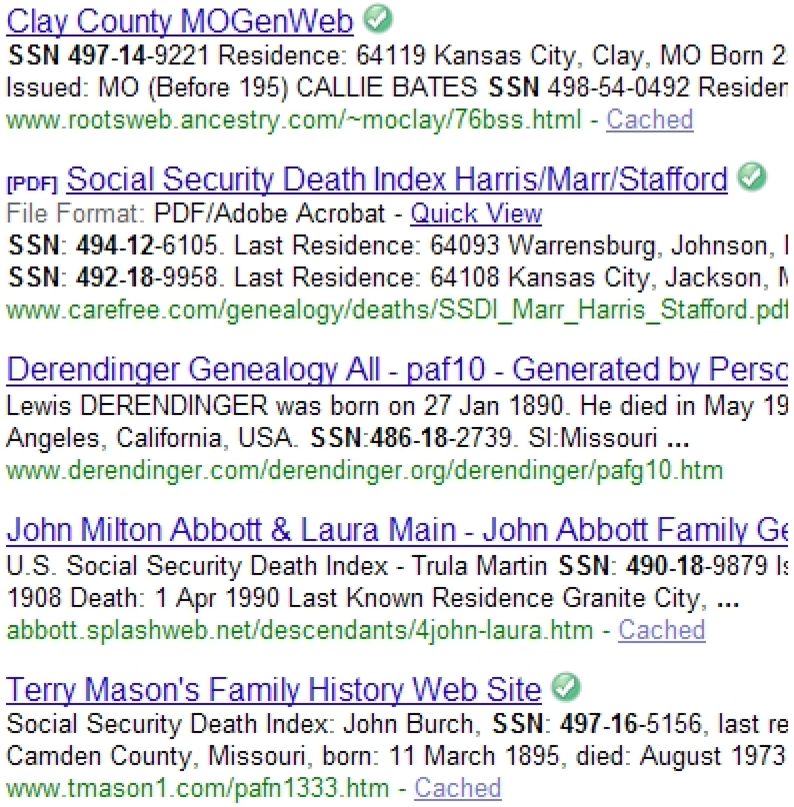
site:[your_domain] ssn "486..500-01..99"

A simple query might work too:

site:[your_domain] ssn OR “social security number”

Social Security Number Allocations

http://www.socialsecurity.gov/employer/stateweb.htm



Clay County MOGenWeb ✓
SSN 497-14-9221 Residence: 64119 Kansas City, Clay, MO Born 2
Issued: MO (Before 195) CALLIE BATES **SSN** 498-54-0492 Resider
www.rootsweb.ancestry.com/~moclay/76bss.html - Cached

[PDF] Social Security Death Index Harris/Marr/Stafford ✓
File Format: PDF/Adobe Acrobat - Quick View
SSN: **494-12**-6105. Last Residence: 64093 Warrensburg, Johnson,
SSN: **492-18**-9958. Last Residence: 64108 Kansas City, Jackson,
www.carefree.com/genealogy/deaths/SSDI_Marr_Harris_Stafford.pd

Derendinger Genealogy All - paf10 - Generated by Perso
Lewis DERENDINGER was born on 27 Jan 1890. He died in May 19
Angeles, California, USA. **SSN**:**486-18**-2739. SI:Missouri **...**
www.derendinger.com/derendinger.org/derendinger/pafg10.htm

John Milton Abbott & Laura Main - John Abbott Family Ge
U.S. Social Security Death Index - Trula Martin **SSN**: **490-18**-9879 Is
1908 Death: 1 Apr 1990 Last Known Residence Granite City, **...**
abbott.splashweb.net/descendants/4john-laura.htm - Cached

Terry Mason's Family History Web Site ✓
Social Security Death Index: John Burch, **SSN**: **497-16**-5156, last re
Camden County, Missouri, born: 11 March 1895, died: August 1973
www.tmason1.com/pafn1333.htm - Cached

Example above:
ssn "486..500-12..23”

MOREnet
Community • Technology • Results

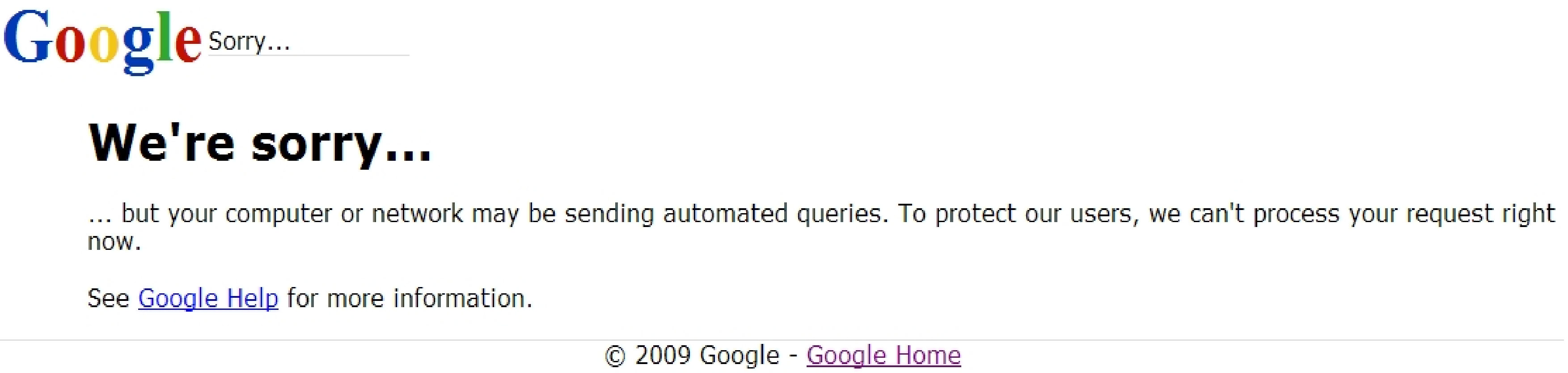
Credit Card Information

When the numrange operator was first introduced, you could use it to search for credit card numbers. This no longer works

Example: mastercard 0000000000000000..9999999999999999

Credit Card Information

When you attempt to use numrange to find credit card numbers, you are likely to see this:



We're sorry...

... but your computer or network may be sending automated queries. To protect our users, we can't process your request right now.

See Google Help for more information.

© 2009 Google - Google Home

Credit Card Information

- Think about how your credit card data is stored. Are there field names or other keywords used by your applications or staff that can be queried?

- Is the data stored in a specific file type that can be queried?

We have found credit cards numbers on a member site using the most simple query!

Simple Query
site:[your_domain] “credit card”

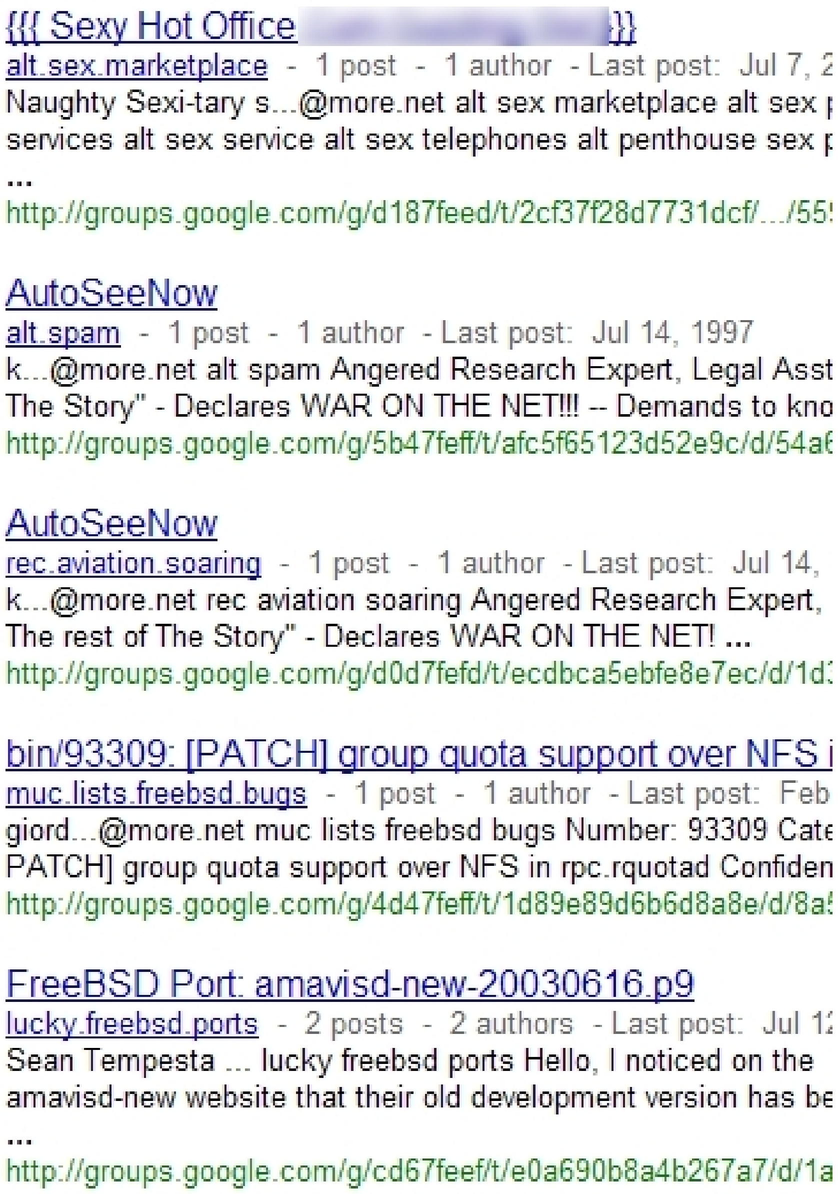
How could you customize this query to better suit your data?

MOREnet

Community • Technology • Results

Google Groups

- What information is being disclosed about your organization in Google Groups?
- Are staff members creating inappropriate posts using their work addresses?
- Are spammers using your domain name?
- Is confidential information being leaked?



{{{ Sexy Hot Office [redacted] }}}
alt.sex.marketplace - 1 post - 1 author - Last post: Jul 7, 2
Naughty Sexi-tary s...@more.net alt sex marketplace alt sex p
services alt sex service alt sex telephones alt penthouse sex p
...
http://groups.google.com/g/d187feed/t/2cf37f28d7731dcf/.../55

AutoSeeNow
alt.spam - 1 post - 1 author - Last post: Jul 14, 1997
k...@more.net alt spam Angered Research Expert, Legal Asst
The Story" - Declares WAR ON THE NET!!! -- Demands to kno
http://groups.google.com/g/5b47feff/t/afc5f65123d52e9c/d/54a6

AutoSeeNow
rec.aviation.soaring - 1 post - 1 author - Last post: Jul 14,
k...@more.net rec aviation soaring Angered Research Expert,
The rest of The Story" - Declares WAR ON THE NET! ...
http://groups.google.com/g/d0d7fefd/t/ecdbca5ebfe8e7ec/d/1d3

bin/93309: [PATCH] group quota support over NFS i
muc.lists.freebsd.bugs - 1 post - 1 author - Last post: Feb
giord...@more.net muc lists freebsd bugs Number: 93309 Cate
PATCH] group quota support over NFS in rpc.rquotad Confiden
http://groups.google.com/g/4d47feff/t/1d89e89d6b6d8a8e/d/8a5

FreeBSD Port: amavisd-new-20030616.p9
lucky.freebsd.ports - 2 posts - 2 authors - Last post: Jul 12
Sean Tempesta ... lucky freebsd ports Hello, I noticed on the
amavisd-new website that their old development version has be
...
http://groups.google.com/g/cd67feef/t/e0a690b8a4b267a7/d/1a

MOREnet
Community • Technology • Results

Google Groups

Query by email domain

author:@[your_domain]

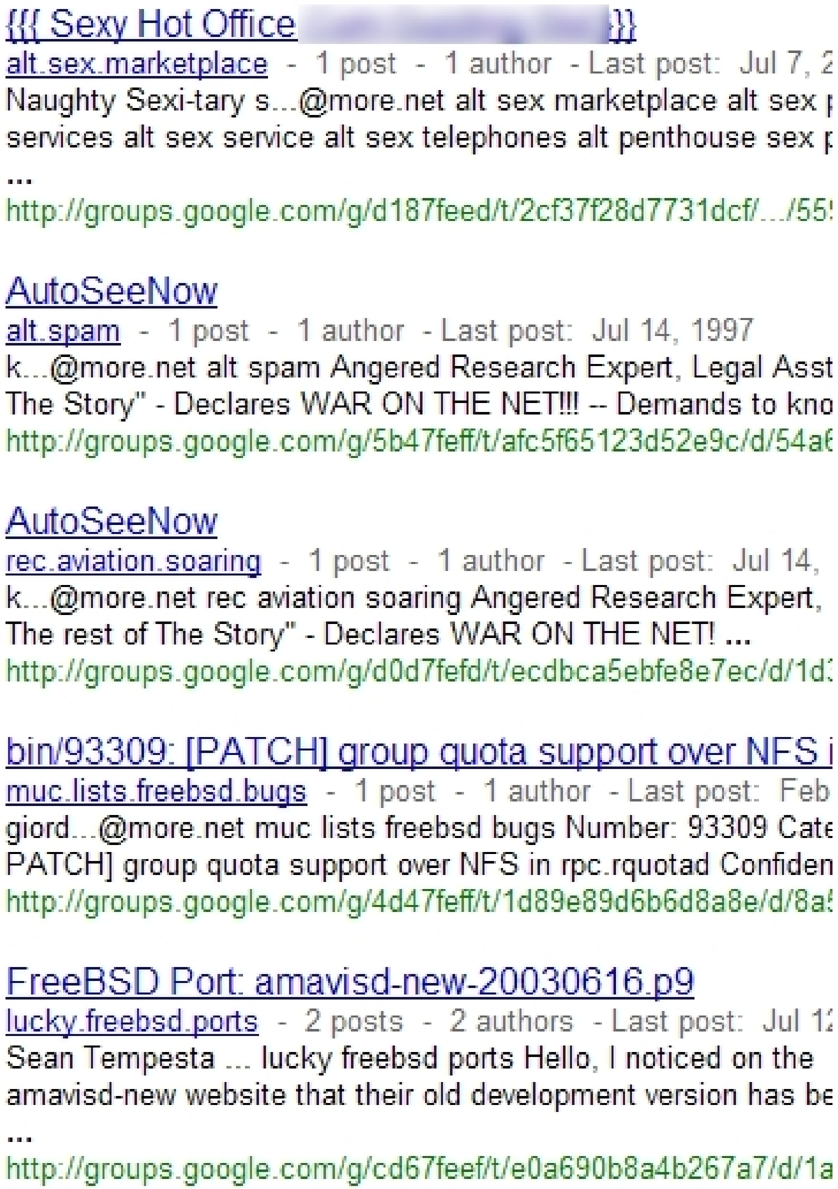
Query by author name

author:[first] author:[last]

Examples

author:@more.net

author:Beth author:Young



{{{ Sexy Hot Office [redacted] }}}
alt.sex.marketplace - 1 post - 1 author - Last post: Jul 7, 2
Naughty Sexi-tary s...@more.net alt sex marketplace alt sex p
services alt sex service alt sex telephones alt penthouse sex p
...
http://groups.google.com/g/d187feed/t/2cf37f28d7731dcf/.../55

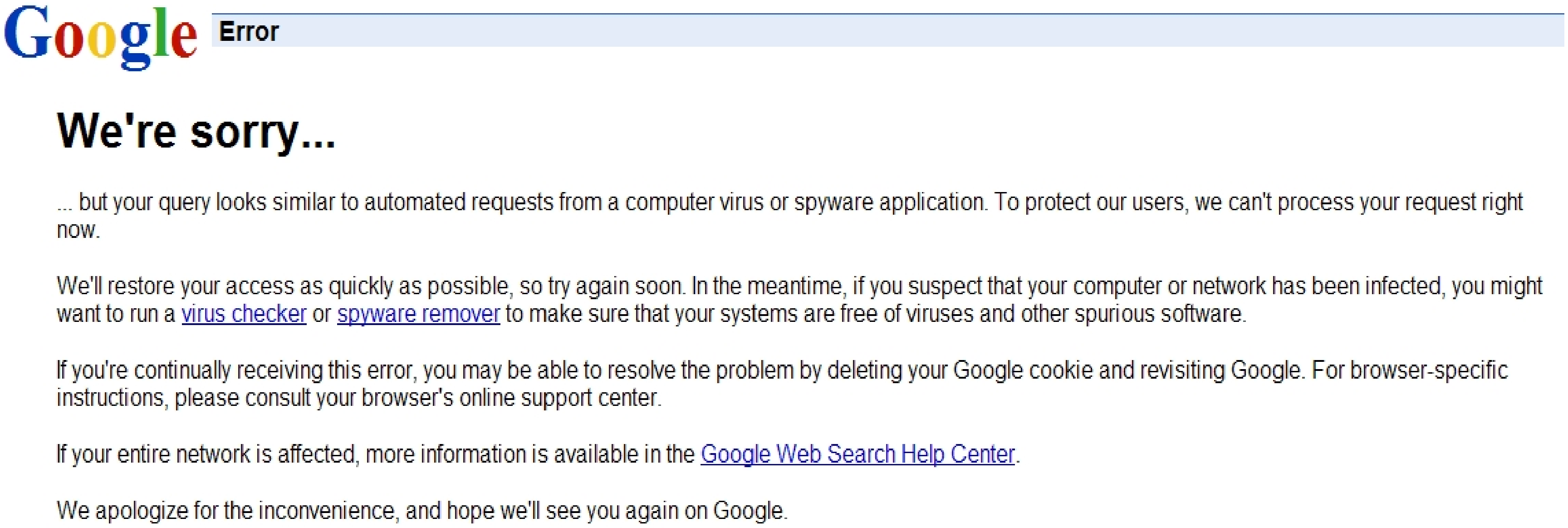
AutoSeeNow
alt.spam - 1 post - 1 author - Last post: Jul 14, 1997
k...@more.net alt spam Angered Research Expert, Legal Asst
The Story" - Declares WAR ON THE NET!!! -- Demands to kno
http://groups.google.com/g/5b47feff/t/afc5f65123d52e9c/d/54a6

AutoSeeNow
rec.aviation.soaring - 1 post - 1 author - Last post: Jul 14,
k...@more.net rec aviation soaring Angered Research Expert,
The rest of The Story" - Declares WAR ON THE NET! ...
http://groups.google.com/g/d0d7fefd/t/ecdbca5ebfe8e7ec/d/1d3

bin/93309: [PATCH] group quota support over NFS i
muc.lists.freebsd.bugs - 1 post - 1 author - Last post: Feb
giord...@more.net muc lists freebsd bugs Number: 93309 Cate
PATCH] group quota support over NFS in rpc.rquotad Confiden
http://groups.google.com/g/4d47feff/t/1d89e89d6b6d8a8e/d/8a5

FreeBSD Port: amavisd-new-20030616.p9
lucky.freebsd.ports - 2 posts - 2 authors - Last post: Jul 12
Sean Tempesta ... lucky freebsd ports Hello, I noticed on the
amavisd-new website that their old development version has be
...
http://groups.google.com/g/cd67feef/t/e0a690b8a4b267a7/d/1a

Can Google Hacking searches be automated?



Google Error

We're sorry...

... but your query looks similar to automated requests from a computer virus or spyware application. To protect our users, we can't process your request right now.

We'll restore your access as quickly as possible, so try again soon. In the meantime, if you suspect that your computer or network has been infected, you might want to run a virus checker or spyware remover to make sure that your systems are free of viruses and other spurious software.

If you're continually receiving this error, you may be able to resolve the problem by deleting your Google cookie and revisiting Google. For browser-specific instructions, please consult your browser's online support center.

If your entire network is affected, more information is available in the Google Web Search Help Center.

We apologize for the inconvenience, and hope we'll see you again on Google.

3rd Party Tools

Use at your own risk!

- SALSA Google hacking tool
  https://spaces.internet2.edu/display/SalsaCSI2WG/GoogleHackingDescription
- SiteDigger
  http://www.foundstone.com/us/resources/proddesc/sitedigger.htm
- Wikto
  http://www.sensepost.com/research/wikto/
- BiDiBLAH
  http://www.sensepost.com/research/bidiblah/
- Goolag (Cult of the Dead Cow)
  http://www.goolag.org/
- Athena
  http://snakeoillabs.com/wordpress/2004/11/07/athena-20-is-go/
- Gooscan
  Authored by Johnny Long of the Google Hacking Database. There doesn't appear to be an official download location anymore, but it may still be found on various tool archives.

Google Alerts

- Google provides their own tool that can be used to automate queries:

  http://www.google.com/alerts?hl=en

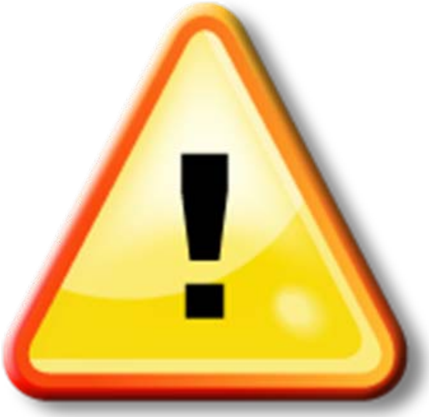
- You can have up to 1000 active alerts

- Maximum number of terms per search is 32 (Google default)

- The results of your alerts are emailed to you at an interval you specify

- It may not alert you to a particular result more than once

The Alert Symbol

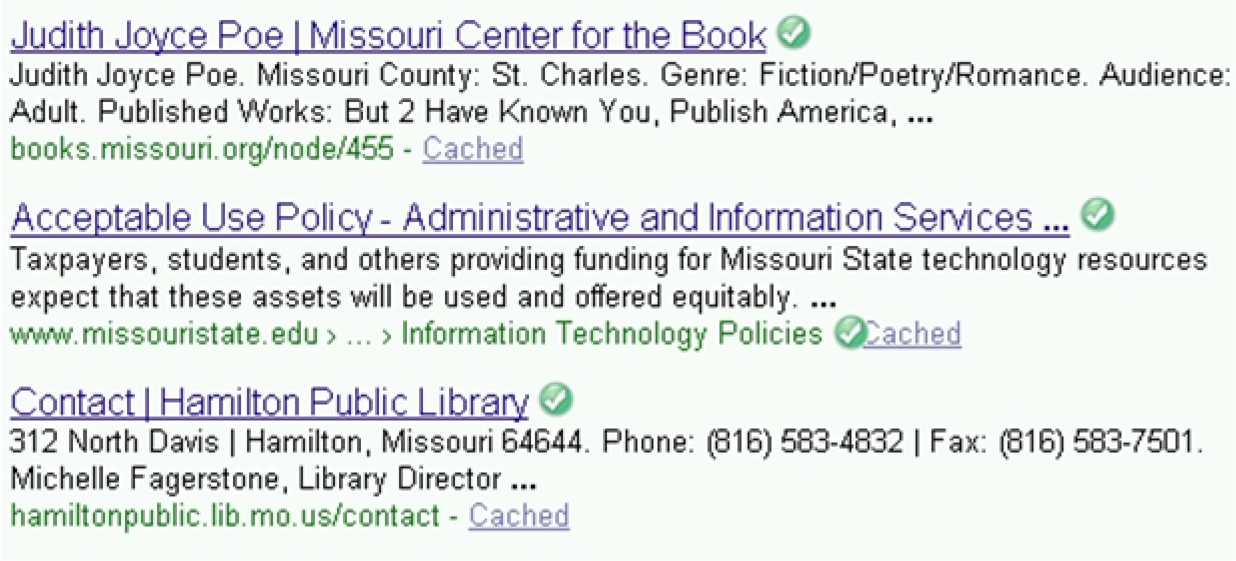


This symbol denotes the queries that frequently return results.
These queries are excellent candidates for a Google Alert!

http://www.google.com/alerts

Who is linking to your site?

link:[your_domain]



[Judith Joyce Poe | Missouri Center for the Book](books.missouri.org/node/455) ✓
Judith Joyce Poe. Missouri County: St. Charles. Genre: Fiction/Poetry/Romance. Audience: Adult. Published Works: But 2 Have Known You, Publish America, ...
books.missouri.org/node/455 - Cached

Acceptable Use Policy - Administrative and Information Services ... ✓
Taxpayers, students, and others providing funding for Missouri State technology resources expect that these assets will be used and offered equitably. ...
www.missouristate.edu › ... › Information Technology Policies ✓ Cached

Contact | Hamilton Public Library ✓
312 North Davis | Hamilton, Missouri 64644. Phone: (816) 583-4832 | Fax: (816) 583-7501. Michelle Fagerstone, Library Director ...
hamiltonpublic.lib.mo.us/contact - Cached

Examples:

- link:http://www.more.net *or* link:www.more.net
- link:http://www.more.net/content/security-2

Using other operators in conjunction will provide unpredictable results.

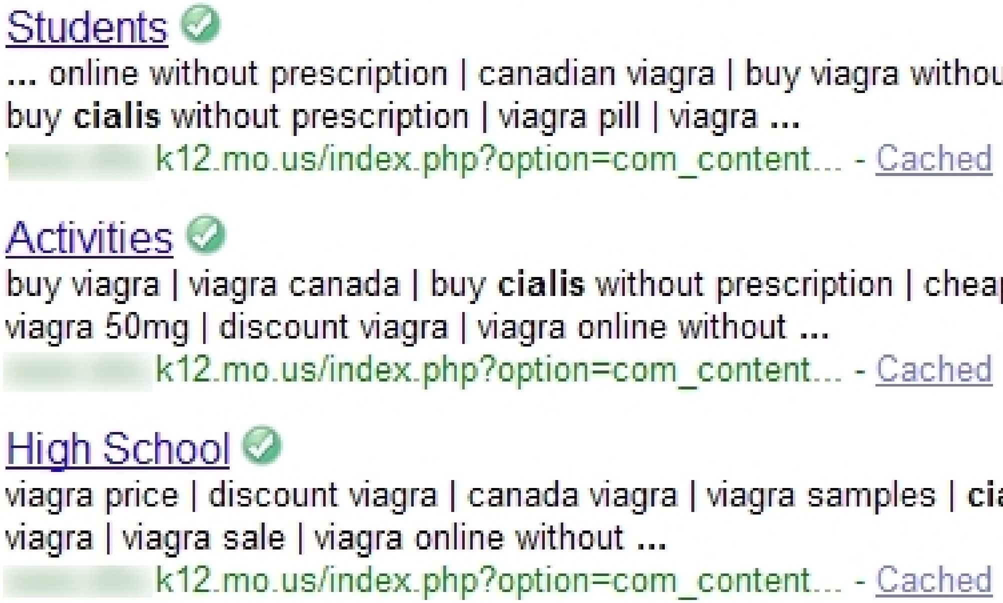
Website Spam

Website Spam

- If you find that your site has been spammed, you have a vulnerability!
  - ➢ Content management software and modules
  - ➢ Forum software
  - ➢ Guestbook software
  - ➢ Code vulnerable to injection
- Gives spammers another outlet to get their message across.
- It may also be used to increase page rank in searches.
- It may not always be visible to you when looking at the site directly; Buried in the source code.

Customize these URLs as you see new spam trends and terms.

An example of SEO spam



Students

... online without prescription | canadian viagra | buy viagra without prescription | canada viagra | buy **cialis** without prescription | viagra pill | viagra ...

[redacted]k12.mo.us/index.php?option=com_content... - Cached

Activities

buy viagra | viagra canada | buy **cialis** without prescription | cheapest viagra | viagra professional | viagra 50mg | discount viagra | viagra online without ...

[redacted]k12.mo.us/index.php?option=com_content... - Cached

High School

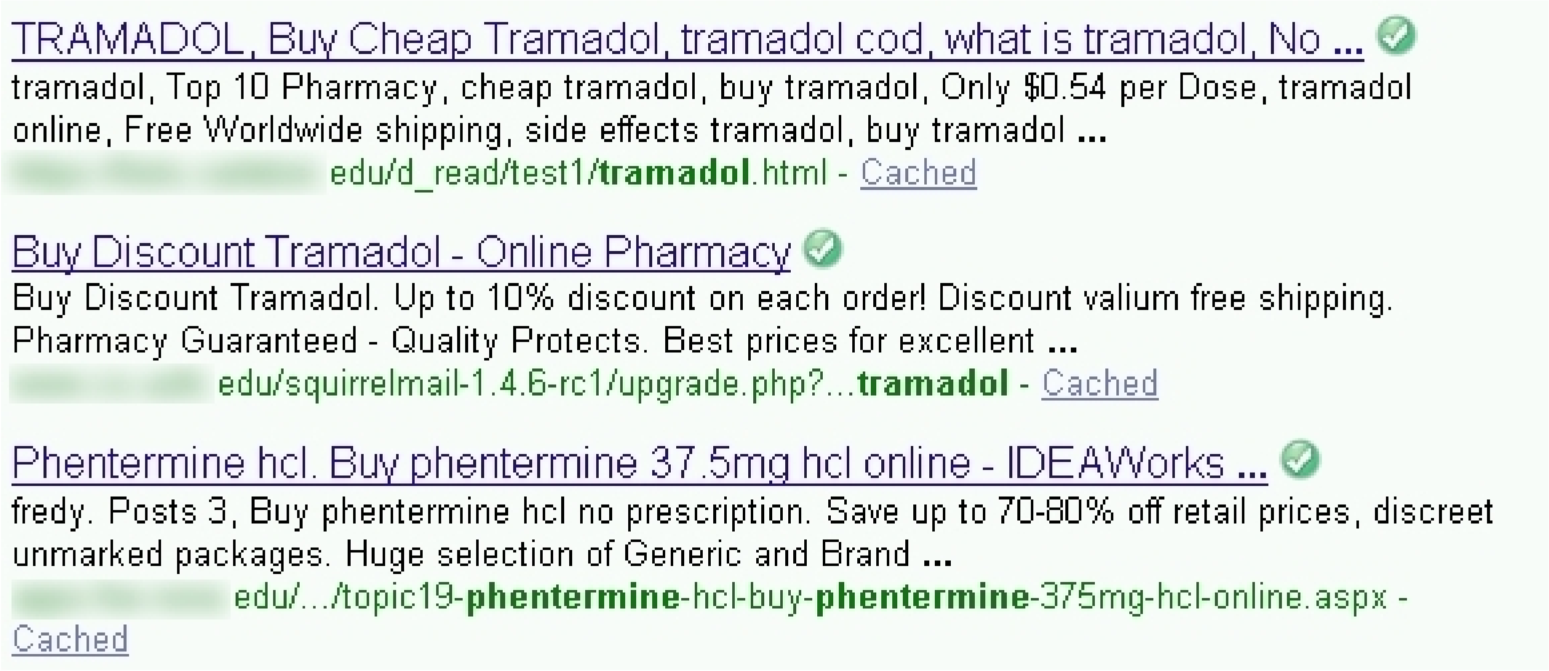
viagra price | discount viagra | canada viagra | viagra samples | **cialis** without prescription | buy viagra | viagra sale | viagra online without ...

[redacted]k12.mo.us/index.php?option=com_content... - Cached

site:[your_domain] phentermine OR viagra OR cialis OR vioxx OR oxycontin OR levitra OR ambien OR xanax OR paxil OR "slot-machine" OR "texas holdem" OR ringtones

site:[your_domain] porn OR hardcore OR exotic OR erotica OR lingerie OR underage OR nude OR "preteen girls"

Spam: Injected Pages



TRAMADOL, Buy Cheap Tramadol, tramadol cod, what is tramadol, No **...** ✓
tramadol, Top 10 Pharmacy, cheap tramadol, buy tramadol, Only $0.54 per Dose, tramadol online, Free Worldwide shipping, side effects tramadol, buy tramadol **...**
[redacted] edu/d_read/test1/**tramadol**.html - Cached

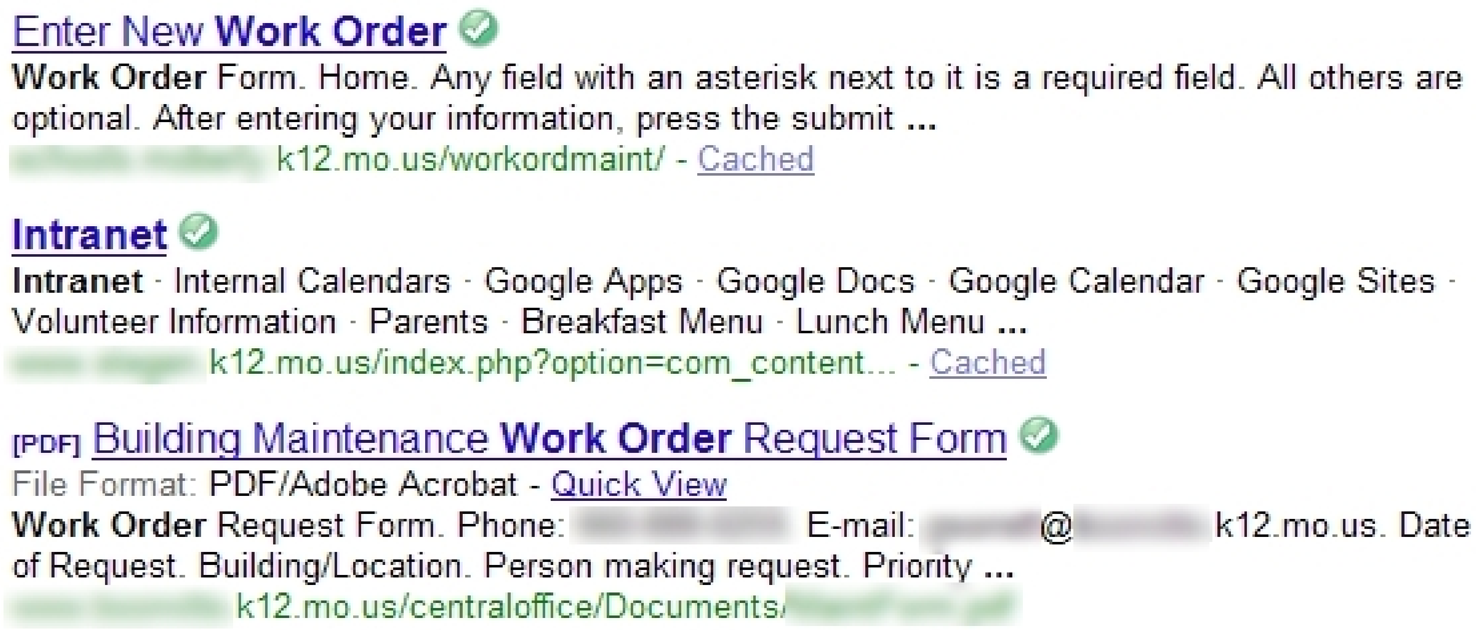
Buy Discount Tramadol - Online Pharmacy ✓
Buy Discount Tramadol. Up to 10% discount on each order! Discount valium free shipping. Pharmacy Guaranteed - Quality Protects. Best prices for excellent **...**
[redacted] edu/squirrelmail-1.4.6-rc1/upgrade.php?...**tramadol** - Cached

Phentermine hcl. Buy phentermine 37.5mg hcl online - IDEAWorks **...** ✓
fredy. Posts 3, Buy phentermine hcl no prescription. Save up to 70-80% off retail prices, discreet unmarked packages. Huge selection of Generic and Brand **...**
[redacted] edu/.../topic19-**phentermine**-hcl-buy-**phentermine**-375mg-hcl-online.aspx - Cached

site:[your_domain] inurl:buy.php OR inurl:phentermine OR inurl:tramadol OR inurl:meridia OR inurl:adipex OR inurl:xenical OR inurl:ionamin OR inurl:tenuate OR inurl:buy-online OR inurl:hold-em OR inurl:viagra OR inurl:online-casino OR inurl:levitra

MOREnet
Community • Technology • Results

Help Desk and Intranet Exposure



Enter New **Work Order** ✓
Work Order Form. Home. Any field with an asterisk next to it is a required field. All others are optional. After entering your information, press the submit ...
[redacted] k12.mo.us/workordmaint/ - Cached

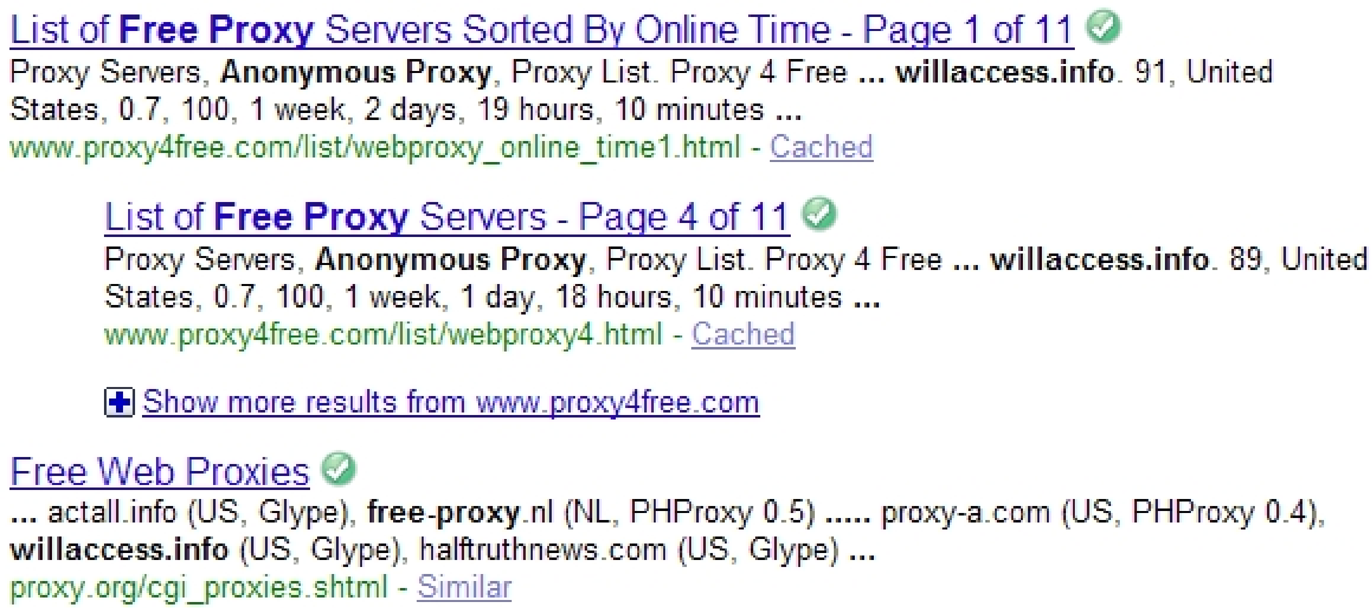
Intranet ✓
Intranet - Internal Calendars - Google Apps - Google Docs - Google Calendar - Google Sites - Volunteer Information - Parents - Breakfast Menu - Lunch Menu ...
[redacted] k12.mo.us/index.php?option=com_content... - Cached

[PDF] Building Maintenance **Work Order** Request Form ✓
File Format: PDF/Adobe Acrobat - Quick View
Work Order Request Form. Phone: [redacted] E-mail: [redacted]@[redacted] k12.mo.us. Date of Request. Building/Location. Person making request. Priority ...
[redacted] k12.mo.us/centraloffice/Documents/[redacted]

site:[your_domain] "work order" OR helpdesk OR "help desk" OR intranet OR "help ticket“ OR inurl:intranet

Customize this query to contain the name of your help desk ticket system, and include inurl operators for words that are unique to your intranet/ticket system!

Open / Anonymous Proxy



List of **Free Proxy** Servers Sorted By Online Time - Page 1 of 11
Proxy Servers, **Anonymous Proxy**, Proxy List. Proxy 4 Free ... **willaccess.info**. 91, United States, 0.7, 100, 1 week, 2 days, 19 hours, 10 minutes ...
www.proxy4free.com/list/webproxy_online_time1.html - Cached

List of **Free Proxy** Servers - Page 4 of 11
Proxy Servers, **Anonymous Proxy**, Proxy List. Proxy 4 Free ... **willaccess.info**. 89, United States, 0.7, 100, 1 week, 1 day, 18 hours, 10 minutes ...
www.proxy4free.com/list/webproxy4.html - Cached

Show more results from www.proxy4free.com

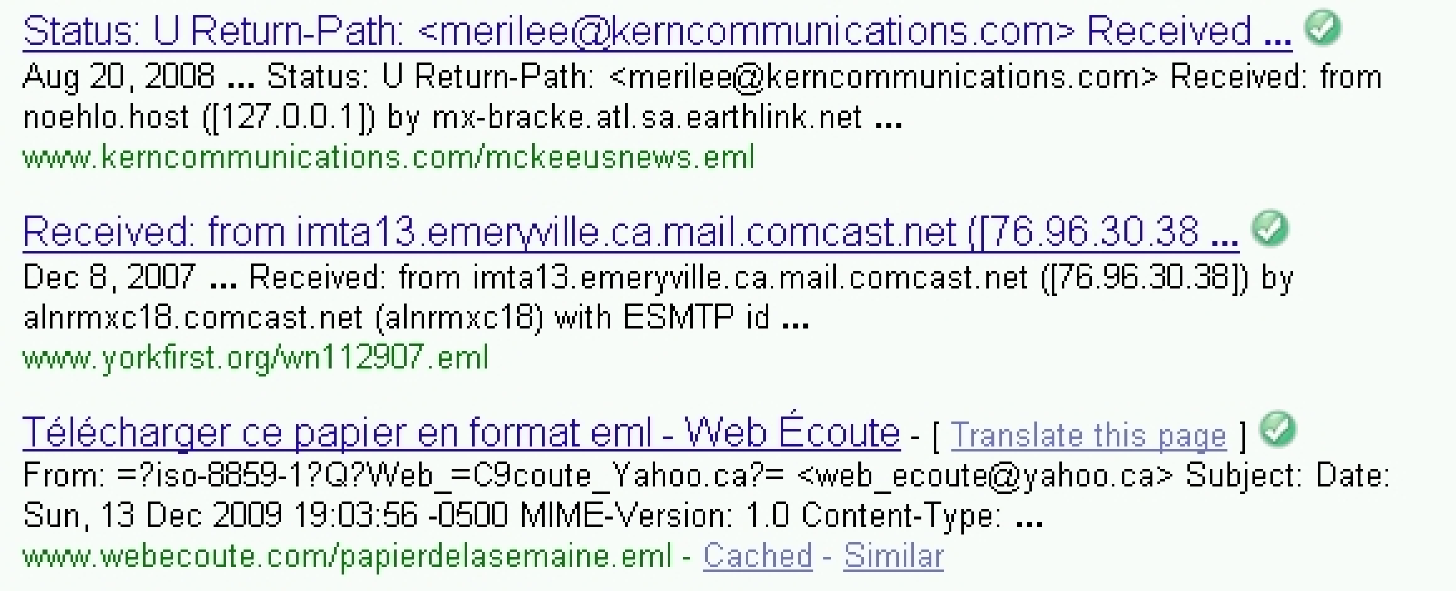
Free Web Proxies
... actall.info (US, Glype), **free-proxy**.nl (NL, PHProxy 0.5) ..... proxy-a.com (US, PHProxy 0.4), **willaccess.info** (US, Glype), halftruthnews.com (US, Glype) ...
proxy.org/cgi_proxies.shtml - Similar

+[your_domain] "anonymous proxy" OR "anonymous proxies" OR "free proxy" OR "free proxies“

+[your_IP] "anonymous proxy" OR "anonymous proxies" OR "free proxy" OR "free proxies“

In this example, we are including plural versions of the words because we are using quotes. Using quotes tells it to look for the exact phrase.

Email Exposure



Status: U Return-Path: <merilee@kerncommunications.com> Received ... ✓
Aug 20, 2008 ... Status: U Return-Path: <merilee@kerncommunications.com> Received: from noehlo.host ([127.0.0.1]) by mx-bracke.atl.sa.earthlink.net ...
www.kerncommunications.com/mckeeusnews.eml

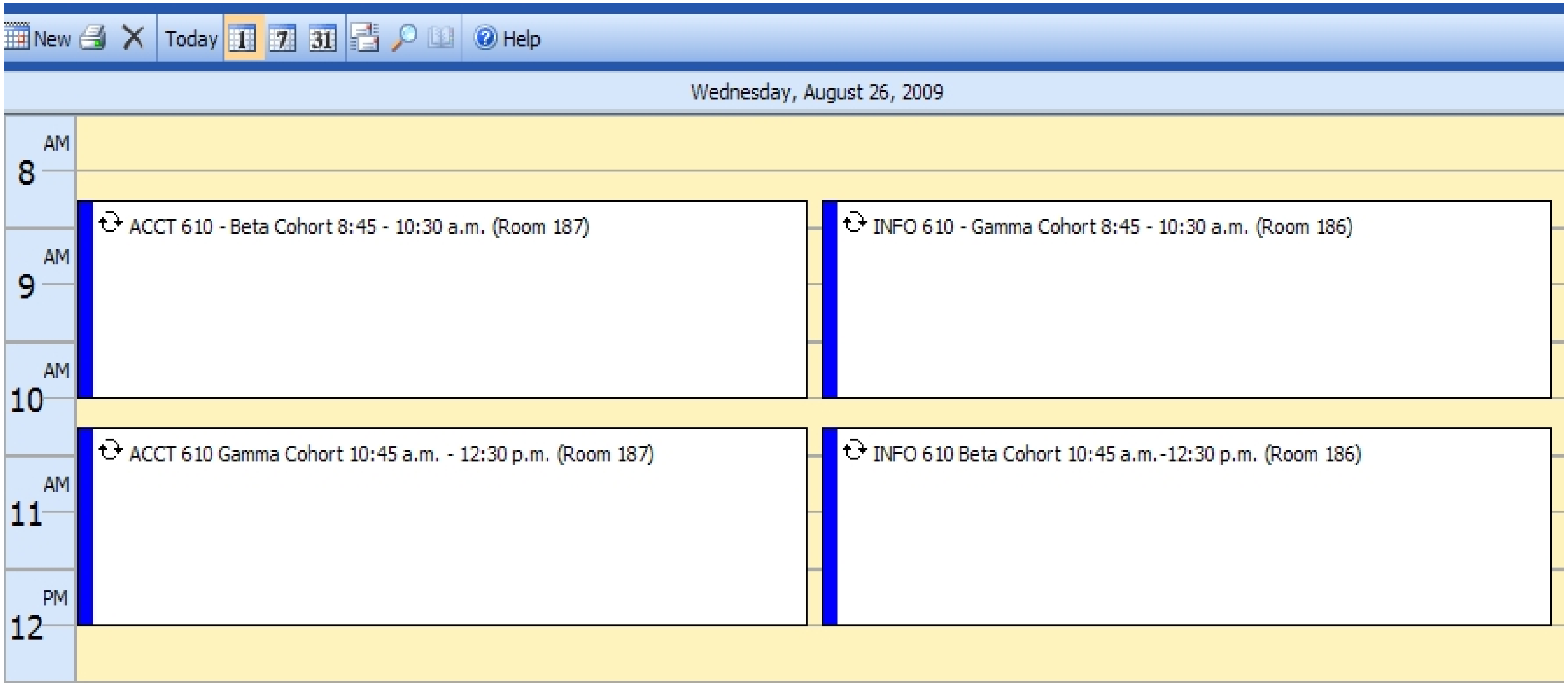
Received: from imta13.emeryville.ca.mail.comcast.net ([76.96.30.38 ... ✓
Dec 8, 2007 ... Received: from imta13.emeryville.ca.mail.comcast.net ([76.96.30.38]) by alnrmxc18.comcast.net (alnrmxc18) with ESMTP id ...
www.yorkfirst.org/wn112907.eml

Télécharger ce papier en format eml - Web Écoute - [Translate this page] ✓
From: =?iso-8859-1?Q?Web_=C9coute_Yahoo.ca?= <web_ecoute@yahoo.ca> Subject: Date: Sun, 13 Dec 2009 19:03:56 -0500 MIME-Version: 1.0 Content-Type: ...
www.webecoute.com/papierdelasemaine.eml - Cached - Similar

site:[your_domain] filetype:mbx OR filetype:eml OR filetype:pst OR filetype:dbx OR filetype:wab

Customize this query to contain the extensions that are used by your email applications!

Outlook Web Access Public Folders



Wednesday, August 26, 2009

ACCT 610 - Beta Cohort 8:45 - 10:30 a.m. (Room 187)

INFO 610 - Gamma Cohort 8:45 - 10:30 a.m. (Room 186)

ACCT 610 Gamma Cohort 10:45 a.m. - 12:30 p.m. (Room 187)

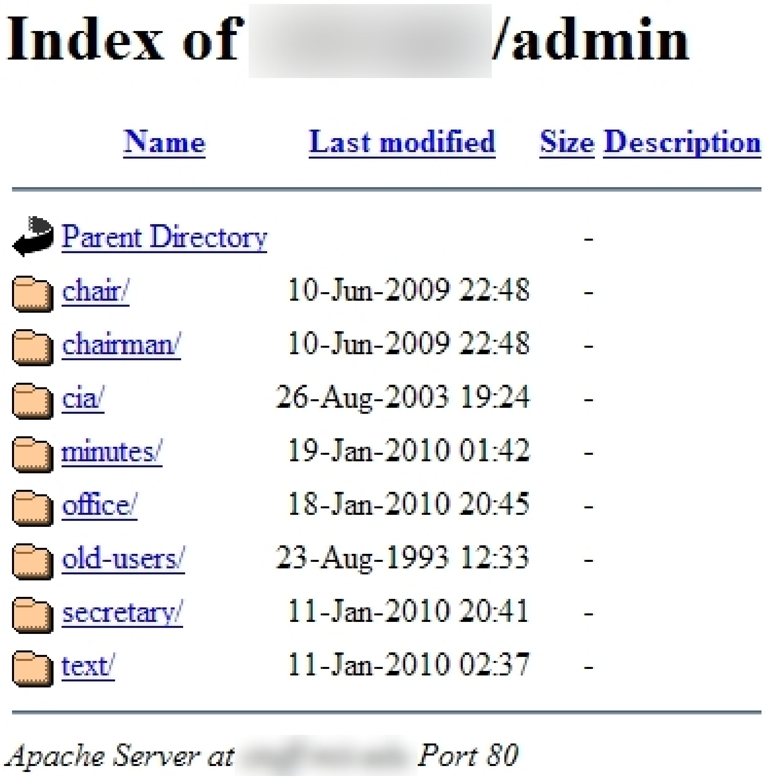
INFO 610 Beta Cohort 10:45 a.m.-12:30 p.m. (Room 186)

site:[your_domain] inurl:/public/?Cmd=contents

Directory Indexing

If there is no index.html file (in the various subdirectories of the web server) and directory indexing is enabled, visitors will see a listing of all of the files and web pages within that directory.

This can be helpful for software repositories and users who need to locate unknown files, but it is often used by attackers to locate sensitive data (e.g., passwords, spreadsheets, databases, etc.).

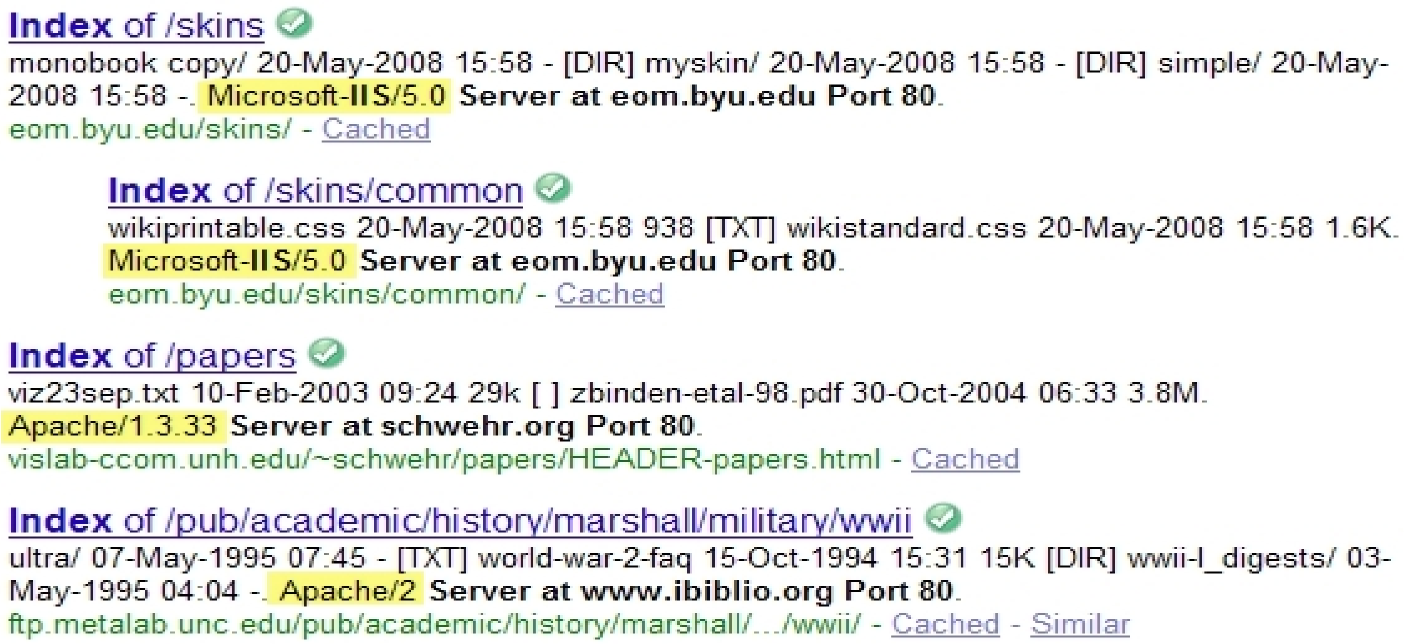


Index of [redacted]/admin

```
        Name              Last modified      Size Description
------------------------------------------------------------
   Parent Directory                            -
   chair/              10-Jun-2009 22:48       -
   chairman/           10-Jun-2009 22:48       -
   cia/                26-Aug-2003 19:24       -
   minutes/            19-Jan-2010 01:42       -
   office/             18-Jan-2010 20:45       -
   old-users/          23-Aug-1993 12:33       -
   secretary/          11-Jan-2010 20:41       -
   text/               11-Jan-2010 02:37       -
------------------------------------------------------------
Apache Server at [redacted] Port 80
```

site:[your_domain] intitle:index.of "parent directory"

Directory Indexing



Index of /skins
monobook copy/ 20-May-2008 15:58 - [DIR] myskin/ 20-May-2008 15:58 - [DIR] simple/ 20-May-2008 15:58 -. Microsoft-IIS/5.0 **Server at eom.byu.edu Port 80**.
eom.byu.edu/skins/ - Cached

Index of /skins/common
wikiprintable.css 20-May-2008 15:58 938 [TXT] wikistandard.css 20-May-2008 15:58 1.6K. Microsoft-IIS/5.0 **Server at eom.byu.edu Port 80**.
eom.byu.edu/skins/common/ - Cached

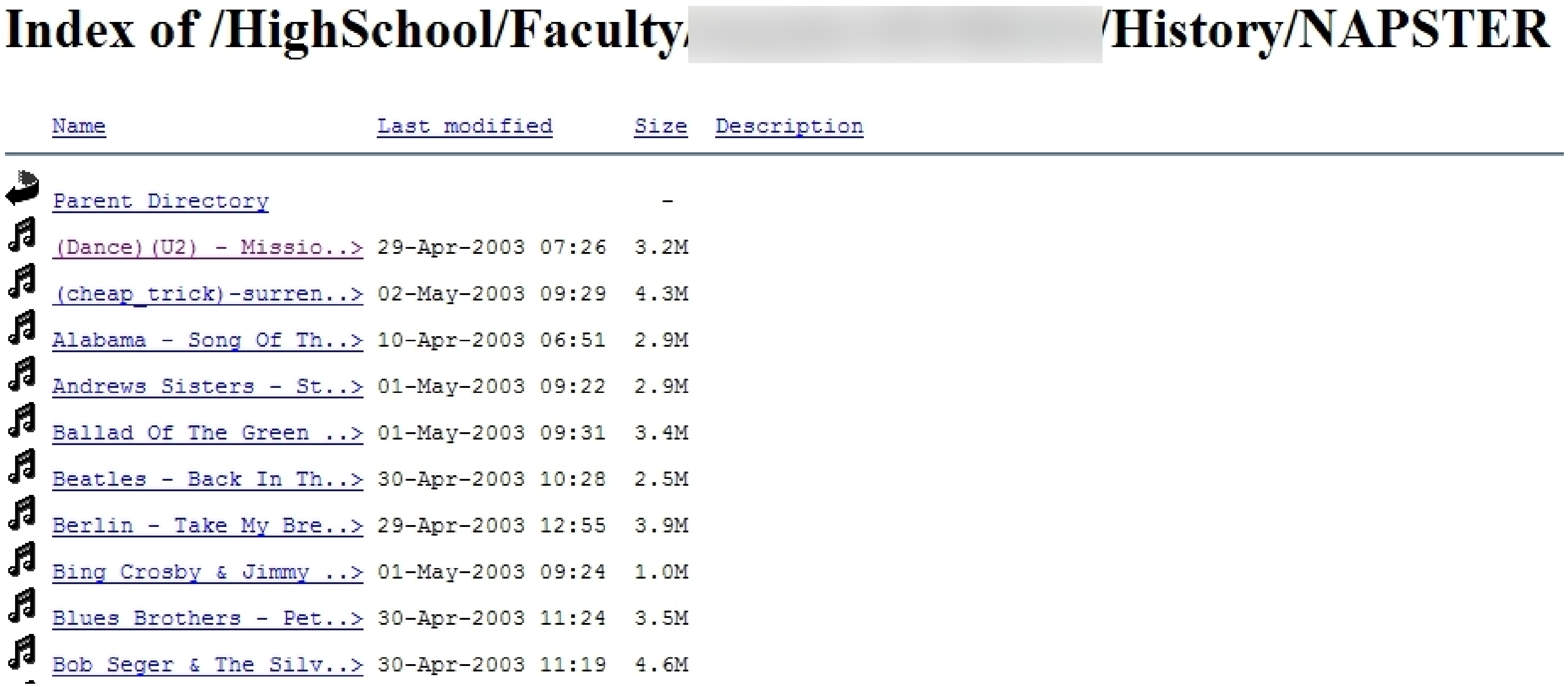
Index of /papers
viz23sep.txt 10-Feb-2003 09:24 29k [ ] zbinden-etal-98.pdf 30-Oct-2004 06:33 3.8M. Apache/1.3.33 **Server at schwehr.org Port 80**.
vislab-ccom.unh.edu/~schwehr/papers/HEADER-papers.html - Cached

Index of /pub/academic/history/marshall/military/wwii
ultra/ 07-May-1995 07:45 - [TXT] world-war-2-faq 15-Oct-1994 15:31 15K [DIR] wwii-l_digests/ 03-May-1995 04:04 -. Apache/2 **Server at www.ibiblio.org Port 80**.
ftp.metalab.unc.edu/pub/academic/history/marshall/.../wwii/ - Cached - Similar

The query can also be crafted a little differently to have Google display the server type and version in the results.

site:[your_domain] intitle:index.of "server at"

Directory Indexing: What’s wrong with this picture?

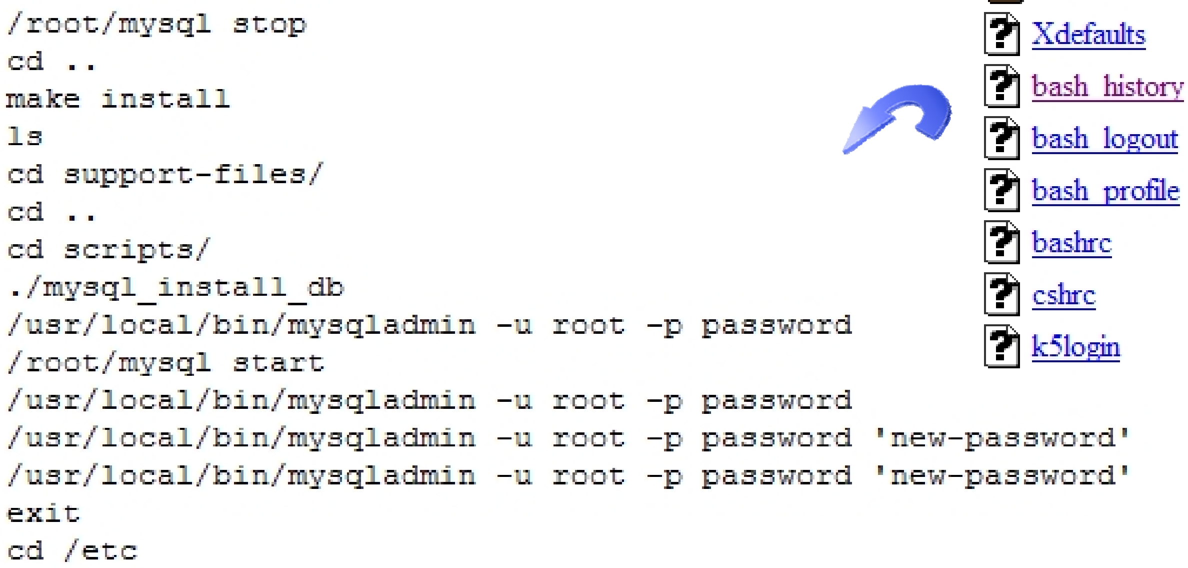


Index of /HighSchool/Faculty/[redacted]/History/NAPSTER

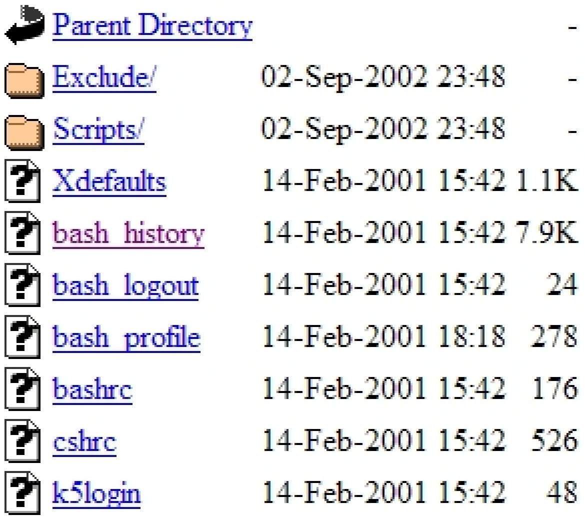
```
   Name                     Last modified      Size  Description
--------------------------------------------------------------------
   Parent Directory                             -
   (Dance)(U2) - Missio..> 29-Apr-2003 07:26  3.2M
   (cheap_trick)-surren..> 02-May-2003 09:29  4.3M
   Alabama - Song Of Th..> 10-Apr-2003 06:51  2.9M
   Andrews Sisters - St..> 01-May-2003 09:22  2.9M
   Ballad Of The Green ..> 01-May-2003 09:31  3.4M
   Beatles - Back In Th..> 30-Apr-2003 10:28  2.5M
   Berlin - Take My Bre..> 29-Apr-2003 12:55  3.9M
   Bing Crosby & Jimmy ..> 01-May-2003 09:24  1.0M
   Blues Brothers - Pet..> 30-Apr-2003 11:24  3.5M
   Bob Seger & The Silv..> 30-Apr-2003 11:19  4.6M
```

This faculty member’s directory also contained classroom tests for the school year, and letters written about students!!

Directory Indexing: Bash History Files



```
/root/mysql stop
cd ..
make install
ls
cd support-files/
cd ..
cd scripts/
./mysql_install_db
/usr/local/bin/mysqladmin -u root -p password
/root/mysql start
/usr/local/bin/mysqladmin -u root -p password
/usr/local/bin/mysqladmin -u root -p password 'new-password'
/usr/local/bin/mysqladmin -u root -p password 'new-password'
exit
cd /etc
```

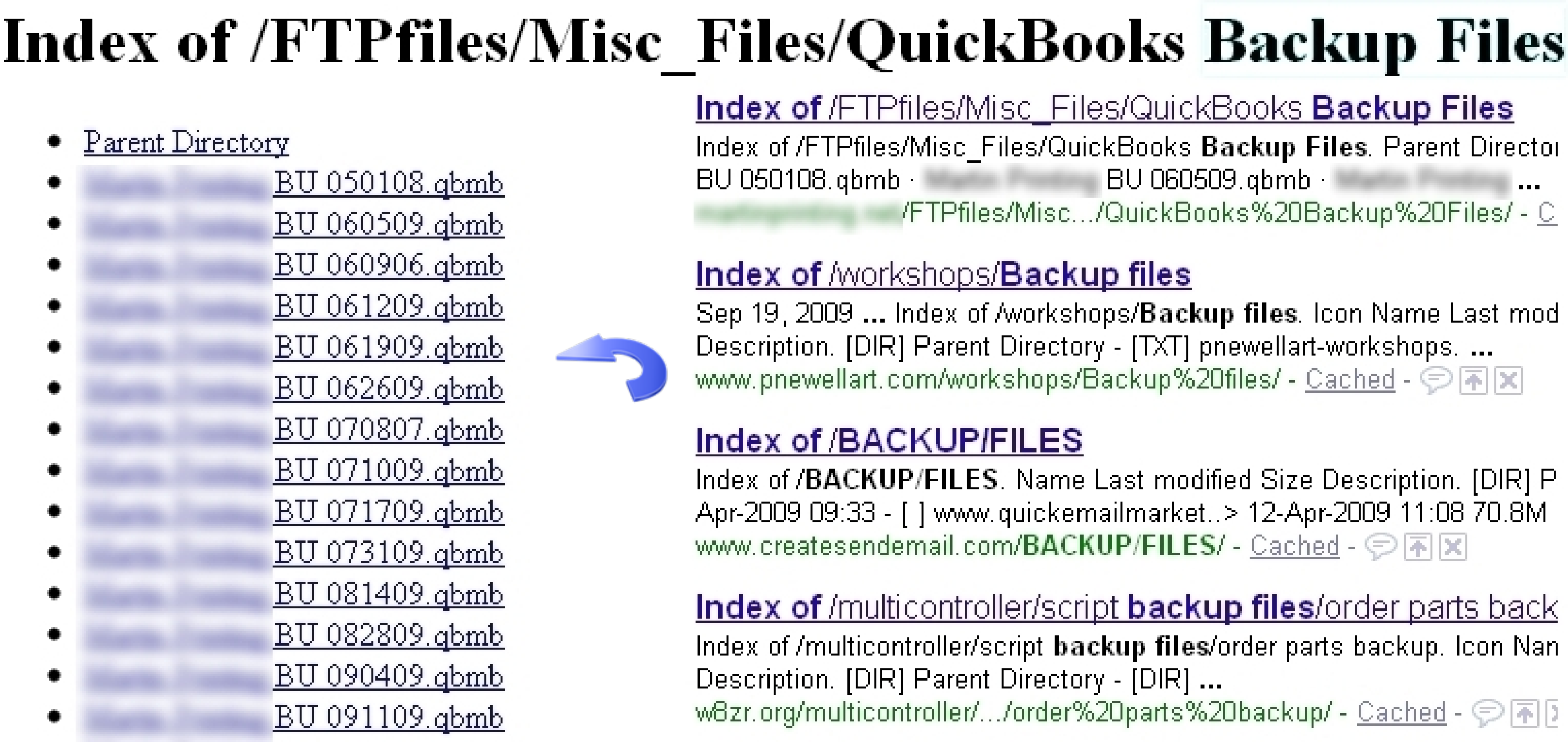


Parent Directory		-
Exclude/	02-Sep-2002 23:48	-
Scripts/	02-Sep-2002 23:48	-
Xdefaults	14-Feb-2001 15:42	1.1K
bash_history	14-Feb-2001 15:42	7.9K
bash_logout	14-Feb-2001 15:42	24
bash_profile	14-Feb-2001 18:18	278
bashrc	14-Feb-2001 15:42	176
cshrc	14-Feb-2001 15:42	526
k5login	14-Feb-2001 15:42	48

site:[your_domain] intitle:index.of .bash_history

MOREnet
Community • Technology • Results

Directory Indexing: Sensitive Information



Index of /FTPfiles/Misc_Files/QuickBooks Backup Files

- Parent Directory
- [redacted] BU 050108.qbmb
- [redacted] BU 060509.qbmb
- [redacted] BU 060906.qbmb
- [redacted] BU 061209.qbmb
- [redacted] BU 061909.qbmb
- [redacted] BU 062609.qbmb
- [redacted] BU 070807.qbmb
- [redacted] BU 071009.qbmb
- [redacted] BU 071709.qbmb
- [redacted] BU 073109.qbmb
- [redacted] BU 081409.qbmb
- [redacted] BU 082809.qbmb
- [redacted] BU 090409.qbmb
- [redacted] BU 091109.qbmb

Index of /FTPfiles/Misc_Files/QuickBooks Backup Files
Index of /FTPfiles/Misc_Files/QuickBooks **Backup Files**. Parent Directo
BU 050108.qbmb · [redacted] BU 060509.qbmb · [redacted] ...
[redacted]/FTPfiles/Misc.../QuickBooks%20Backup%20Files/ - C

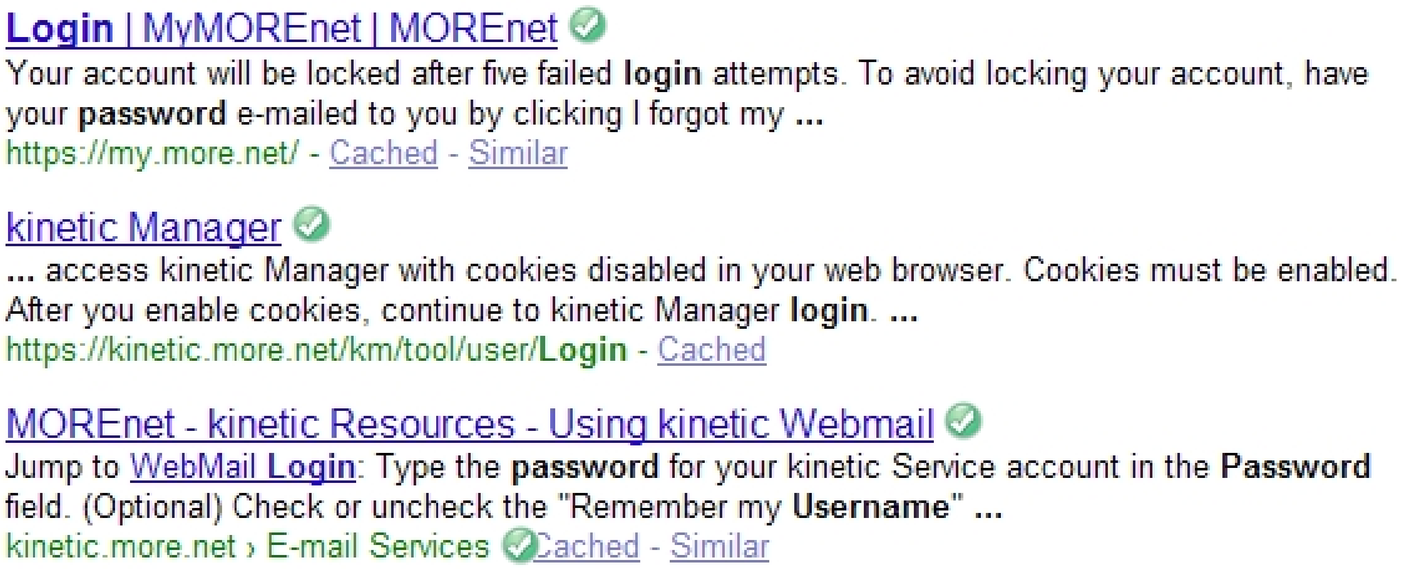
Index of /workshops/Backup files
Sep 19, 2009 ... Index of /workshops/**Backup files**. Icon Name Last mod
Description. [DIR] Parent Directory - [TXT] pnewellart-workshops. ...
www.pnewellart.com/workshops/Backup%20files/ - Cached -

Index of /BACKUP/FILES
Index of /**BACKUP**/**FILES**. Name Last modified Size Description. [DIR] P
Apr-2009 09:33 - [ ] www.quickemailmarket..> 12-Apr-2009 11:08 70.8M
www.createsendemail.com/**BACKUP**/**FILES**/ - Cached -

Index of /multicontroller/script backup files/order parts back
Index of /multicontroller/script **backup files**/order parts backup. Icon Nan
Description. [DIR] Parent Directory - [DIR] ...
w8zr.org/multicontroller/.../order%20parts%20backup/ - Cached -

site:[your_domain] intitle:"index of" admin OR private OR "backup files“ OR backup

Login Portals



[Login | MyMOREnet | MOREnet](https://my.more.net/)
Your account will be locked after five failed **login** attempts. To avoid locking your account, have your **password** e-mailed to you by clicking I forgot my **...**
https://my.more.net/ - Cached - Similar

kinetic Manager
... access kinetic Manager with cookies disabled in your web browser. Cookies must be enabled. After you enable cookies, continue to kinetic Manager **login**. **...**
https://kinetic.more.net/km/tool/user/**Login** - Cached

MOREnet - kinetic Resources - Using kinetic Webmail
Jump to WebMail **Login**: Type the **password** for your kinetic Service account in the **Password** field. (Optional) Check or uncheck the "Remember my **Username**" **...**
kinetic.more.net › E-mail Services - Cached - Similar

Should the world have access to them?

site:[your_domain] inurl:admin | inurl:login | intitle:admin | intitle:login | intext:“log on” | intext:login | intext:username | intext:password | intext:portal

What else can you find?

- Devices (printers, switches, routers, etc.)
- Usernames and passwords
- Database dumps/information exposures/misconfigurations
- Service and device configuration files
- Log files
- Much more than you should!

The queries we have provided are a great start. Take some time to look around and see what Google knows about you!

MOREnet
Community • Technology • Results

How Do I Keep Google From Indexing Content?

Robots.txt

- A robots.txt file specifies the portions of your web site that you don’t want indexed by search engines.
- Respectable bots will abide by the directives in your robots.txt file but troublemakers may ignore it.
- Attackers will actively search it out to learn where you don’t want them to go.

Robots.txt

- It is not a recommended way to prevent access to confidential data! If it is confidential, it should be password protected or removed.
- Google won't crawl or index the content of pages restricted by robots.txt, but they may still index the URLs if they find links to them on other pages on the web.
- While useful, it is not a foolproof means of keeping your information out of web searches.

Robots.txt syntax

Instead of telling the web crawlers which directories not to index (since this will tell an attacker exactly which directories contain sensitive information), disallow all directories and then allow the specific directories you would like the general public to have access to:

```
User-agent: *                          # * - Wildcard to address all bots
Allow: /directory1/file.html           # Allow a specific file
Allow: /directory2/                    # Allow all files in a specific directory
Disallow: /directory2/*.gif$           # Disallow all .gif files in the allowed directory
Allow: /directory3/dogs.jpg            # Allow a specific image
Disallow: /                            # Disallow indexing the entire site
```

Test robots.txt tool

- Part of the Google Webmasters Tool Set.
- Allows you to see if your robots.txt file is accidentally blocking Googlebot from a file or directory on your site.
- Allows you to see if it's permitting Googlebot to crawl files that should not be indexed.
- The tool checks your syntax, lists the effects of the file, and notifies you of possible problems.

Test robots.txt tool

http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=156449

Meta Tags

- Meta tags are inserted into the source code of individual web pages, but are not visible to users.

```
<html>
<head>
<title>...</title>
<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">
</head>
```

- They allow you to control access on a page-by-page basis.
- Respectable bots will abide by the meta tags but troublemakers may ignore them.
- When Google sees a meta tag on a page, they will completely drop the page from the search results, even if other pages link to it.
- Can be used in conjunction with Robots.txt, but disallows in Robots.txt will override.

Google Web Search

To prevent all bots from indexing a page on your site, use the noindex meta tag:

```
<meta name="robots" content="noindex">
```

To allow all bots except Google’s to index a page on your site:

```
<meta name="googlebot" content="noindex">
```

Google Web Search

To prevent bots from following and indexing the links on your page, use nofollow:

```
<meta name="robots" content="nofollow">
```

To prevent search engines from displaying a description under your search listing, use nosnippet:

```
<meta name="robots" content="nosnippet">
```

Google Cache

To prevent all search engines from keeping a cached version of a page on your site, use the noarchive meta tag:

```
<meta name="robots" content="noarchive">
```

To allow all search engines except Google to cache a page on your site:

```
<meta name="googlebot" content="noarchive">
```

Combining Meta Tags

- Meta tags can be combined

- To prevent web indexing, caching, and following/indexing the links of a page:

```
<meta name="robots" content="noindex,noarchive,nofollow">
```

- To allow web indexing but prevent caching and following/indexing the links of a page:

```
<meta name="robots" content="noarchive,nofollow">
```

Google Groups

- Posts to Google Groups can provide valuable information to an attacker about your environment.
- If you have posted a help inquiry for software or system configurations, you may have exposed information that is helpful to someone profiling your network.
- You can prevent a message from displayed after seven days, or from being searchable in Google Groups after the initial seven day period.
- Posts must contain the text 'X-No-Archive: Yes' in the message header or on the very first line of the message body.

What If You Find Something That Shouldn’t Be Indexed?

Removing pages from the Google index and cache

- It is necessary to add the appropriate meta tags on the pages your are trying to remove from the Google Index or cache.

- Removing results may take several days to a few weeks.

- This process can be expedited by using Google's Webmaster Tools. https://www.google.com/webmasters/tools/home?hl=en

- If you do not want to use the Webmaster Tool's, you can use the 'Web Page Removal Request Tool': https://www.google.com/webmasters/tools/removals

- If the meta tags are in place, you can also just wait for the Google bot to revisit your page. No other action is necessary.

Removing content you don’t own

- Google requires you to work with the site owner to remove offending content or to insert the appropriate meta tags in their pages.

- If the site owner makes the changes, you can expedite the update process by submitting through the ‘Web Page Removal Request Tool’:
  https://www.google.com/webmasters/tools/removals

Removing content you don’t own

Google will provide you with assistance if a page contains:

- Your social security number or Government ID number
- Your bank account or credit card information
- An image of your handwritten signature
- Adult content in results when SafeSearch is enabled
- Inappropriate images on their featured video results
- Your full name or the name of your business appearing on an adult content site that's spamming Google's search results

Removing residential and business phone numbers

- Only applies to Google's phonebook listings; not phone numbers contained within 3rd party web pages

- To remove a residential phone number:

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

- To remove a business number, you have to send a signed written request

- Removal is permanent and it won't be possible to get the number relisted

Don’t Forget About…

The WayBackMachine
http://www.archive.org/index.php

Besides being indexed by search engines, your website may have been indexed by archive.org.

- Contains pages going back to 1996
- Doesn’t offer pages until the archived version is 6 months old
- Will abide by robots.txt and remove archived documents it disallows (User-agent: ia_archiver)

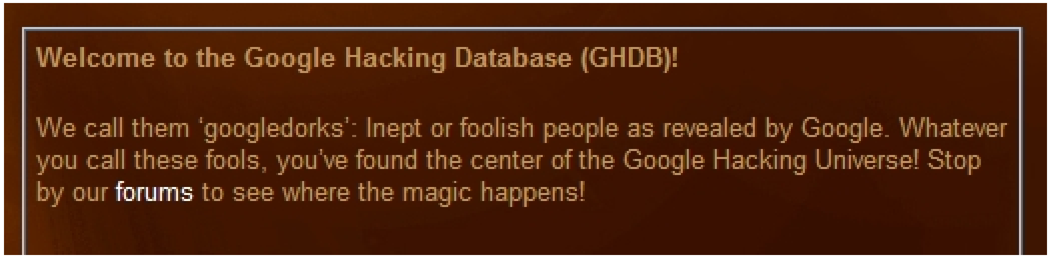
MOREnet
Community • Technology • Results

Want to learn more?

The Google Hacking Database

- Repository for queries used by the Google Hacking community.
- Used to be free, now requires a subscription.
- Wasn't frequently updated since 2006. This may change now that they require a subscription.
- Moderately useful, but a great way to learn how the community uses Google and what they are looking for.

http://johnny.ihackstuff.com/ghdb/



Welcome to the Google Hacking Database (GHDB)!

We call them 'googledorks': Inept or foolish people as revealed by Google. Whatever you call these fools, you've found the center of the Google Hacking Universe! Stop by our forums to see where the magic happens!

Great Resources

Google Hacking for Penetration Testers, Volume 2 (By Johnny Long)

http://www.amazon.com/Google-Hacking-Penetration-Testers-Johnny/dp/1597491764/ref=sr_1_3?ie=UTF8&s=books&qid=1262197372&sr=1-3

Book preview: Google Hacking for Penetration Testers, Volume 2 (By Johnny Long)

http://books.google.com/books?id=bvB1-MmhEjQC&lpg=PP1&pg=PA50#v=onepage&q=&f=false

Google Guide (syntax and advanced operators)

http://www.googleguide.com/