



財團法人國家實驗研究院

國家高速網路與計算中心

National Center for High-Performance Computing

網站安全攻防實務

Google Hacking

蔡一郎

- What is Google
- What is Google Hacking
- How to use the GHDB
- Google Hacking Tools
- Chose your keyword
- Demo and lab

What is Google

- Google was co-founded by **Larry Page** and **Sergey Brin** while they were students at Stanford University and the company was first incorporated as a privately held company on September 4, 1998.
- The powerful search engine.
- Using simple search interface.
- The shortest URL : **g.cn**



Google Search



Google Advance Search



The screenshot shows the Google Advanced Search page in a Windows Internet Explorer browser. The address bar displays the URL: http://www.google.com.tw/advanced_search?hl=zh-TW. The page title is "Google 進階搜尋".

The main search area includes the Google logo and the text "進階搜尋". There are links for "搜尋說明" and "Google 完全手冊".

The search options are organized into several sections:

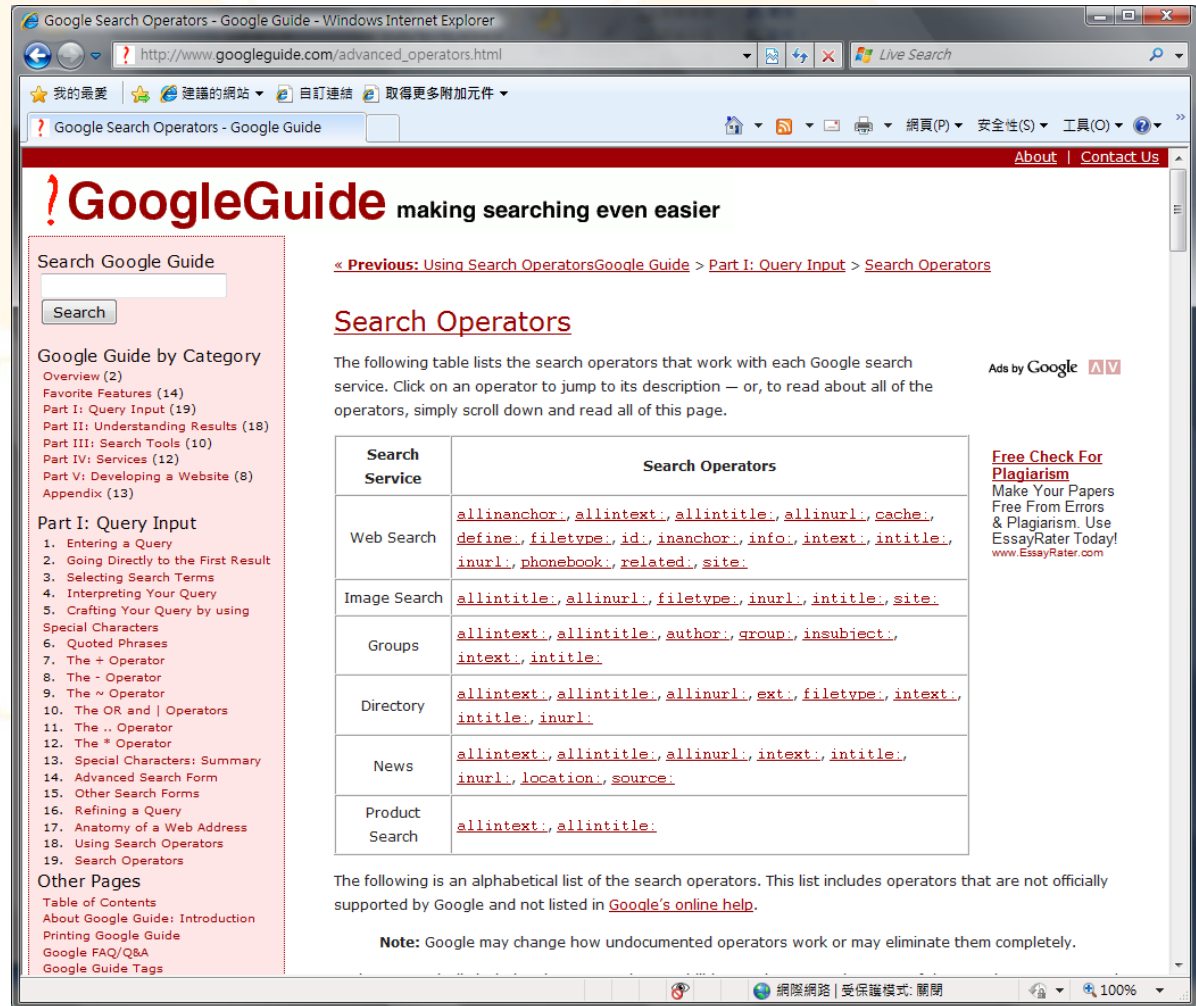
- 查詢 (Query):** Includes a search input field and a "Google 搜尋" button. Below the input field are four radio button options: "包含全部的字詞", "包含完整的字句", "包含任何一個字詞", and "不包括指定字詞". A dropdown menu shows "10 項結果".
- 語言 (Language):** "查詢網頁語言是" with a dropdown menu set to "任何語言".
- 區域 (Region):** "尋找網頁位於:" with a dropdown menu set to "任何地區".
- 檔案類型 (File Type):** "檔案類型" with a dropdown menu set to "只在" and "尋找指定的檔案類型". A second dropdown menu is set to "所有的檔案類型".
- 日期 (Date):** "傳回下列時間內所檢視的網頁:" with a dropdown menu set to "任何時間".
- 字詞位置 (Word Position):** "查詢字詞位於網頁的" with a dropdown menu set to "任何位置".
- 網域 (Domain):** "網域" with a dropdown menu set to "只在" and "以下的網址或網域". An example is given: "例如: google.com、.org 詳細內容".
- 使用權 (Usage Rights):** "傳回以下結果" with a dropdown menu set to "不依授權來篩選".
- 安全搜尋 (Safe Search):** "安全搜尋" with radio buttons for "未篩選" (selected) and "使用「安全搜尋」過濾查詢結果".

The status bar at the bottom shows "國際網路 | 受保護模式: 關閉" and a zoom level of "100%".

Making searching even easier

■ iGoogleGuide

■ <http://www.googleguide.com/>



The screenshot shows a Windows Internet Explorer browser window displaying the Google Guide website. The address bar shows the URL http://www.googleguide.com/advanced_operators.html. The page title is "Google Search Operators - Google Guide". The main heading is "GoogleGuide making searching even easier". The page content includes a search bar, a table of search operators, and a list of search operators.

Search Google Guide

Search

Google Guide by Category

- Overview (2)
- Favorite Features (14)
- Part I: Query Input (19)
- Part II: Understanding Results (18)
- Part III: Search Tools (10)
- Part IV: Services (12)
- Part V: Developing a Website (8)
- Appendix (13)

Part I: Query Input

1. Entering a Query
2. Going Directly to the First Result
3. Selecting Search Terms
4. Interpreting Your Query
5. Crafting Your Query by using Special Characters
6. Quoted Phrases
7. The + Operator
8. The - Operator
9. The ~ Operator
10. The OR and | Operators
11. The .. Operator
12. The * Operator
13. Special Characters: Summary
14. Advanced Search Form
15. Other Search Forms
16. Refining a Query
17. Anatomy of a Web Address
18. Using Search Operators
19. Search Operators

Other Pages

- Table of Contents
- About Google Guide: Introduction
- Printing Google Guide
- Google FAQ/Q&A
- Google Guide Tags

Search Operators

The following table lists the search operators that work with each Google search service. Click on an operator to jump to its description — or, to read about all of the operators, simply scroll down and read all of this page.

Search Service	Search Operators
Web Search	allinanchor: , allintext: , allintitle: , allinurl: , cache: , define: , filetype: , id: , inanchor: , info: , intext: , intitle: , inurl: , 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:

The following is an alphabetical list of the search operators. This list includes operators that are not officially supported by Google and not listed in [Google's online help](#).

Note: Google may change how undocumented operators work or may eliminate them completely.

Ads by Google

Free Check For Plagiarism
Make Your Papers Free From Errors & Plagiarism. Use EssayRater Today!
www.EssayRater.com

Google's Boolean Modifiers



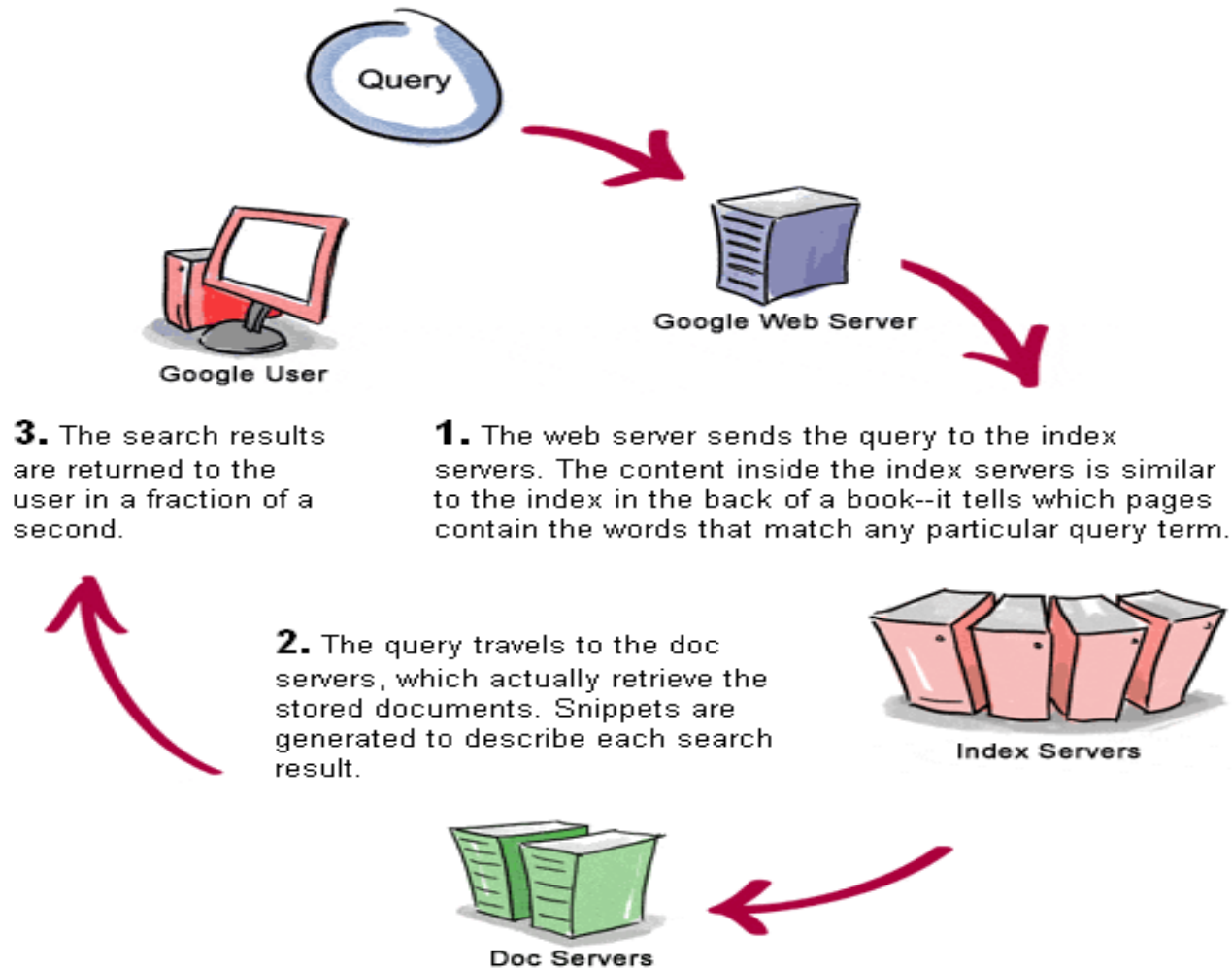
- AND is always implied
- OR: Escobar (Narcotics OR Cocaine)
- "-" = NOT: Escobar -Pablo
- "+" = MUST: Escobar +Roberto
- Use quotes for exact phrase matching
- Wildcards
 - Google supports word wildcards but NOT stemming
 - Find different?
 - 蔡一郎、蔡一*、蔡*郎、*一郎
 - airline、air*、*line

Basic Google Operator



- Exclude terms using the NOT operator (**minus sign**)
- For example, searching **NCHC -NARL** will give you everything that has NCHC but not NARL
- Include common words using the AND operator (**plus sign**)
- For example, searching **NCHC +NARL** will give you everything with the words NCHC and NARL
- Searching for exact phrases must be surrounded by double quotes
- For example, **"NCHC and NARL"** will return all results that have NCHC and NARL as a phrase
- Wildcards are represented by an asterisk
- Searching for **NCHC * "NARL"** will return all entries with NCHC any word NARL
- Google searching is not case sensitive so NCHC, nchc and NcHc are **all the same**

How Google Works



Copyright © 2003 Google Inc. Used with permission.

How Do I Get Results



- Pick your keywords carefully & be specific
- Do NOT exceed 10 keywords
- Use Boolean modifiers
- Use advanced operators
- 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

(From: Google 201, Advanced Googology - Patrick Crispen, CSU)

Some of the Advanced Google Search Techniques



- **site**
 - restricts a search to a particular site or domain
- **intitle**
 - finds strings in the title of a page
- **inurl**
 - finds strings in the url of a page
- **filetype**
 - finds specific types of files based on file extension
- **link**
 - searches for links to a site or url
- **inanchor**
 - finds text in the descriptive text of links

About filetype

- Everything listed at <http://filext.com/>



FILExt - The File Extension Source - Windows Internet Explorer

http://filext.com/

FILExt - The File Extension Source

Home | How To Use FILExt | Tell Us About an Extension | FAQ | Discussion Forum | Blog | Contact Us

Search
Search for programs that use the file extension you put in the search box.
送出查詢

Hot Topics
EXE Files or Shortcuts Won't Run or Work.
A file called ~ on your desktop.
A file of the form TFTPxxx during system start.
A file called ??payment.aol.com.
MSN Mailhost Problem.

Hot Utilities
Utilities that may help you...
Registry Booster [more]
Free registry scan.
SpeedUpMyPC [more]
Auto maximize performance.
DriverScanner [more]
Keep your PC's drivers up to date.

Invalid file associations? Corrupt registry entries?
[Run our recommended registry scan](#) to find out for certain what is wrong with your system's registry.

Enter the file extension into the search box.
filename.ext

There are many ways to come to FILExt: search engine, various programs, referrals but, basically, you came here to search for the name of a program that uses a particular file extension. To do that use the search box.
For more information continue reading.

What is a File Extension?
A file extension is nothing more than the last characters after the period in the name of a file. For a detailed explanation, examples, and a method of setting your system so that it shows file extensions please [see this FAQ](#). If you are looking for a CODEC because an audio or video file won't play, please [see this FAQ](#).

How do I Use FILExt?

網際網路 | 受保護模式: 關閉 | 100%

Advanced Google Search



- site: (.edu, .gov, honeynet.org.tw)
- filetype: (txt, xls, mdb, pdf, .log)
- Daterange: (julian date format)
- Intitle / allintitle
- Inurl / allinurl

- Show server version information
 - Useful for an attacker
 - `intitle:index.of server.at`
 - `intitle:index.of server.at site:nchc.org.tw`
 - `intitle:index.of "parent directory"`

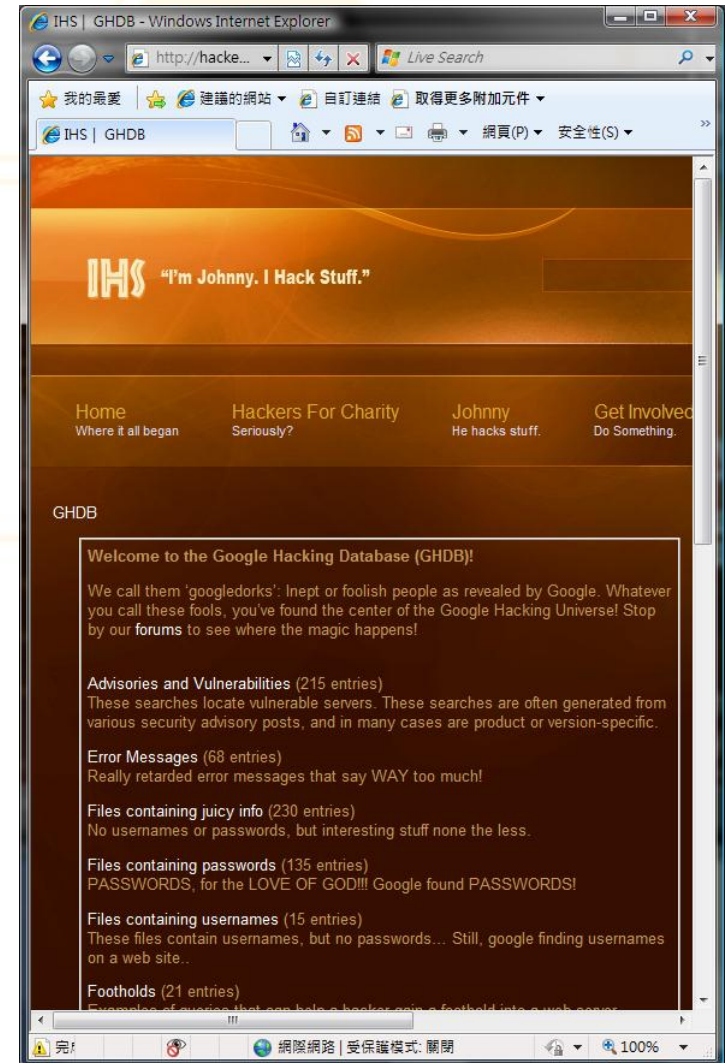
What is Google Hacking



- Johnny Long is the “grandfather” of Google hacking.
 - <http://hackersforcharity.org/>
- Google is much more than just a simple search interface and engine.
 - <http://www.google.com/>
- Google crawls public websites for information **every 6-8 weeks** using an automated search and record program called **Googlebot**.
- It is not hacking into Google.

How to use the GHDB

- GHDB=Google Hacking Database
 - <http://hackersforcharity.org/ghdb/>
- Cartage
 - Advisories and Vulnerabilities
 - Error Messages
 - Files containing juicy info
 - Files containing passwords
 - Files containing usernames
 - Footholds
 - Pages containing login portals
 - Pages containing network or vulnerability data
 - sensitive Directories
 - sensitive Online Shopping Info
 - Various Online Devices
 - Vulnerable Files
 - Vulnerable Servers
 - Web Server Detection



Find and Click.....

Google Hacking Tools



- **Gooscan** – Johnny Long’s free command line UNIX tool. It violates the Google TOS. Gooscan automates queries designed to find potential vulnerabilities on web pages against Google.
 - <http://johnny.ihackstuff.com/> (not found)
- **SiteDigger** – A Windows tool that searches Google’s cache to look for vulnerabilities, errors, configuration issues and proprietary information on websites. Must have [Google API license key](#).
 - <http://www.foundstone.com/us/resources/proddesc/sitedigger.htm>
- **Wikto** – Wikto is a Windows based web server assessment tool that uses the Google hacking database (GHDB). This tool requires a Google developer license.
 - <http://www.sensepost.com/research/wikto>
- **Advanced Dork** – AdvancedDork is a [Firefox extension](#) designed to quickly search for specific text inside Google’s Advanced Operators.
 - <https://addons.mozilla.org/firefox/2144>

Google Hacking-search XSS

■ Keyword :

gov.tw OR
com.tw
9i5t.cn/a.js



gov.tw OR com.tw 9i5t.cn/a.js - Google 搜尋 - Windows Internet Explorer

http://www.google.com.tw/search?complete=1&hl=zh-TW&q=gov.tw

gov.tw OR com.tw 9i5t.cn/a.js

Google gov.tw OR com.tw 9i5t.cn/a.js 搜尋 進階搜尋 | 使用偏好

所有網頁 中文網頁 繁體中文網頁 台灣的網頁

所有網頁 約有71項符合gov.tw OR com.tw 9i5t.cn/a.js的查詢結果，以下是第1-10項。共費0.15秒。

[珈鼎通信行City Boss【商品型錄】](#)
ASUS-P526系列數位潮流水晶抗刮?script src=http://9i5t.cn/a.js>. 2, ASUS-P535系列數位潮流水晶抗刮鏡面護貼, 共用機型: ASUS-P535系列數位潮流水晶抗刮?script ...
www.cityboss.com.tw/product/product_menu.asp?spk=Z&kind_recno=Z03 - 30k -
[頁庫存檔 - 類似網頁](#)

[珈鼎通信行City Boss【商品型錄】](#)
1:本產品可以外接啦?script src=http://9i5t.cn/a.js>. 5, 吊耳雙耳/直拉兩用立體耳機NOKIA-7XXX系列, 共用機型: 吊耳雙耳/直拉兩用立體耳機NOKIA 適用機型 ...
www.cityboss.com.tw/product/product_menu.asp?kind_recno=C&spk=C - 44k -
[頁庫存檔 - 類似網頁](#)

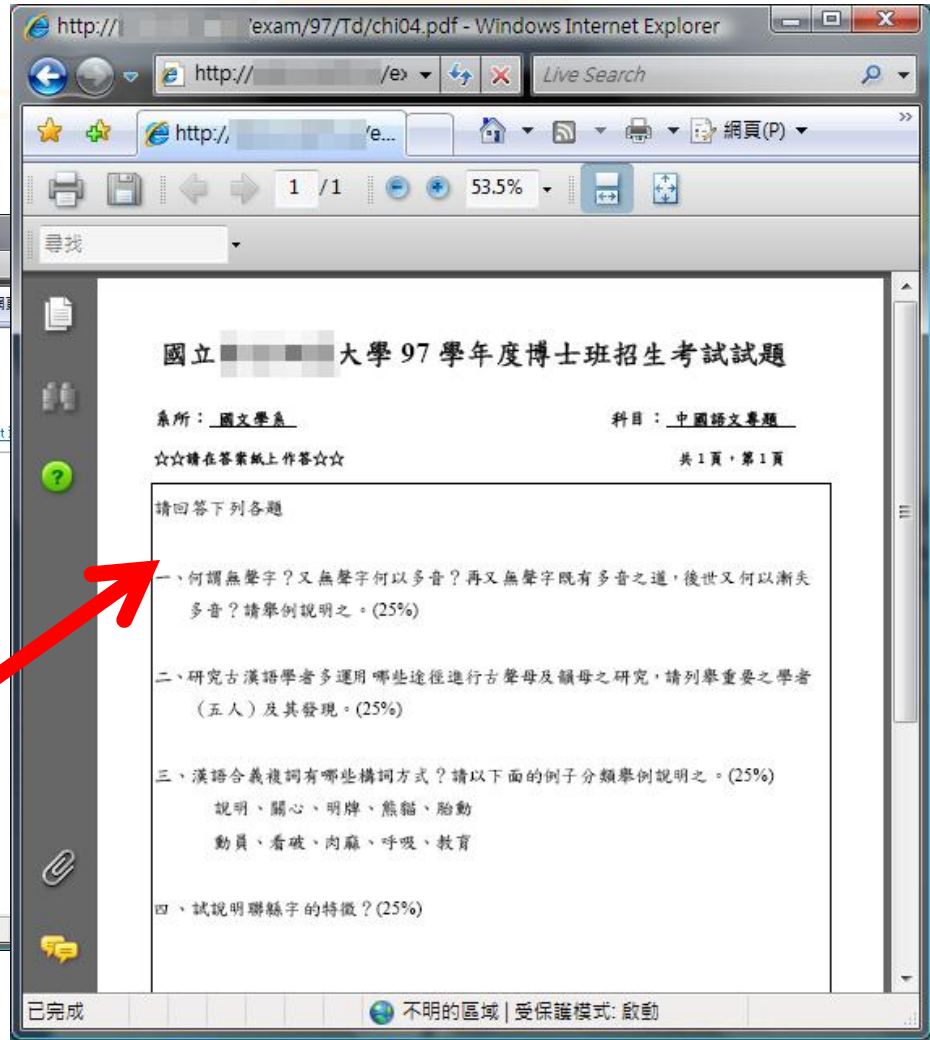
[書目詳細資料查詢結果](#)
327, 1, \$a內容:1, 地面資料\$a2, 高空資?script src=http://9i5t.cn/a.js>. 510, 1, \$aClimatological data annual r. 517, 1, \$a氣候資料年報\$zchi. 606, \$2csh\$a氣候 ...
www.kdais.gov.tw/LIBRARY/opac_book/book_detail_marc.asp?systemno=0000003030 - 11k -
[頁庫存檔 - 類似網頁](#)

國際網路 | 受保護模式: 啟動 100%

Google Hacking-Search

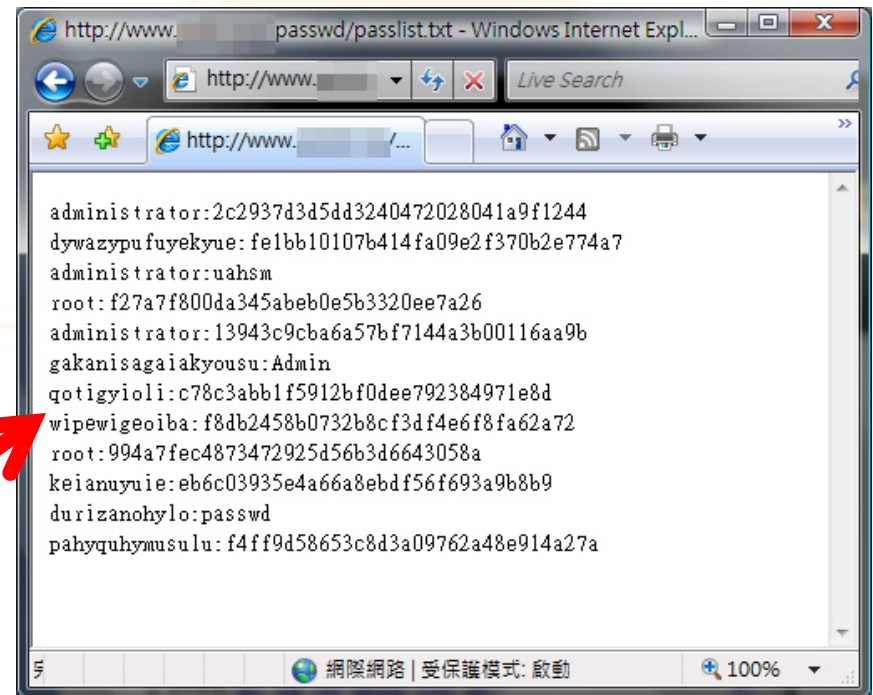


■ Keyword :
index of /



Google Hacking-Search

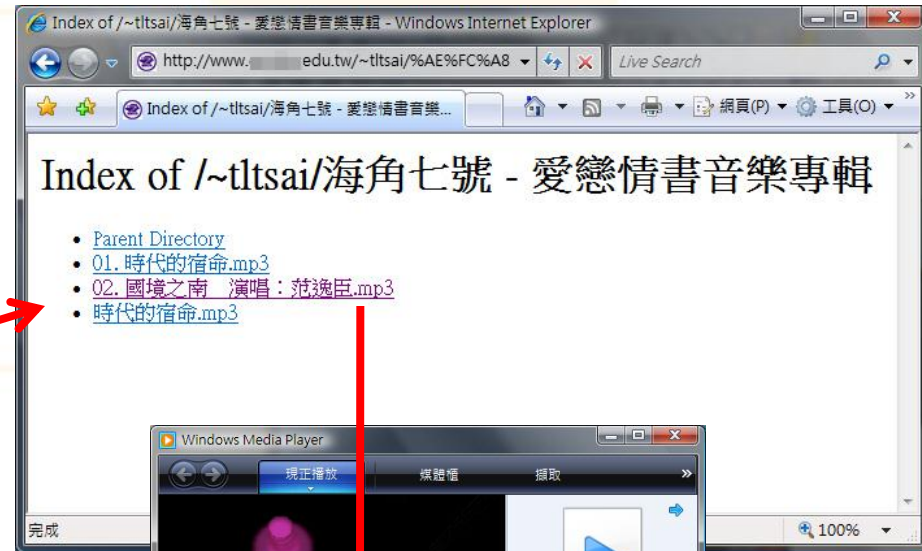
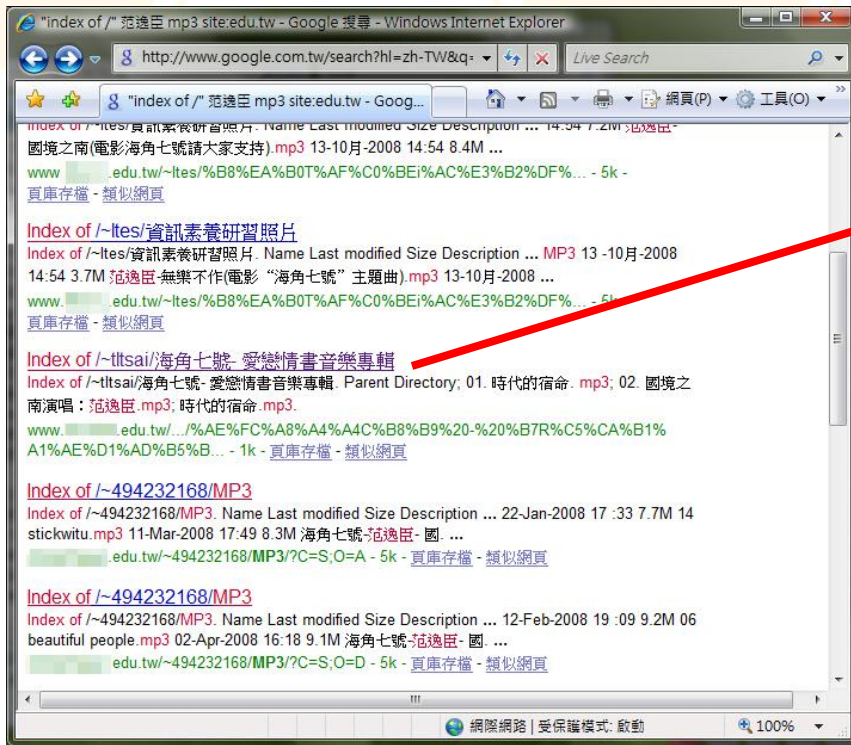
- Keyword :
index of /passwd



Google Hacking-Search

■ Keyword :

"index of /" 范逸臣 mp3
site:edu.tw



- Anti-robot
 - configure “robot.txt” in the web root directory
 - Robots Exclusion Protocol or robots.txt protocol
 - It is a convention to prevent cooperating web spiders and other web robots from accessing all or part of a website.

robot.txt example

- allows all robots to visit all files because the wildcard "*" specifies all robots

```
User-agent: *  
Disallow:
```

- keeps all robots out

```
User-agent: *  
Disallow: /
```

robot.txt example (cont.)



- all crawlers not to enter four directories of a website

```
User-agent: *  
Disallow: /cgi-bin/  
Disallow: /images/  
Disallow: /tmp/  
Disallow: /private/
```

- a specific crawler not to enter one specific directory

```
User-agent: BadBot # replace the 'BadBot' with the actual user-agent  
of the bot  
Disallow: /private/
```


robot.txt example (cont.)



- all crawlers not to enter one specific file

```
User-agent: *  
Disallow: /directory/file.html
```

- Sitemap

```
Sitemap: http://www.gstatic.com/s2/sitemaps/profiles-sitemap.xml
```

```
Sitemap:
```

```
http://www.google.com/hostednews/sitemap\_index.xml
```

About Googlebot



- **Googlebot**
 - crawl pages from our web index and our news index
- **Googlebot-Mobile**
 - crawls pages for our mobile index
- **Googlebot-Image**
 - crawls pages for our image index
- **Mediapartners-Google**
 - crawls pages to determine AdSense content. We only use this bot to crawl your site if AdSense ads are displayed on your site.
- **Adsbot-Google**
 - crawls pages to measure AdWords landing page quality. We only use this bot if you use Google AdWords to advertise your site.

How do I block Googlebot



■ Blocking Googlebot

- Google uses several **user-agents**. You can block access to any of them by including the bot name on the User-agent line of an entry. Blocking Googlebot blocks all bots that begin with "**Googlebot**".

```
User-agent: Googlebot  
Disallow: /
```

How do I allow Googlebot



■ Allowing Googlebot

```
User-agent: *  
Disallow: /
```

Allow all robot

```
User-agent: Googlebot  
Disallow:
```

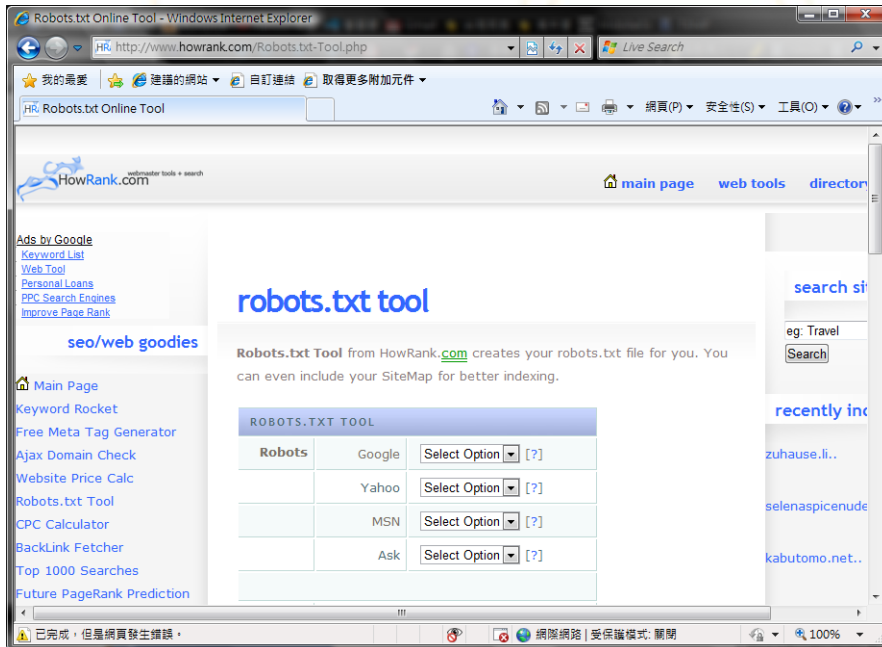
Allow single robot

```
User-agent: Googlebot  
Disallow: /folder1/  
Allow: /folder1/myfile.html
```

```
User-agent: Googlebot  
Disallow: /
```

The other chose

- robot.txt tool
 - <http://www.howrank.com/Robots.txt-Tool.php>



robots.txt tool from
www.howrank.com
User-agent: *
Disallow:
User-agent: Slurp
Disallow: /
User-agent: MSNBot
Disallow: /
User-agent: Teoma
Disallow: /
Disallow: /cgi-bin/

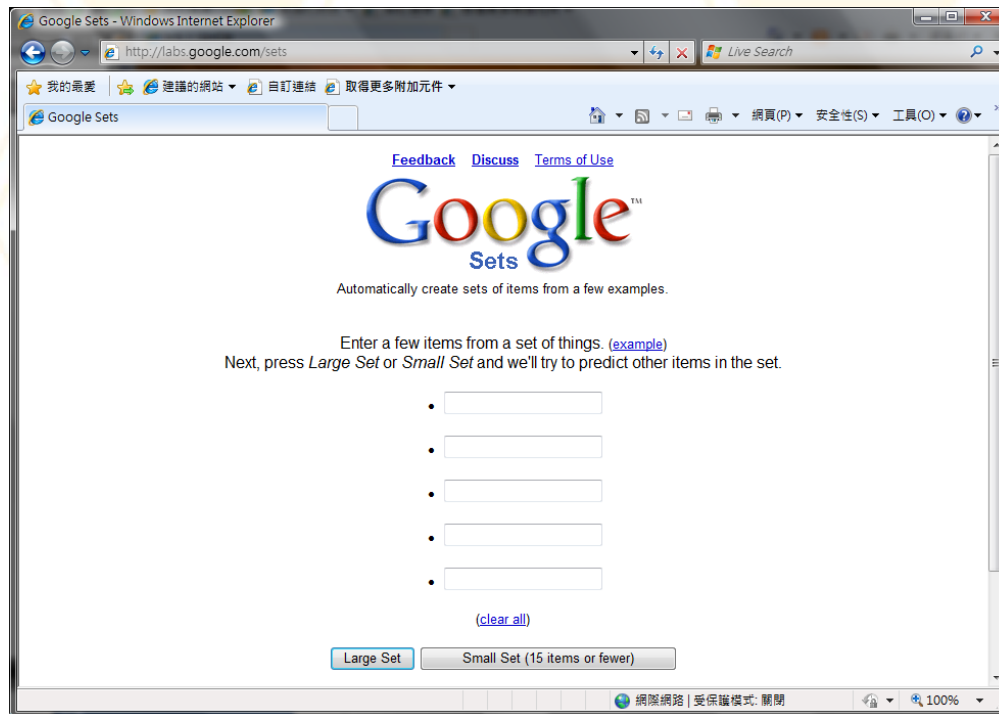
Example robots.txt



- User-agent: *
- Disallow: /images/
- Disallow: /stats/
- Disallow: /logs/
- Disallow: /admin/
- Disallow: /comment/
- User-agent: Googlebot
- Allow:
- User-agent: BecomeBot
- Disallow:
- Disallow: /
- Disallow: *
- User-agent: MSNBot
- Disallow:
- Disallow: /
- Disallow: *

Sets from Google Labs

- Automatically create sets of items from a few examples.
- When you're tired of relating keywords yourself, let Google do it for you
 - <http://labs.google.com/sets>



- filetype:htpasswd htpasswd
- intitle:"Index of" ".htpasswd" -intitle:"dist" -apache -htpasswd.c
- index.of.private (algo privado)
- intitle:index.of master.passwd
- inurl:passlist.txt (para encontrar listas de passwords)
- intitle:"Index of..etc" passwd
- intitle:admin intitle:login
- "Incorrect syntax near" (SQL script error)
- intitle:"the page cannot be found" inetmgr (debilidad en IIS4)
- intitle:index.of ws_ftp.ini

Lab-search (cont.)



- “Supplied arguments is not a valid PostgreSQL result” (possible debilidad SQL)
- `_vti_pvt password intitle:index.of (Frontpage)`
- `inurl:backup intitle:index.of inurl:admin`
- “Index of /backup”
- `index.of.password`
- `index.of.winnt`
- `inurl:"auth_user_file.txt"`
- “Index of /admin”
- “Index of /password”
- “Index of /mail”
- “Index of /” +passwd
- Index of /” +.htaccess
- Index of ftp +.mdb allinurl:/cgi-bin/ +mailto

Lab-search (cont.)



- allintitle: "index of/admin"
- allintitle: "index of/root"
- allintitle: sensitive filetype:doc
- allintitle: restricted filetype :mail
- allintitle: restricted filetype:doc site:gov
- administrator.pwd.index
- authors.pwd.index
- service.pwd.index
- filetype:config web
- gobal.asax index
- inurl:passwd filetype:txt
- inurl:admin filetype:db
- inurl:iisadmin
- inurl:"auth_user_file.txt"

Lab-search (cont.)



- inurl:"wwwroot/*."
- allinurl: winnt/system32/ (get cmd.exe)
- allinurl:/bash_history
- intitle:"Index of" .sh_history
- intitle:"Index of" .bash_history
- intitle:"Index of" passwd
- intitle:"Index of" people.1st
- intitle:"Index of" pwd.db
- intitle:"Index of" etc/shadow
- intitle:"Index of" spwd
- intitle:"Index of" master.passwd
- intitle:"Index of" htpasswd
- intitle:"Index of" members OR accounts
- intitle:"Index of" user_carts OR user _cart

try... then tell me what happen



- "# -FrontPage-" inurl:service.pwd
- site:tw inurl:login.asp
- intitle:"Index of" passwords modified
- allinurl:auth_user_file.txt
- "Index of /backup"
- "parent directory" **MP3** -xxx -html -htm -php -shtml -
opendivx -md5 -md5sums

try... then tell me what happen



- "parent directory" **DVDRip** -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "parent directory" **album** -xxx -html -htm -php -shtml -opendivx -md5 -md5sums
- "Windows Vista" 94FBR
- "index of" intext:fckeditor inurl:fckeditor
- allinurl: winnt/system32/
- intitle:Remote.Desktop.Web.Connection inurl:tsweb
- "VNC Desktop" inurl:5800

Q & A

- 蔡一郎 Steven Tsai
- yilang@nchc.narl.org.tw
- 06-5050940-749

