Foundstone SiteDigger 2.0™
Identifying Information Leakage Using Search Engines

By Kartik Trivedi

January 2005

Public Information Gathering Tools

The Story So Far…

Search engines are the main stay of the Internet. Throughout the economic cycles in the dot com boom and bust, they have continued to grow as legitimate "for profit" businesses reaching out to all of those who use the Internet. As relevance becomes an increasingly important concept, cutting through the volume of data that has migrated itself to the web, companies like Yahoo and Google have become household brands in the same way that it took Coca-Cola decades.

And the search engine revolution has certainly not been missed by hackers and security professionals looking for new ways to ferret out security nuggets on the digital matrix. At Foundstone we have long used public search engines to 'footprint' companies and examine public data that they may not want to have in the public eye. Embarrassing and confidential information too often finds its way onto the Internet opening up security holes and allowing information leakage without anyone's attention.

Recently, growing attention of hacking with Google, known as "Google hacking" has emerged and lists of signatures continue to appear on security sites in an intriguing game of cunning. People are developing new ways to use the limited functionality of search engines and their vast banks of seemingly unlimited data to unveil security flaws that can have a significant impact to businesses.

So exactly how do you use search engines to find security vulnerabilities in web sites? This paper walks you through some of the techniques and will show you some examples using Google. It might just as well have been Yahoo, MSN or any number of the other search engines out there but we chose Google.

We then introduce SiteDigger 2.0™, a free Windows tool that automates Google security queries to the Google web service API.

Information Leakage

As we have all learned, the Internet is a powerful, yet very dangerous environment. Too often, information that is not meant for others to see find its way onto Web sites and hence become free for all to find. This information leakage can have very damaging results and puts the organization at risk, often without anyone knowing it.

Password files, vulnerability scan reports, configuration errors, and other security issues are the treasures that hackers are hunting to find. Hackers have now turned to search engines like Google to make their searches easier.

How Google Works

Google’s search technology works by crawling the Internet, taking snapshots of each web page it examines and caching the results. When a search query is performed using Google, the query is conducted on Google’s cached pages. This is import because it allows for very quick search results and it allows people to use Google’s syntax options to conduct specific types of searches. Thus setting the foundation for what has become known as “Google hacking”.

Using Google

This section provides you with an overview of the syntax that is used by Google. It is not intended to be exhaustive but will explain the basic concepts of how search queries are constructed and how results are derived.

> **Note:** Words and phrases surrounded by square brackets and in italics denote search terms.
> *For example:* [search term here]

The Basic Syntax

Keyword Searches

Basic keyword searching with Google is very simple. You enter the keywords you are looking for and submit the query. It really is that simple. As an example lets use the keywords [*ask the oracle*].

Phrase Searches

You can refine searches by using phrases rather than words. To do so, you simply encapsulate the phrase within double-quotes such as [*“ask the oracle”]*.

The AND Operator

Most search engines default to using the AND operator to concatenate queries. In fact we would use this technique to find Larry Ellison. It is actually the same as asking for [*Larry AND Ellison*].

As you will come to find out, simple search doesn’t always turn up what you want, so the kind folks in search engine engineering land have provided us with a nice set of Boolean operators to further refine our search.

The Addition Operator

By default Google ignores certain words that are considered not helpful in searching. These are called stop words. These include ‘the’, ‘I’ and ‘a’. The plus operator allows us to search for [*Larry Ellison+the+oracle*] and thus include stop words.

The Minus Operator

The minus operator allows us to search for pages that include Larry Ellison but not Oracle [*“Larry Ellison” –Oracle*]. The negation operator is extremely useful when analyzing false positives when Google hacking.

The Wild Card

Google doesn’t support a technique called stemming where the search [*mar*+Curphey*] would find Mark Curphey and Marcus Curphey. It does however support full word wild cards using the * operator.

The Advanced Syntax

As well as the basic syntax, Google provides a rich set of additional query terms that allows you to request exactly the data you want. Using this advanced syntax is where Google really comes into its own as a powerful security scanner and is the basis for the SiteDigger 2.0™ tool.

intitle:

Restricts the search term to appear only in the title on the page. This turns out to be very useful, as many technologies create default pages such as "Terminal Services Web Connection", “Welcome to IIS 4.0!” or “Outlook Web Access”. Examples include:

[*intitle:“Terminal Services Web Connection”*]
[*intitle:“Welcome to IIS 4.0!”*]
[*intitle:“Outlook Web Access”*]

intext:

Restricts the search term to the body of the text itself, ignoring titles and URL’s. This is very useful when combined with other strings.

insite:

Restricts the search to a specific domain. This is particularly useful for focusing on vulnerabilities in specific domain names or top level domains such as “.mil”. SiteDigger 2.0 leverages insite: extensively to narrow the search space to web sites of interest. Examples include:

[*insite: gov*]
[*insite: securitycompany.net*]

inurl:

Restricts the search string to be present in the URL of a resource. This is particularly useful when looking for special characters in URL's such as @ sign's used in http password syntax and file locations such as:

[*inurl: /etc/passwd*]

cache:

Can be used to find older versions of sites where content changes frequently and may no longer be available. This technique is very interesting to use on sites that have recently repaired vulnerabilities or had content removed from a site after a security incident.

Filetype

This is a powerful construct and uses the file extension to identify results. Uses include searching for Excel spreadsheets (.xls) and web technology types such as .asp or .cfm.

Some of the most powerful queries include:

[*password filetype: xls*]
[*confidential filetype: doc*]
[*secret filetype: ppt*]
[*administrator filetype: cfm*]

The ALL Syntax

The advanced syntax can be used with multiple search words to encapsulate the query more accuracy. As an example [*allintitle: test login*] will find both the words test and login in the title of a page. Prefixing all is not always a suitable technique when mixing syntax together into complex search queries.

Google Examples

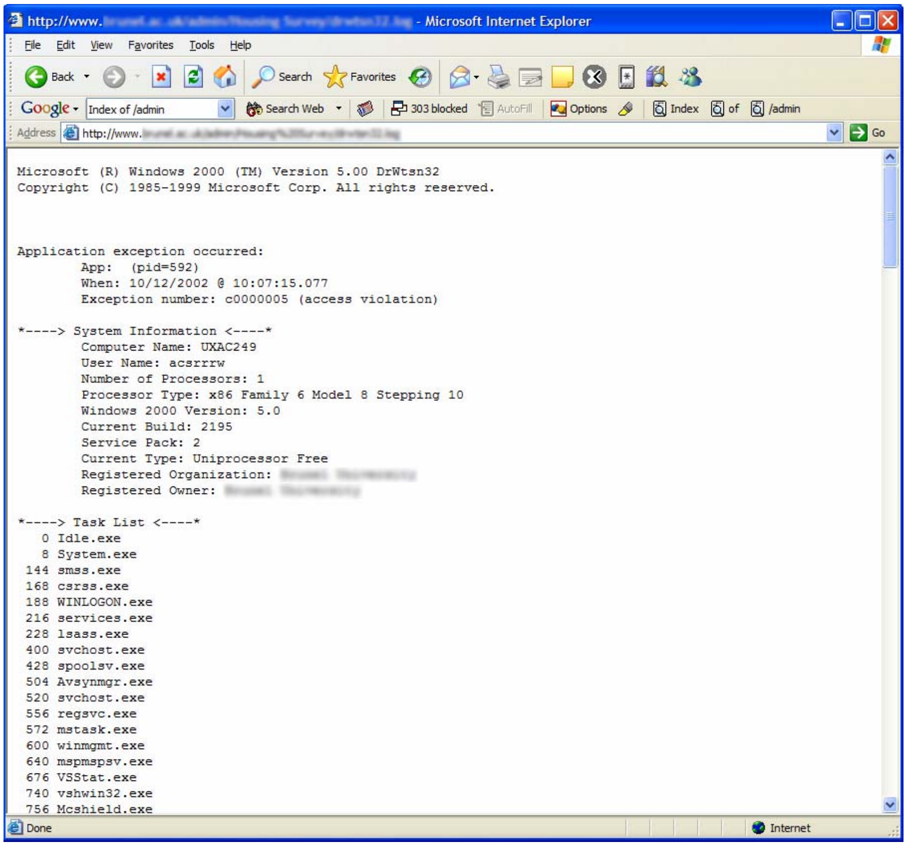
To illustrate the syntax described above we will examine some example queries and results.

Error Handling - "Index of /admin"

Are you feeling lucky?

Many web servers generate errors and exceptions to log files. These often get written into a published directory and therefore become accessible via the web. The "Index of" query is very useful for finding these files. By issuing an [*"Index of /admin"*] query we get a set of results of pages that are directory listings and at a URL in the form "http://domain/admin".

Directory listings are often interesting as they list all files in the directory and not just those that are intentionally hyperlinked. In the screen shot below you can see that inside one of the resulting pages from this query was a Dr. Watson error log.



http://www.[redacted] - Microsoft Internet Explorer

Google: Index of /admin

Address: http://www.[redacted]

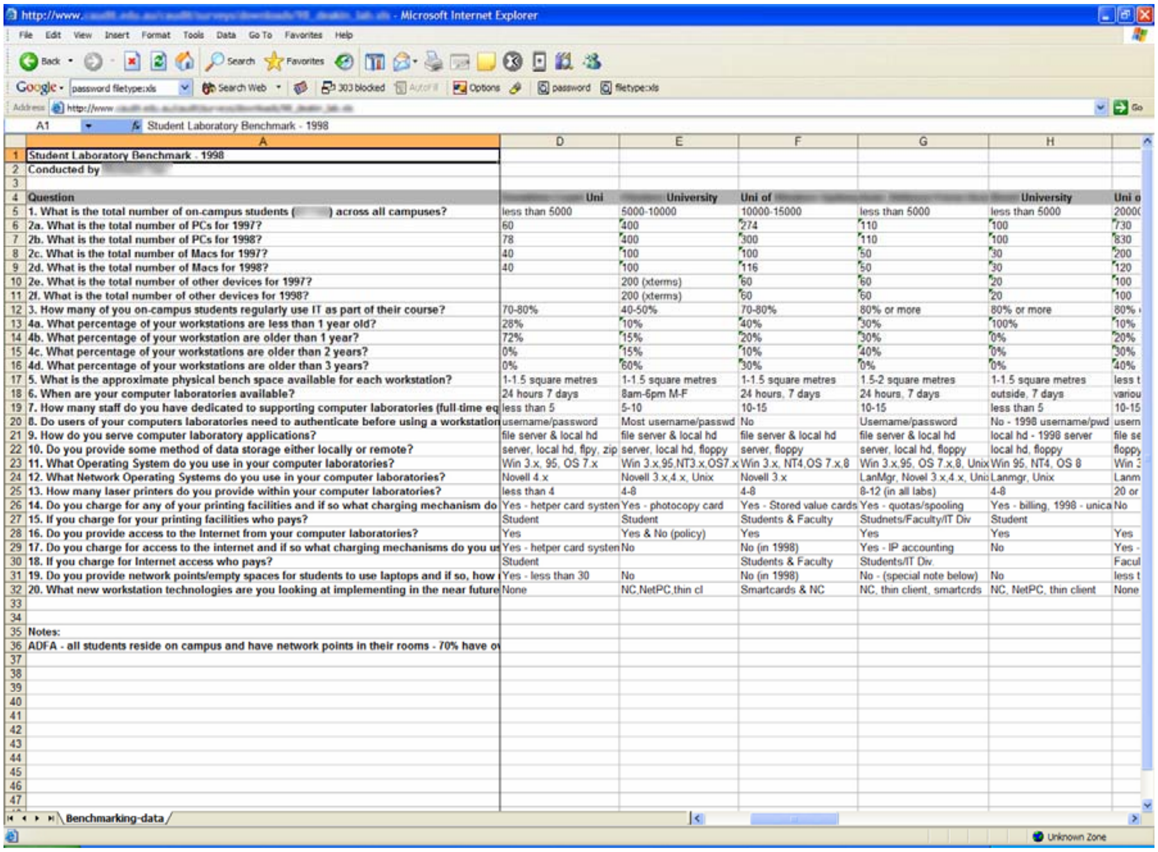
```
Microsoft (R) Windows 2000 (TM) Version 5.00 DrWtsn32
Copyright (C) 1985-1999 Microsoft Corp. All rights reserved.



Application exception occurred:
        App:  (pid=592)
        When: 10/12/2002 @ 10:07:15.077
        Exception number: c0000005 (access violation)

*----> System Information <----*
        Computer Name: UXAC249
        User Name: acsrrrw
        Number of Processors: 1
        Processor Type: x86 Family 6 Model 8 Stepping 10
        Windows 2000 Version: 5.0
        Current Build: 2195
        Service Pack: 2
        Current Type: Uniprocessor Free
        Registered Organization: [redacted]
        Registered Owner: [redacted]

*----> Task List <----*
   0 Idle.exe
   8 System.exe
 144 smss.exe
 168 csrss.exe
 188 WINLOGON.exe
 216 services.exe
 228 lsass.exe
 400 svchost.exe
 428 spoolsv.exe
 504 Avsynmgr.exe
 520 svchost.exe
 556 regsvc.exe
 572 mstask.exe
 600 winmgmt.exe
 640 mspmspsv.exe
 676 VSStat.exe
 740 vshwin32.exe
 756 Mcshield.exe
```

You will notice on further inspection that there is a lot of interesting information in this log file. Dr. Watson log files are written when the Windows operating system deals with an exception. This log file tells us that the host is a Windows 2000 machine with service pack two, the username of the logged in user, the fact that McAfee anti-virus is running and Microsoft SMS is used to manage the host. And that was just for starters.

While on client engagements, Foundstone Professional Services consultants have found web server log files that actually contained the credit card details of the clients' customers (GET requests can be logged by the web server), database errors including the connection strings, and administrative instructions on how to login and administer the system.

Office Files – [password filetype:xls]

Search engines don't just index html. In fact Google boasts the ability to parse over 200 MIME types (message transport types for files) from Microsoft Word documents to PDF files. By issuing the [*password filetype:xls*] query we found a selection of Excel spreadsheets that contained the text password.



A1 Student Laboratory Benchmark - 1998

	A	D	E	F	G	H	
1	Student Laboratory Benchmark . 1998						
2	Conducted by [redacted]						
3							
4	Question	[redacted] Uni	[redacted] University	Uni of [redacted]	[redacted]	[redacted] University	Uni o
5	1. What is the total number of on-campus students ([redacted]) across all campuses?	less than 5000	5000-10000	10000-15000	less than 5000	less than 5000	2000(
6	2a. What is the total number of PCs for 1997?	60	400	274	110	100	730
7	2b. What is the total number of PCs for 1998?	78	400	300	110	100	830
8	2c. What is the total number of Macs for 1997?	40	100	100	50	30	200
9	2d. What is the total number of Macs for 1998?	40	100	116	50	30	120
10	2e. What is the total number of other devices for 1997?		200 (xterms)	60	60	20	100
11	2f. What is the total number of other devices for 1998?		200 (xterms)	60	60	20	100
12	3. How many of you on-campus students regularly use IT as part of their course?	70-80%	40-50%	70-80%	80% or more	80% or more	80% (
13	4a. What percentage of your workstations are less than 1 year old?	28%	10%	40%	30%	100%	10%
14	4b. What percentage of your workstation are older than 1 year?	72%	15%	20%	30%	0%	20%
15	4c. What percentage of your workstations are older than 2 years?	0%	15%	10%	40%	0%	30%
16	4d. What percentage of your workstations are older than 3 years?	0%	60%	30%	0%	0%	40%
17	5. What is the approximate physical bench space available for each workstation?	1-1.5 square metres	1-1.5 square metres	1-1.5 square metres	1.5-2 square metres	1-1.5 square metres	less t
18	6. When are your computer laboratories available?	24 hours 7 days	8am-6pm M-F	24 hours, 7 days	24 hours, 7 days	outside, 7 days	variou
19	7. How many staff do you have dedicated to supporting computer laboratories (full-time eq	less than 5	5-10	10-15	10-15	less than 5	10-15
20	8. Do users of your computers laboratories need to authenticate before using a workstation	username/password	Most username/passwd	No	Username/password	No - 1998 username/pwd	usern
21	9. How do you serve computer laboratory applications?	file server & local hd	file server & local hd	file server & local hd	file server & local hd	local hd - 1998 server	file se
22	10. Do you provide some method of data storage either locally or remote?	server, local hd, flpy, zip	server, local hd, floppy	server, floppy	server, local hd, floppy	local hd, floppy	floppy
23	11. What Operating System do you use in your computer laboratories?	Win 3.x, 95, OS 7.x	Win 3.x,95,NT3.x,OS7.x	Win 3.x, NT4,OS 7.x,8	Win 3.x,95, OS 7.x,8, Unix	Win 95, NT4, OS 8	Win 3
24	12. What Network Operating Systems do you use in your computer laboratories?	Novell 4.x	Novell 3.x,4.x, Unix	Novell 3.x	LanMgr, Novel 3.x,4.x, Unix	Lanmgr, Unix	Lanm
25	13. How many laser printers do you provide within your computer laboratories?	less than 4	4-8	4-8	8-12 (in all labs)	4-8	20 or
26	14. Do you charge for any of your printing facilities and if so what charging mechanism do	Yes - hetper card systen	Yes - photocopy card	Yes - Stored value cards	Yes - quotas/spooling	Yes - billing, 1998 - unica	No
27	15. If you charge for your printing facilities who pays?	Student	Student	Students & Faculty	Studnets/Faculty/IT Div	Student	
28	16. Do you provide access to the Internet from your computer laboratories?	Yes	Yes & No (policy)	Yes	Yes	Yes	Yes
29	17. Do you charge for access to the internet and if so what charging mechanisms do you us	Yes - hetper card systen	No	No (in 1998)	Yes - IP accounting	No	Yes -
30	18. If you charge for Internet access who pays?	Student		Students & Faculty	Students/IT Div.		Facul
31	19. Do you provide network points/empty spaces for students to use laptops and if so, how	Yes - less than 30	No	No (in 1998)	No - (special note below)	No	less t
32	20. What new workstation technologies are you looking at implementing in the near future	None	NC,NetPC,thin cl	Smartcards & NC	NC, thin client, smartcrds	NC, NetPC, thin client	None
33							
34							
35	Notes:						
36	ADFA - all students reside on campus and have network points in their rooms - 70% have o						
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							

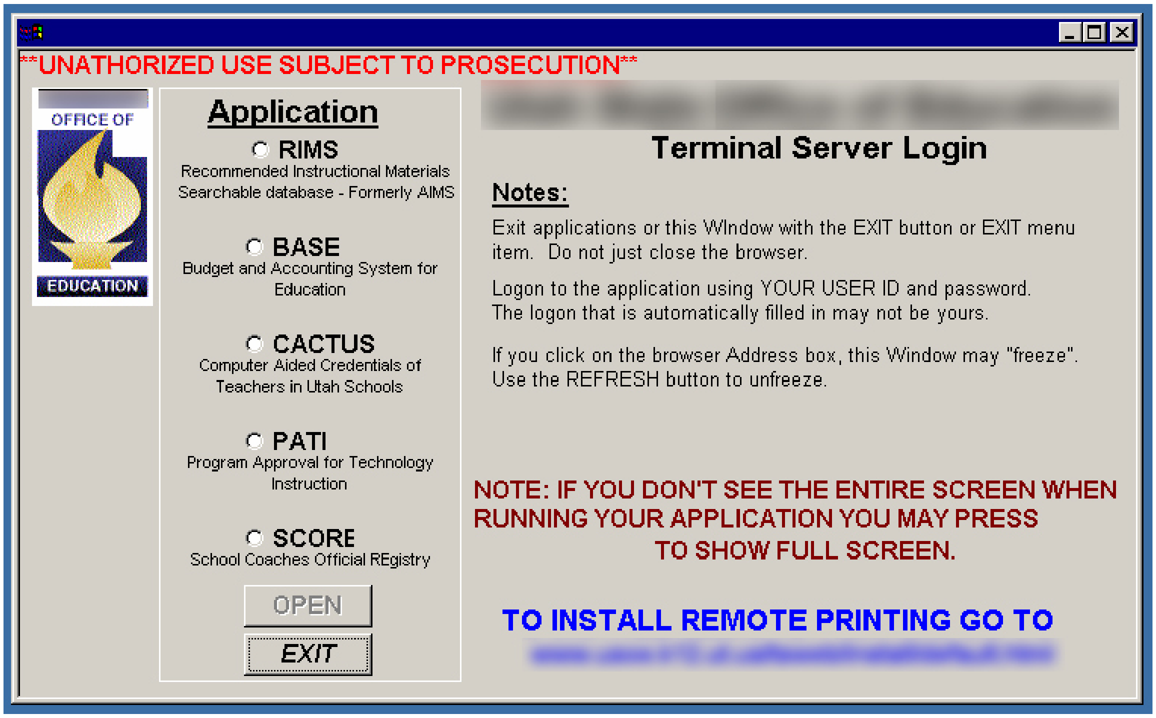
As you can see from the screen shot, one of the results is a spreadsheet of a survey that contained a question asking whether a username and password was required to login to laboratory workstations. This is great information for the hacker that wants to set up his network of zombies for his next mass denial of service attack.

Again, Foundstone has found Excel spreadsheets containing username and password lists, school grades and even medical records. In one case we found a Microsoft PowerPoint presentation with wiring diagrams and GPS location information to a very sensitive piece of the national critical infrastructure.

Other favorites include filetype:log [*inurl: password.log*] and an all time classic [*filetype:htpasswd htpasswd*].

Remote Administration - intitle:"Terminal Services Web Connection"

Many web servers, and associated technology like Microsoft's Terminal Services, create a default Title on the login page. Google offers the ability to restrict the search to information in the title and so by issuing the query [*intitle:"Terminal Services Web Connection"*] we are able to find URL's that offer logins to companies' terminal services.



UNATHORIZED USE SUBJECT TO PROSECUTION

Application

○ RIMS
Recommended Instructional Materials Searchable database - Formerly AIMS

○ BASE
Budget and Accounting System for Education

○ CACTUS
Computer Aided Credentials of Teachers in Utah Schools

○ PATI
Program Approval for Technology Instruction

○ SCORE
School Coaches Official REgistry

[redacted]

Terminal Server Login

Notes:

Exit applications or this WIndow with the EXIT button or EXIT menu item. Do not just close the browser.

Logon to the application using YOUR USER ID and password.
The logon that is automatically filled in may not be yours.

If you click on the browser Address box, this Window may "freeze".
Use the REFRESH button to unfreeze.

NOTE: IF YOU DON'T SEE THE ENTIRE SCREEN WHEN RUNNING YOUR APPLICATION YOU MAY PRESS
TO SHOW FULL SCREEN.

TO INSTALL REMOTE PRINTING GO TO
[redacted]

As you can see from the screenshot below the query resulted in a direct login to a Microsoft terminal services server at an educational establishment. It even has a handy set of applications that you can login to for budgeting, accounting, and instructional materials.

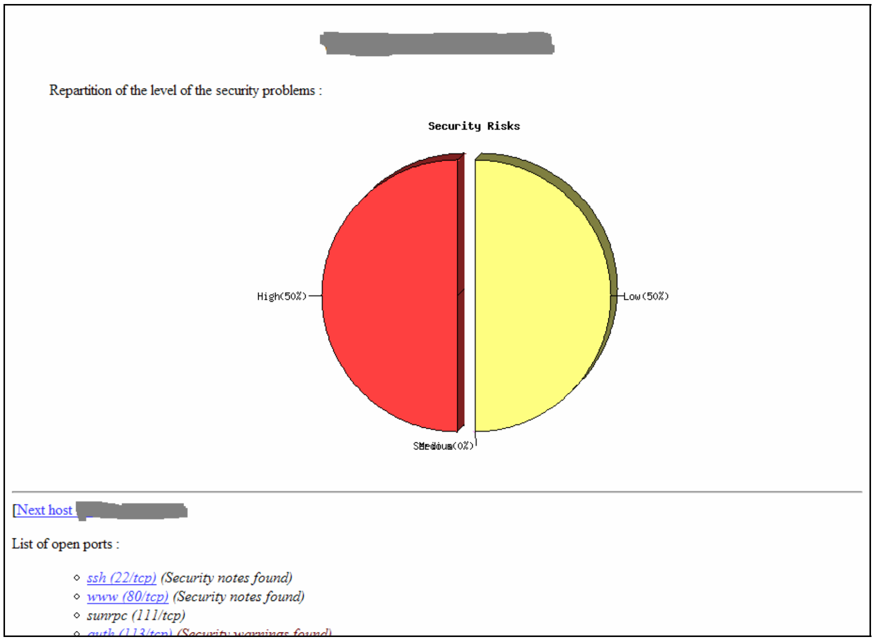
This is a good example of a remote administration interface that is all too commonly placed on the Internet today. Other examples include web server administration interfaces (often left open with blank passwords), PBX telephony systems, and application configuration management interfaces.

This same technique can be used to find default installations of Microsoft's IIS 4.0 web server, Cisco devices with the web interface enabled and ColdFusion admin tools.

[*intitle:"Welcome to IIS 4.0"*]
[*inurl:tech-support inurl:show Cisco*]
[*intitle:"ColdFusion Administrator Login"*]

Security Reports - "This file was generated by Nessus"

Google indexes everything including security reports. Using search engines, hackers are often able to find vulnerability scanning reports and intrusion detection alerts or logs. Using the query [*"This file was generated by Nessus"*] we are able to find vulnerability scanning reports, some with details of significant remote root compromises from the Internet to the internal corporate network. The screen shot below highlights one such report.



[redacted]

Repartition of the level of the security problems :

[Next host [redacted]

List of open ports :

- *ssh (22/tcp) (Security notes found)*
- *www (80/tcp) (Security notes found)*
- *sunrpc (111/tcp)*
- *auth (113/tcp) (Security warnings found)*

Potential SQL Injection – [ORA-00921: unexpected end of SQL command]

SQL Injection has become one of the most powerful application security attacks, where an attacker is able to run database queries of his/her choosing on the web sites backend data store. Not only can the attacker often read the entire database, but in some circumstances he/she can overwrite data such as prices and purchase items at a discount.

One sign where SQL injection is often a potential problem is database errors on web pages. Google allows you to construct a query looking for these. Foundstone has found numerous SQL injection points to online banks, ecommerce and password databases using this technique.

[*ORA-00921: unexpected end of SQL command*]

Automated Scanning – SiteDigger 2.0™

As you can see from the examples in the previous sections, the potential for malicious users to use search engines to find really bad stuff is very real indeed.

To that effect Foundstone Professional Services consultants decided to write a tool that automates the process of profiling Web sites using the Google search engine so that our customers can ensure that they do not have any vulnerabilities available through public search engines.

Google provides a web services interface (API) that allows users to create queries and send then to Google as web service requests. This programmatic method is ideal for generating a series of requests and analyzing the results. Using the Microsoft .NET Framework and the C# language, the developers Kartik Trivedi and Eric Heitzman created a Windows desktop application to automate the process called SiteDigger 2.0.

Installing SiteDigger 2.0

Step One

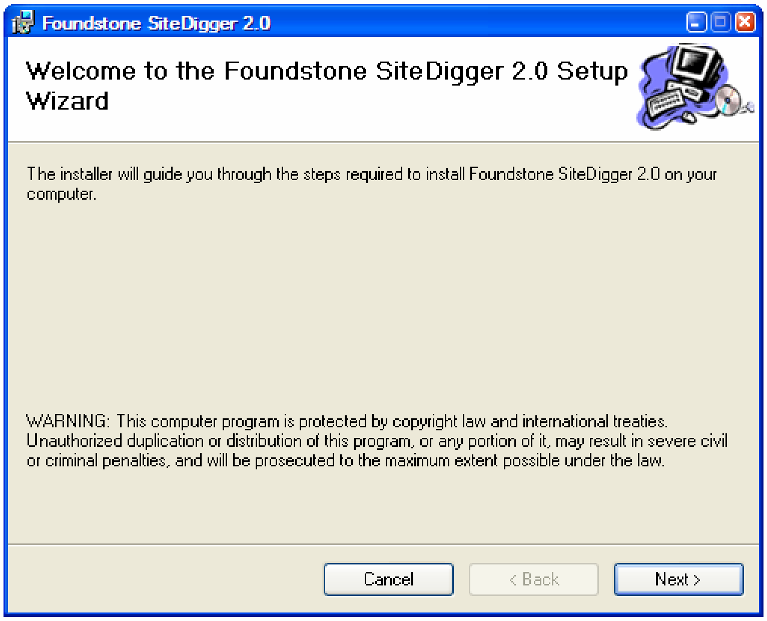
You can download SiteDigger 2.0 from the Foundstone web site. Navigate your web browser to http://www.foundstone.com/s3i and the SiteDigger 2.0 download can be found in the left hand box.

Before running the installation program setup.exe you must have the .NET Framework Version 1.1 or later installed on a Windows XP PC. The application should work on other variants of Windows although has not been tested. You can download the .NET framework here:

http://msdn.microsoft.com/netframework/howtoget/default.aspx

When you double click the installation file for SiteDigger 2.0, you will be presented with an installation splash screen like the one on the following page.

Step Two

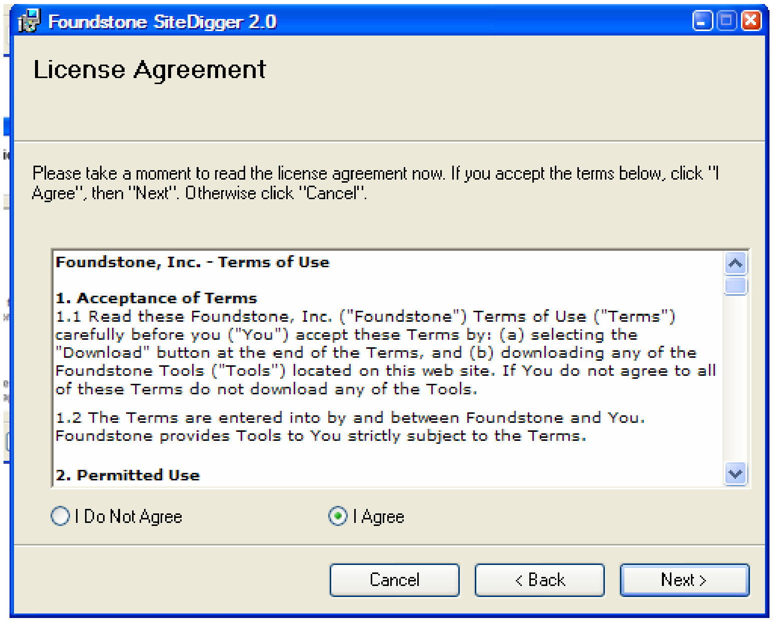


Welcome to the Foundstone SiteDigger 2.0 Setup Wizard

The installer will guide you through the steps required to install Foundstone SiteDigger 2.0 on your computer.

WARNING: This computer program is protected by copyright law and international treaties. Unauthorized duplication or distribution of this program, or any portion of it, may result in severe civil or criminal penalties, and will be prosecuted to the maximum extent possible under the law.

You will need to accept the Foundstone, Inc. Terms of Use. Click next and you will be presented with the following screen.



License Agreement

Please take a moment to read the license agreement now. If you accept the terms below, click "I Agree", then "Next". Otherwise click "Cancel".

Foundstone, Inc. - Terms of Use

1. Acceptance of Terms
1.1 Read these Foundstone, Inc. ("Foundstone") Terms of Use ("Terms") carefully before you ("You") accept these Terms by: (a) selecting the "Download" button at the end of the Terms, and (b) downloading any of the Foundstone Tools ("Tools") located on this web site. If You do not agree to all of these Terms do not download any of the Tools.

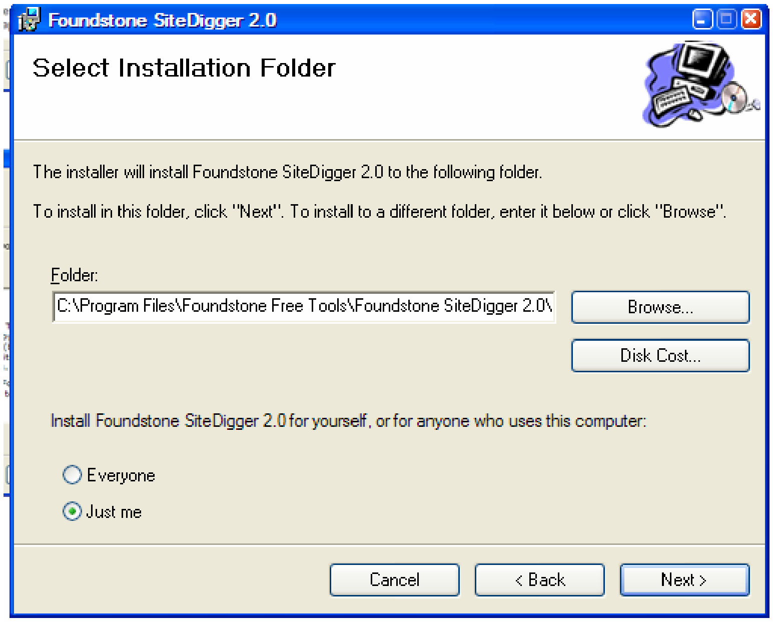
1.2 The Terms are entered into by and between Foundstone and You. Foundstone provides Tools to You strictly subject to the Terms.

2. Permitted Use

○ I Do Not Agree ● I Agree

Step Three

Select the Folder where you would like to install SiteDigger 2.0. The default location is C:\Program Files\Foundstone\SiteDigger 2.0\.



Select Installation Folder

The installer will install Foundstone SiteDigger 2.0 to the following folder.

To install in this folder, click "Next". To install to a different folder, enter it below or click "Browse".

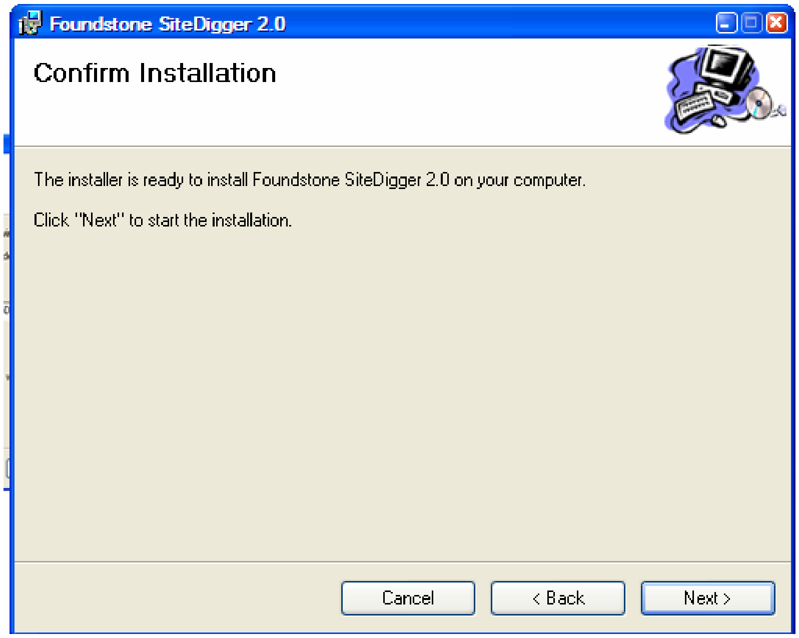
Folder: C:\Program Files\Foundstone Free Tools\Foundstone SiteDigger 2.0\

Install Foundstone SiteDigger 2.0 for yourself, or for anyone who uses this computer:

○ Everyone
● Just me

Step Four

The installation progress bar should take less than 10 seconds to complete.



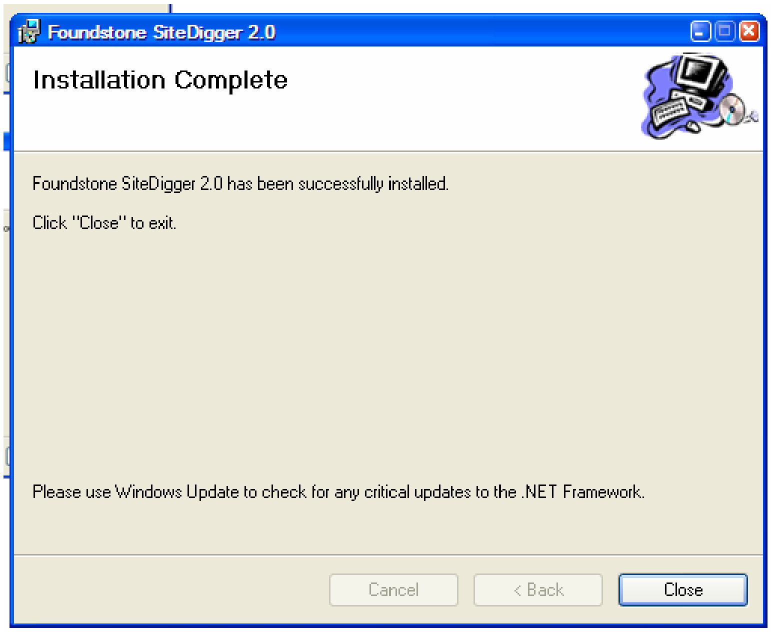
Confirm Installation

The installer is ready to install Foundstone SiteDigger 2.0 on your computer.

Click "Next" to start the installation.

Step Five

SiteDigger 2.0 requires Google a web services API license key. Instructions on how to obtain your license key from Google is discussed in the next section of this paper.



Foundstone SiteDigger 2.0

Installation Complete

Foundstone SiteDigger 2.0 has been successfully installed.

Click "Close" to exit.

Please use Windows Update to check for any critical updates to the .NET Framework.

Click Next and if no errors occurred during the installation process, you will be presented with an installation complete splash screen.

What to do if the Installer Fails

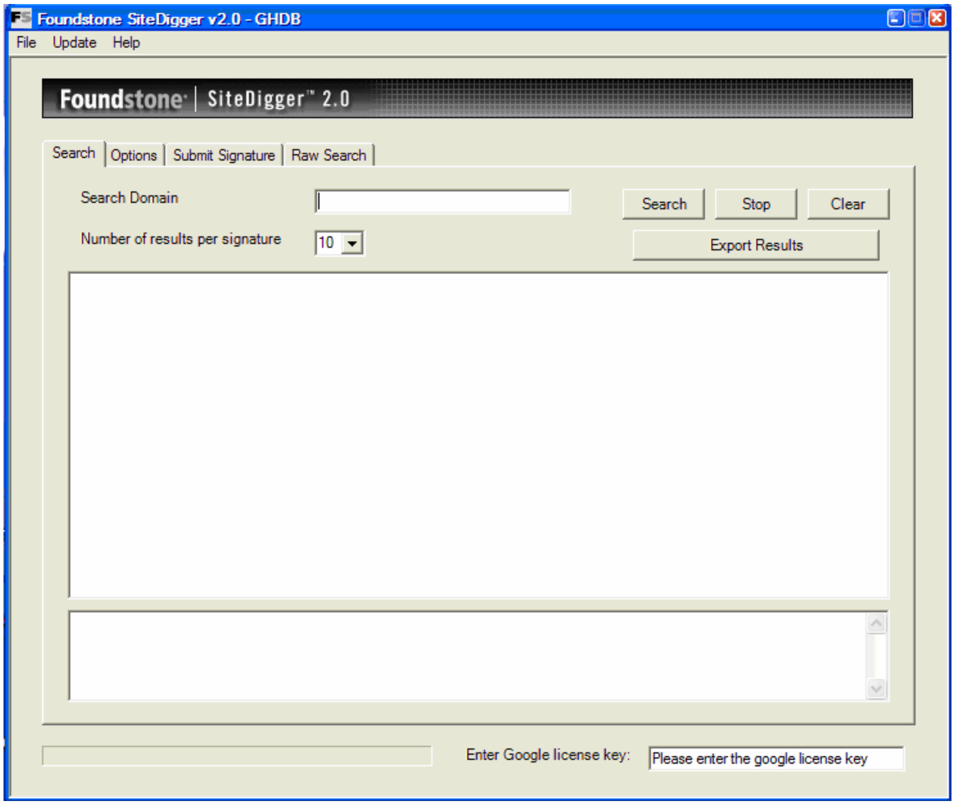
The installer checks to ensure you have the .NET framework installed and that you meet the minimum requirements for installation. However, if installation fails for any reason we suggest running windows Update (http://windowsupdate.microsoft.com) and reinstalling the latest .NET framework.

Using SiteDigger 2.0™

The SiteDigger 2.0 User Interface

To open SiteDigger 2.0™ navigate to the Start Menu > Programs > Foundstone Free Tools. Look for other free tools from Foundstone in the coming months that will also be found under the Foundstone Free Tools menu.

SiteDigger 2.0 is a windows GUI tool. All navigation and options are set from the GUI. There are four main tabs and a drop down menu at the top of the interface.



Search Domain

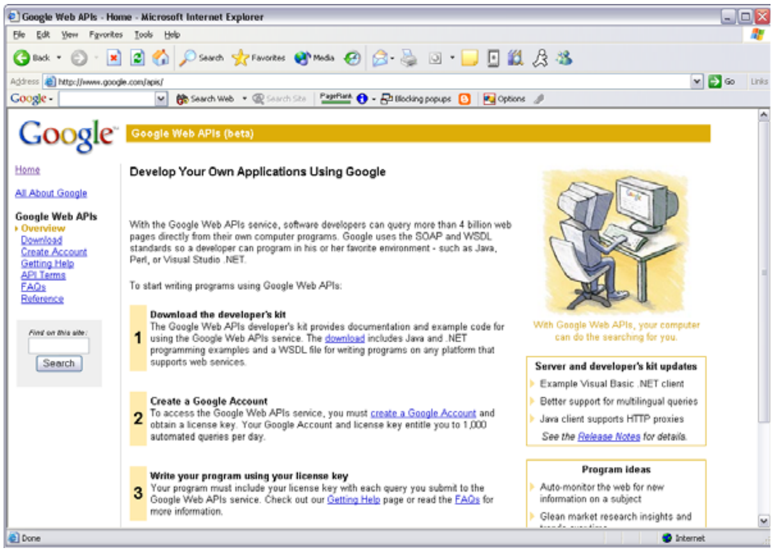
Number of results per signature: 10

Enter Google license key: Please enter the google license key

Getting a Google License Key

Before using SiteDigger 2.0 you must obtain your own Google web services API from Google by navigating to http://www.google.com/apis/ and following the instructions. You will be required to accept the Google API terms and conditions and have a valid email address to generate a key. Google limits this key to 1,000 search queries per day.

When you have obtained your key you should copy and paste the license string into the bottom left hand corner of the SiteDigger 2.0 GUI.


Google Web APIs - Home - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://www.google.com/apis/
Go Links
Google - Search Web Search Site PageRank Blocking popups Options
Google
Google Web APIs (beta)
Home
All About Google
Google Web APIs
Overview
Download
Create Account
Getting Help
API Terms
FAQs
Reference
Find on this site:
Search
Done
Internet

Develop Your Own Applications Using Google

With the Google Web APIs service, software developers can query more than 4 billion web pages directly from their own computer programs. Google uses the SOAP and WSDL standards so a developer can program in his or her favorite environment - such as Java, Perl, or Visual Studio .NET.

To start writing programs using Google Web APIs:

1 **Download the developer's kit**
The Google Web APIs developer's kit provides documentation and example code for using the Google Web APIs service. The download includes Java and .NET programming examples and a WSDL file for writing programs on any platform that supports web services.

2 **Create a Google Account**
To access the Google Web APIs service, you must create a Google Account and obtain a license key. Your Google Account and license key entitle you to 1,000 automated queries per day.

3 **Write your program using your license key**
Your program must include your license key with each query you submit to the Google Web APIs service. Check out our Getting Help page or read the FAQs for more information.

With Google Web APIs, your computer can do the searching for you.

Server and developer's kit updates
- Example Visual Basic .NET client
- Better support for multilingual queries
- Java client supports HTTP proxies

See the Release Notes for details.

Program ideas
- Auto-monitor the web for new information on a subject
- Glean market research insights and

Updating Signatures

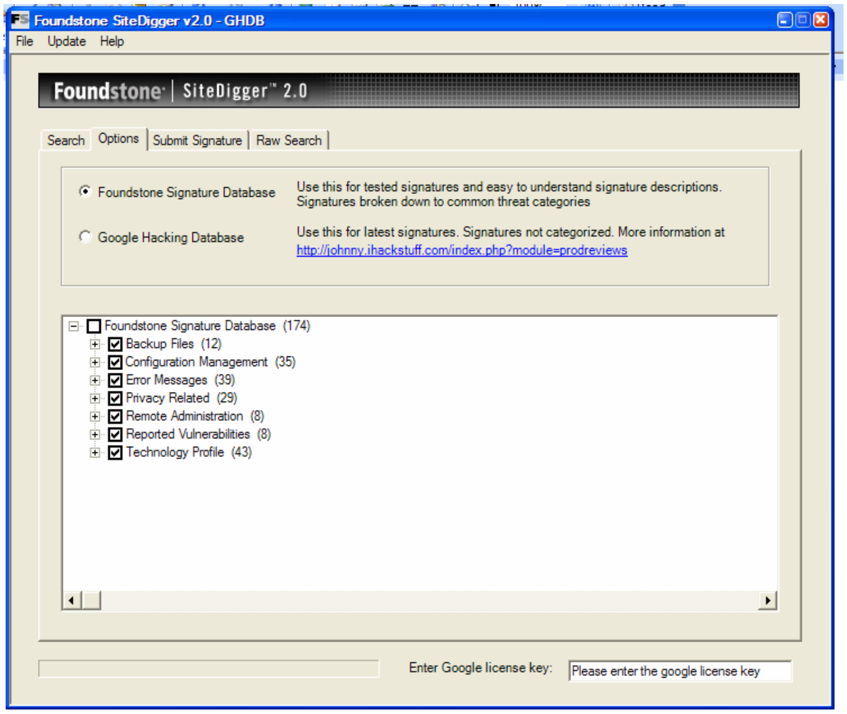
Before running SiteDigger 2.0 for the first time, and periodically afterwards, you should run the Update Signatures option from the Options drop down menu. SiteDigger 2.0 has two sets of Signature Databases; Foundstone Signature Database and Google Hacking Database. The Foundstone Signature Database is includes signatures that have been tested by Foundstone, have Foundstone signature descriptions, and are categorized by threat category. Google Hacking Database includes the latest signatures that have been submitted to the Google Hacking Database hosted at the http://johnny.ihackstuff.com website. These signatures have not been tested and verified by Foundstone and include signature descriptions that are a bit more humorous.

To update either database, select your desired database from the Update menu, this option connects to the appropriate Web site and determines whether you have the latest signature file. If you do not it will automatically download and install the latest signatures.

Configuring Scan Options

You will need to configure both a target site and the signatures that you wish to execute. To configure the target site simply enter the domain into the text field that says “Search Domain”. Select the number of results per signature that you desire.

To select the signatures you wish to execute you navigate to the Options tab and check or uncheck the desired signatures.


Foundstone SiteDigger v2.0 - GHDB
File
Update
Help
Foundstone | SiteDigger™ 2.0
Search
Options
Submit Signature
Raw Search

● Foundstone Signature Database — Use this for tested signatures and easy to understand signature descriptions. Signatures broken down to common threat categories

○ Google Hacking Database — Use this for latest signatures. Signatures not categorized. More information at http://johnny.ihackstuff.com/index.php?module=prodreviews

- □ Foundstone Signature Database (174)
    - ☑ Backup Files (12)
    - ☑ Configuration Management (35)
    - ☑ Error Messages (39)
    - ☑ Privacy Related (29)
    - ☑ Remote Administration (8)
    - ☑ Reported Vulnerabilities (8)
    - ☑ Technology Profile (43)

Enter Google license key: Please enter the google license key

For the Foundstone Signature Database, you can select the entire set of signatures by checking the top level “All” checkbox or select signatures by category or individually. Signatures are divided into core groups:

- Back Up Files
- Configuration Management
- Error Messages
- Privacy Related
- Remote Administration Interfaces
- Reported Vulnerabilities
- Technology Profile

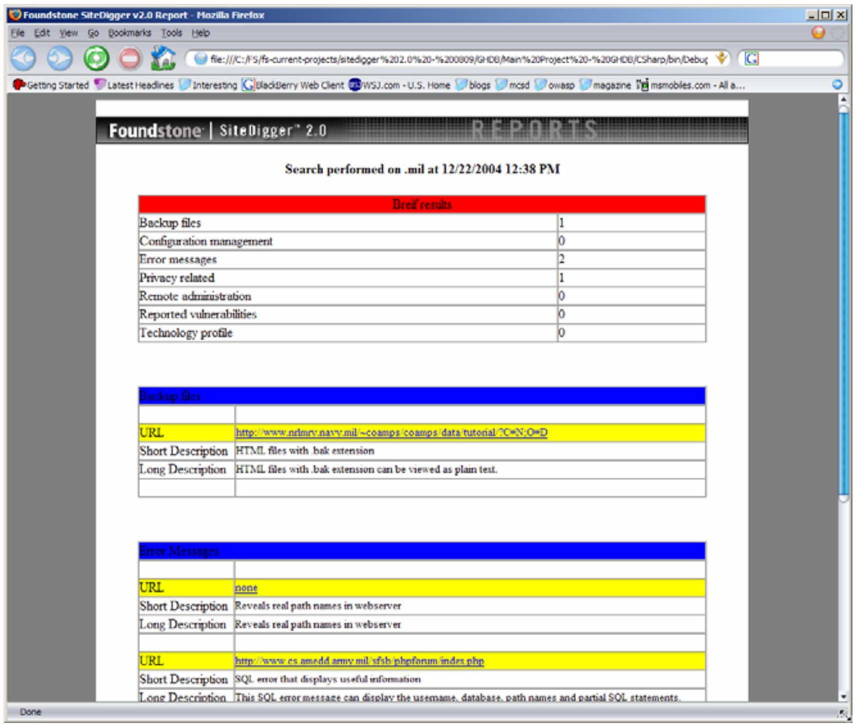
For the Google Hacking Database, since they are not categorized, you must select all of the signatures or individually select the signatures you wish to test.

Launching the Scan

After you have installed your key, updated and configured your signatures and entered your target site, you are ready to scan. Simply press the Search button.

Analyzing Results

After SiteDigger 2.0 has completed its scan you can generate a report using the Export button. This will create an HTML report and open it in your default web browser. If this does not open automatically look for the results.html file in the Program Files\Foundstone\SiteDigger 2.0\output\ directory.


Foundstone SiteDigger v2.0 Report - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
file:///C:/FS/fs-current-projects/sitedigger%202.0%20-%200809/GHDB/Main%20Project%20-%20GHDB/CSharp/bin/Debuç
Getting Started Latest Headlines Interesting BlackBerry Web Client WSJ.com - U.S. Home blogs mcsd owasp magazine msmobiles.com - All a...
Foundstone | SiteDigger™ 2.0 REPORTS
Done

Search performed on .mil at 12/22/2004 12:38 PM

Breif results	
Backup files	1
Configuration management	0
Error messages	2
Privacy related	1
Remote administration	0
Reported vulnerabilities	0
Technology profile	0

Backup files	
URL	http://www.nrlmry.navy.mil/~coamps/coamps/data/tutorial/?C=N;O=D
Short Description	HTML files with .bak extension
Long Description	HTML files with .bak extension can be viewed as plain text.

Error Messages	
URL	none
Short Description	Reveals real path names in webserver
Long Description	Reveals real path names in webserver
URL	http://www.cs.amedd.army.mil/sfsb/phpforum/index.php
Short Description	SQL error that displays useful information
Long Description	This SQL error message can display the username, database, path names and partial SQL statements.

Each result is listed with the URL of the vulnerability, a summary of the search query and a short description of the issue. Google hacking is especially prone to false positive and false negatives and so each issue needs to be manually validated by clicking on the URL in question.

After taking appropriate remediation steps, Foundstone suggests that you run SiteDigger 2.0 on a regular basis to ensure that other information leakage issues have not surfaced.

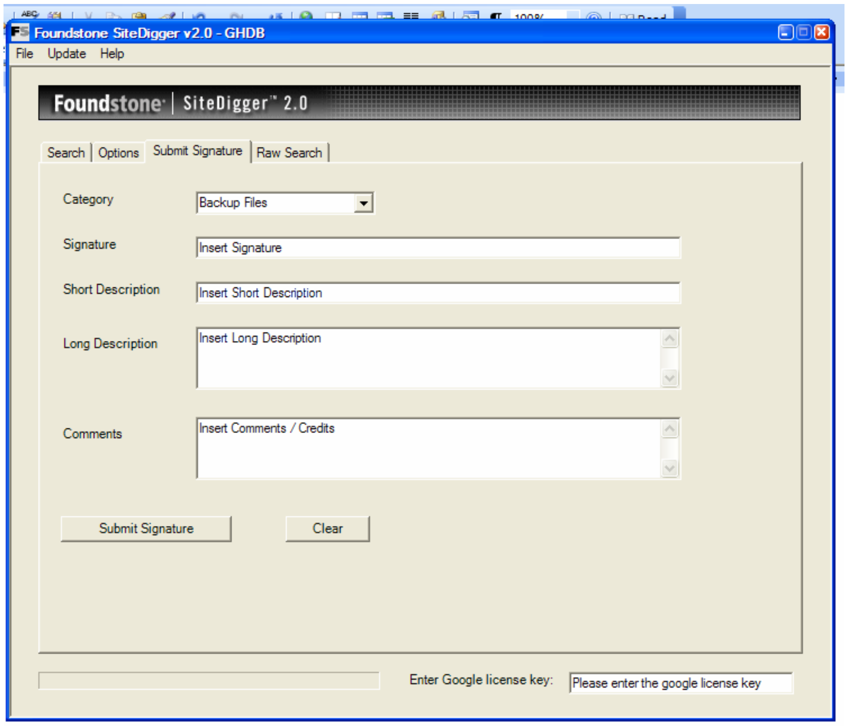
Writing Your Own Signatures

All of the signatures are stored in a simple XML file called schema.xml located at Drive Letter:\Program Files\Foundstone\SiteDigger 2.0\schema.xml.

You can write your own signatures to be used by SiteDigger 2.0 by simply adding a new entry to the main signature file in the format below. We strongly suggest cloning an existing entry and using a simple text editor such as Notepad that doesn’t add additional meta data.

```
<signature>
	<signatureReferenceNumber>Incremental Sig Number Goes 
Here</signatureReferenceNumber>
	<categoryref>Ref Goes Here</categoryref>
	<category>Category Goes Here</category>
	<querytype>Query Type Goes Here</querytype>
	<querystring>Search Query Goes Here</querystring>
	<shortDescription>Short Description Goes Here</shortDescription>
	<textualDescription>Textual Description Goes Here</textualDescription>
	<cveNumber>CVE Number Goes Here</cveNumber>
	<cveLocation>CVE URL Goes Here</cveLocation>
</signature>
```

If you would like to share your updated signatures, please submit your signature using the Submit Signature feature.



Category: Backup Files

Signature: Insert Signature

Short Description: Insert Short Description

Long Description: Insert Long Description

Comments: Insert Comments / Credits

Enter Google license key: Please enter the google license key

Performing Raw Searches

You can also perform your own raw searches within SiteDigger 2.0 using the syntax discussed earlier in this document. Simply click on the “Raw Search” tab and enter the signature in the Search Signature field.

Conclusion

Information leakage is serious problem but one that can be easily managed using SiteDigger 2.0. Identifying the disclosure of confidential information is now a less time-intensive process and one that can be added to the prioritized responsibilities of professionals charged with protecting their critical assets from malicious intruders.

References

Web Sites

An excellent site dedicated to Google Hacking: http://johnny.ihackstuff.com

Google Web Services API: www.google.com/apis/

Trademarks

SiteDigger 2.0 is a trademark of Foundstone Inc. All other trademarks are the property of their respective owner.

About Foundstone

Foundstone®, Inc., a division of McAfee, offers a unique combination of software, services, and education to help organizations continuously and measurably protect the most important assets from the most critical threats. Through a strategic approach to security, Foundstone identifies, recommends, and implements the right balance of technology, people, and process to manage digital risk and leverage security investments more effectively.