GOOGLE HACKING

Defiana Arnaldy, M.Si
deff_arnaldy@yahoo.com

Overview

- GOOGLE SEARCH TECHNIQUES
- GOOGLE ADVANCED OPERATORS
- GOOGLE HACKING TECHNIQUES
- ABOUT GOOGLE AUTOMATED SCANNING
- OTHER GOOGLE STUFF

GOOGLE SEARCH TECHNIQUES

- The Google search engine found at www.google.com offers many different features including language and document translation, web, image, newsgroups, catalog and news searches and more
- These features offer obvious benefits to even the most uninitiated web surfer, but these same features allow for far more nefarious possibilities to the most malicious Internet users including hackers, computer criminals, identity thieves and even terrorists

Google web interface

- The Google search engine is fantastically easy to use.
- Despite the simplicity, it is very important to have a firm grasp of these basic techniques in order to fully comprehend the more advanced uses.
- The most basic Google search can involve a single word entered into the search page found at www.google.com.

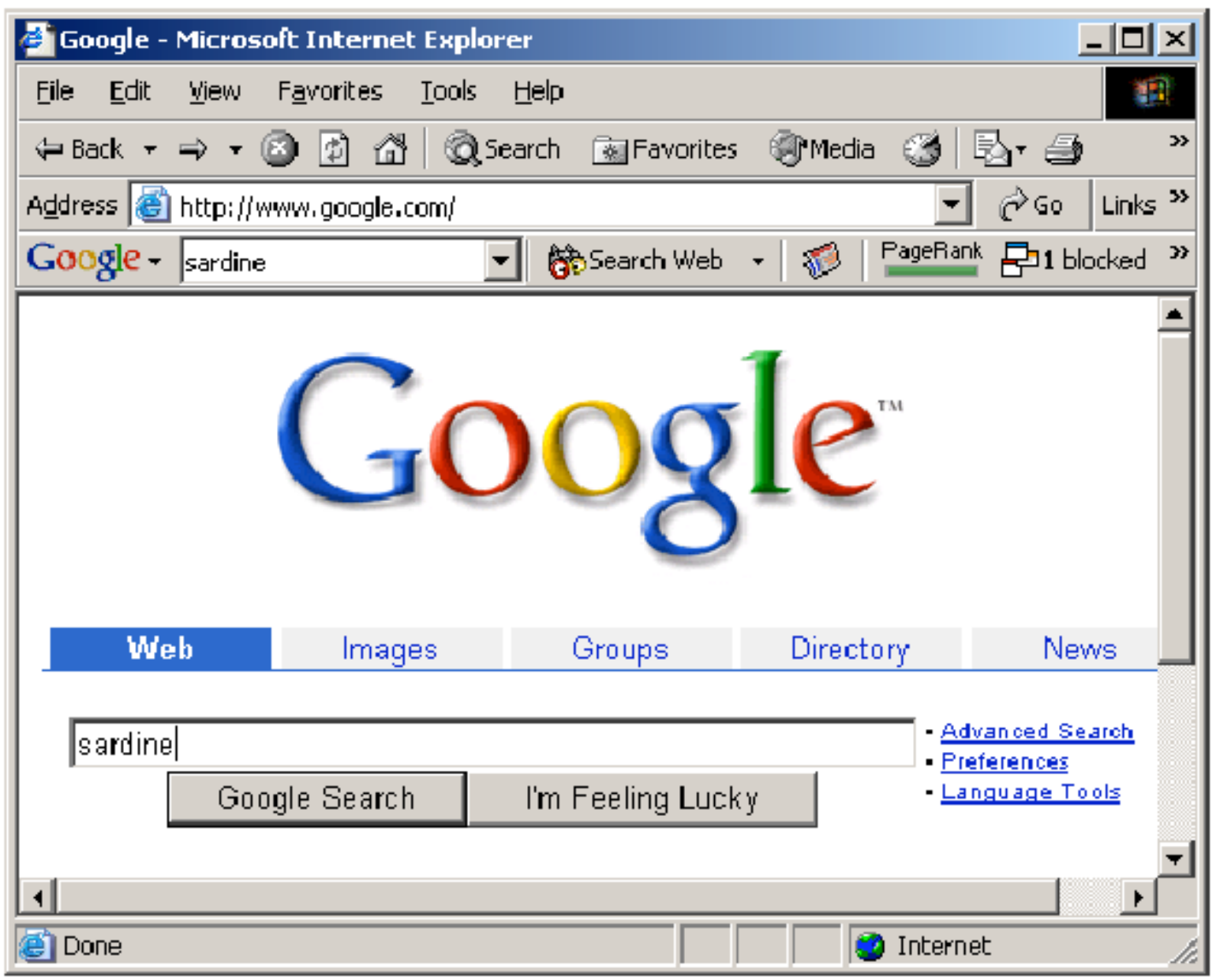

Figure 1: The main Google search page

- As shown in Figure 1. I have entered the word Sardine. into the search screen.
- Figure 1 shows many of the options available from the www.google.com front page.

The Google toolbar	The Internet Explorer browser I am using has a Google "toolbar" (a free download from toolbar.google.com) installed and presented under the address bar. Although the toolbar offers many different features, it is not a required element for performing advanced searches. Even the most advanced search functionality is available to any user able to access the www.google.com web page with any type of browser, including text-based and mobile browsers.

"Web, Images, Groups, Directory and News" tabs	These tabs allow you to search web pages, photographs, message group postings, Google directory listings, and news stories respectively. First-time Google users should consider that these tabs are not always a replacement for the "Submit Search" button.
Search term input field	Located directly below the alternate search tabs, this text field allows the user to enter a Google search term. Search term rules will be described later.
"Submit Search"	This button submits the search term supplied by the user. In many browsers, simply pressing the "Enter/Return" key after typing a search term will activate this button.
"I'm Feeling Lucky"	Instead of presenting a list of search results, this button will forward the user to the highest-ranked page for the entered search term. Often times, this page is the most relevant page for the entered search term.
"Advanced Search"	This link takes the user to the "Advanced Search" page as shown in Figure 2. Much of the advanced search functionality is accessible from this page. Some advanced features are not listed on this page.
"Preferences"	This link allows the user to select several options (which are stored in cookies on the user's machine for later retrieval) including languages, filters, number of results per page, and window options.
"Language tools"	This link allows the user to set many different language options and translate text to and from various languages.

Basic search techniques

- Simple word searches
  - Basic Google searches, as I have already presented, consist of one or more words entered without any quotations or the use of special keywords.
  - Examples:
    - peanut butter
    - butter peanut
    - olive oil popeye

- “+” searches
  - When supplying a list of search terms, Google automatically tries to find every word in the list of terms, making the Boolean operator “AND” redundant
  - Some search engines may use the plus sign as a way of signifying a Boolean “AND”
  - Google uses the plus sign in a different fashion. When Google receives a basic search request that contains a very common word like “the”, ”how”, or “where” the word will often times be removed from the query as shown in Figure 6

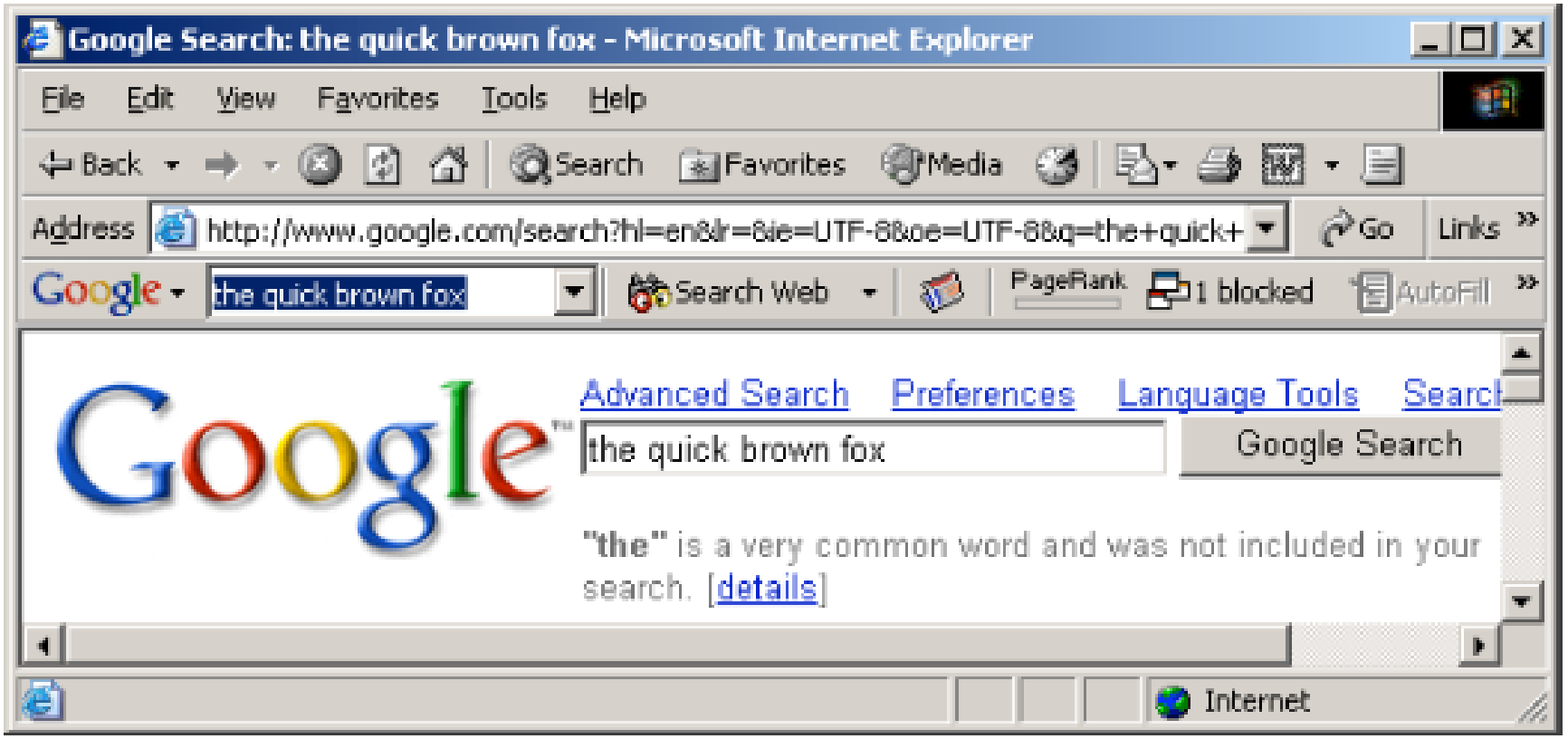

Figure 6: Google removing overly common words

- In order to force Google to include a common word, precede the search term with a plus (+) sign.
- Do not use a space between the plus sign and the search term.
- For example, the following searches produce slightly different results:
  - where quick brown fox
  - +where quick brown fox

- “-” searches
  - Excluding a term from a search query is as simple as placing a minus sign (-) before the term.
  - Do not use a space between the minus sign and the search term.
  - For example, the following searches produce slightly different results:
    - quick brown fox
    - quick –brown fox

- Phrase Searches
  - In order to search for a phrase, supply the phrase surrounded by double-quotes.
  - Examples:
    - “the quick brown fox”
    - “liberty and justice for all”
    - “harry met sally”

- Mixed searches
  - Mixed searches can involve both phrases and individual terms.
  - Example:
    - macintosh "microsoft office“
  - This search will only return results that include the phrase “Microsoft office “, and the term macintosh.

Google advanced operators

- Google allows the use of certain operators to help refine searches.
- The use of advanced operators is very simple as long as attention is given to the syntax.
- The basic format is:
  - operator:search_term
- Notice that there is no space between the operator, the colon and the search term.
- If a space is used after a colon, Google will display an error message.
- If a space is used before the colon, Google will use your intended operator as a search term

- Some advanced operators can be used as a standalone query.
- For example
- ‘cache:www.google.com’ can be submitted to Google as a valid search query
- The ‘site’ operator, by contrast, must be used along with a search term, such as ‘site:www.google.com help’

- **Table 1: Advanced Operator Summary**

Operator	Description	Additional search argument required?
site:	find search term only on site specified by search_term.	YES
filetype:	search documents of type search_term	YES
link:	find sites containing search_term as a link	NO
cache:	display the cached version of page specified by search_term	NO
intitle:	find sites containing search_term in the title of a page	NO
inurl:	find sites containing search_term in the URL of the page	NO

- site: find web pages on a specific web site
  - This advanced operator instructs Google to restrict a search to a specific web site or domain.
  - When using this operator, an addition search argument is required
  - Example

```
site:harvard.edu tuition
```

  - This query will return results from harvard.edu that include the term tuition anywhere on the page

- **filetype: search only within files of a specific type.**
  - This operator instructs Google to search only within the text of a particular type of file.
  - This operator requires an additional search argument.
  - Example:

```
filetype:txt endometriosis
```

  - This query searches for the word “endometriosis” within standard text documents.
  - There should be no period (.) before the filetype and no space around the colon following the word “filetype”

- It is important to note that Google only claims to be able to search *within certain types of files*
- Google can search within most files that present as plain text
- For example, Google can easily find a word within a file of type “txt”, “html” or “php” since the output of these files in a typical web browser window is textual
- The current list of files that Google can search is listed in the filetype FAQ located at http://www.google.com/help/faq_filetypes.html.

- Google can search within the following file types:
  - Adobe Portable Document Format (pdf)
  - Adobe PostScript (ps)
  - Lotus 1-2-3 (wk1, wk2, wk3, wk4, wk5, wki, wks, wku)
  - Lotus WordPro (lwp)
  - MacWrite (mw)
  - Microsoft Excel (xls)
  - Microsoft PowerPoint (ppt)
  - Microsoft Word (doc)
  - Microsoft Works (wks, wps, wdb)
  - Microsoft Write (wri)
  - Rich Text Format (rtf)
  - Text (ans, txt)

- **link: search within links**
  - The hyperlink is one of the cornerstones of the Internet.
  - A hyperlink is a selectable connection from one web page to another.
  - Most often, these links appear as underlined text but they can appear as images, video or any other type of multimedia content.

- This advanced operator instructs Google to search within hyperlinks for a search term.
- This operator requires no other search arguments.
- Example

```
link:www.apple.com
```

- This query query would display web pages that link to Apple.com's main page.
- This special operator is somewhat limited in that the link must appear exactly as entered in the search query.
- The above query would not find pages that link to www.apple.com/ipod, for example.

- **cache: display Google's cached version of a page**
  - This operator displays the version of a web page as it appeared when Google crawled the site.
  - This operator requires no other search arguments.
  - Example:

```
cache:johnny.ihackstuff.com
cache:http://johnny.ihackstuff.com
```

  - These queries would display the cached version of Johnny's web page.
  - Note that both of these queries return the same result

- **intitle: search within the title of a document**
  - This operator instructs Google to search for a term within the title of a document.
  - Most web browsers display the title of a document on the top title bar of the browser window.
  - This operator requires no other search arguments.
  - Example:

```
intitle:gandalf
```

  - This query would only display pages that contained the word ‘gandalf’ in the title

- A derivative of this operator, ‘allintitle’ works in a similar fashion.
- Example:

```
allintitle:gandalf silmarillion
```

- This query finds both the words ‘gandalf’ and ‘silmarillion’ in the title of a page.
- The ‘allintitle’ operator instructs Google to find *every subsequent word in the query only in the* title of the page.
- This is equivalent to a string of individual ‘intitle’ searches.

- **inurl: search within the URL of a page**
  - This operator instructs Google to search only within the URL, or web address of a document.
  - This operator requires no other search arguments.
  - Example:

```
inurl:amidala
```

  - This query would display pages with the word ‘amidala’ inside the web addres. One returned result, ‘http://www.yarwood.org/kell/amidala/’ contains the word ‘amidala’ as the name of a directory

- The word can appear anywhere within the web address, including the name of the site or the name of a file.
- A derivative of this operator, 'allinurl' works in a similar fashion.
- Example:

```
allinurl:amidala gallery
```

- This query finds both the words 'amidala' and 'gallery' in the URL of a page.
- The 'allinurl' operator instructs Google to find *every subsequent word in the query only in the URL of* the page.
- This is equivalent to a string of individual 'inurl' searches.

About Google's URL syntax

- The advanced Google user often times streamlines the search process by use of the Google toolbar or through direct use of Google URL's.
- For example, consider the URL generated by the web search for sardine:

```
http://www.google.com/search?hl=en&ie=UTF-8&oe=UTF-8&q=sardine
```

- First, notice that the base URL for a Google search is
- The question mark denotes the end of the URL and the beginning of the arguments "`http://www.google.com/search`".
- The '&' symbol separates arguments
- The URL presented to the user may vary depending on many factors including whether or not the search was submitted via the toolbar, the native language of the user, etc.

- Arguments to the Google search program are well documented at http://www.google.com/apis.
- The arguments found in the above URL are as follows:

`hl`:	Native language results, in this case “en” or English.
`ie`:	Input encoding, the format of incoming data. In this case “UTF-8”.
`oe`:	Output encoding, the format of outgoing data. In this case “UTF-8”.
`q`:	Query. The search query submitted by the user. In this case “sardine”.

- Most of the arguments in this URL can be omitted, making the URL much more concise.
- For example, the above URL can be shortened to making the URL much more concise

```
http://www.google.com/search?q=sardine
```

- Additional search terms can be appended to the URL with the plus sign.
- For example, to search for ‘sardine’ along with ’peanut’ and ‘butter’ consider using this URL:

```
http://www.google.com/search?q=sardine+peanut+butter
```

- Since simplified Google URLs are simple to read and portable, they are often used as a way to represent a Google search.

- Google (and many other web-based programs) must represent special characters like quotation marks in a URL with a hexadecimal number preceded by a percent (%) sign in order to follow the http URL standard.
- For example, a search for “the quick brown fox” (paying special attention to the quotation marks) is represented as

```
http://www.google.com/search?&q=%22the+quick+brown+fox%22
```

- In that example, a double quote is displayed as ’22’ and spaces are replaced by plus (+) signs.
- Google does not exclude overly common words from phrase searches.
- Overly common words are automatically included when enclosed in double-quotes.

Google hacking techniques

Domain searches using the ‘site’ operator

- The site operator can be expanded to search out entire domains.
- For example:

```
site:gov secret
```

- This query searches every web site in the .gov domain for the word ‘secret’

- Notice that the site operator works on addresses in reverse.
- For example, Google expects the site operator to be used like this:

```
site:www.cia.gov
site:cia.gov
site:gov
```

- Google would not necessarily expect the site operator to be used like this:

```
site:www.cia
site:www
site:cia
```

- The reason for this is simple. 'Cia' and 'www' are not valid top-level domain names.
- This means that as of this writing, Internet names may not *end in 'cia' or ' www'.*
- However sending unexpected queries like these are part of a competent Google hacker's arsenal as we explore in the 'googleturds' section.

How this technique can be used

1. Journalists, snoops and busybodies in general can use this technique to find interesting 'dirt' about a group of websites owned by organizations such as a government or non-profit organization. Remember that top-level domain names are often very descriptive and can include interesting groups such as: the U.S. Government (.gov or .us)
2. Hackers searching for targets. If a hacker harbors a grudge against a specific country or organization, he can use this type of search to find sensitive targets.

Finding ‘googleturds’ using the ‘site’ operator

- Googleturds, are little dirty pieces of Google ‘waste’
- These search results seem to have stemmed from typos Google found while crawling a web page.
- Example:

```
site:csc
site:microsoft
```

- Neither of these queries are valid according to the loose rules of the ‘site’ operator, since they do not end in valid top-level domain names.
- However, these queries produce interesting results as shown in Figure 7.

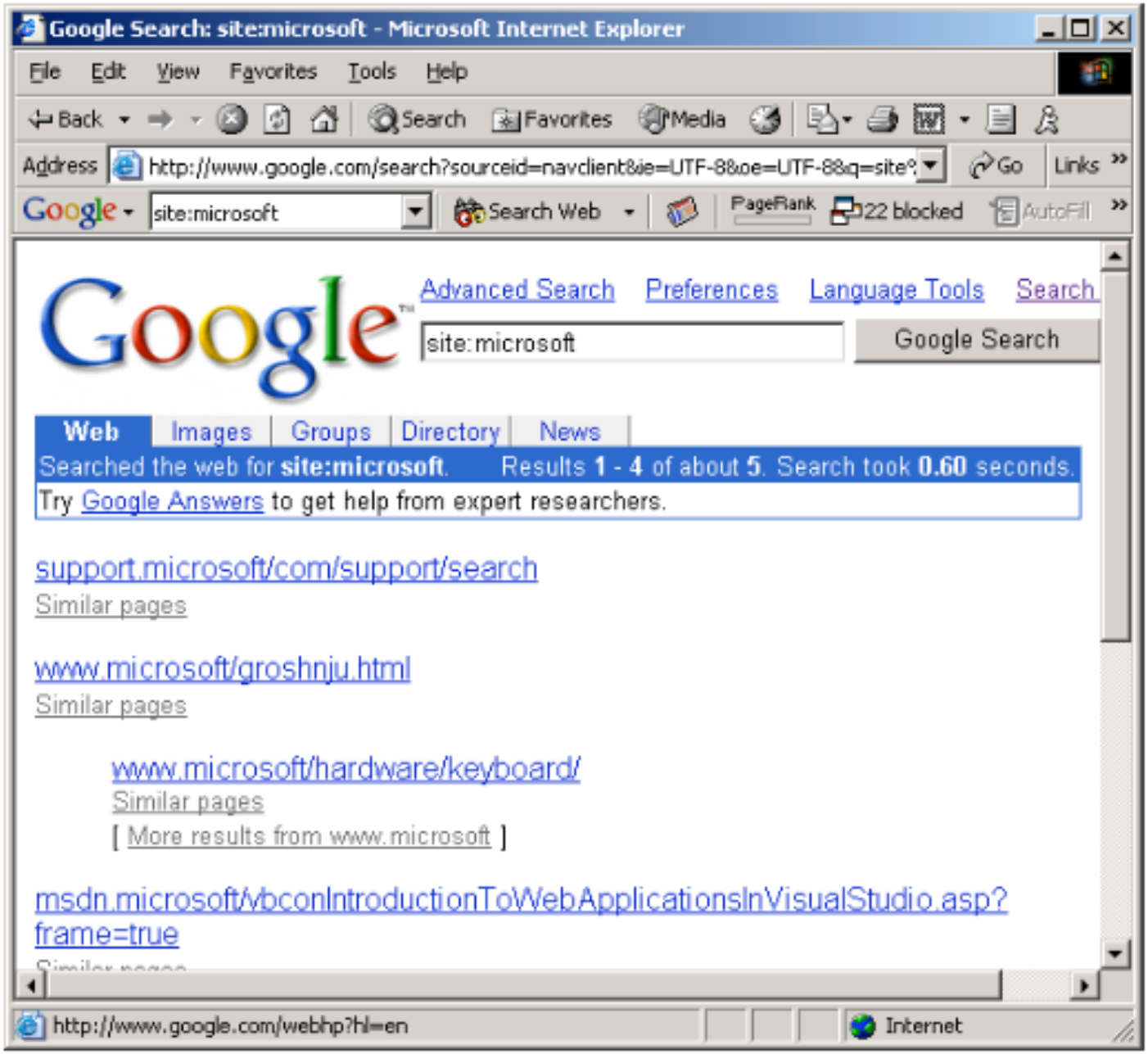

Figure 7: Googleturd example

How this technique can be used

- Hackers investigating a target can use munged site values based on the target's name to dig up Google pages (and subsequently potential sensitive data) that may not be available to Google searches using the valid 'site' operator
- Example:
  - A hacker is interested in sensitive information about ABCD Corporation, located on the web at www.ABCD.com.
  - Using a query like "site:ABCD" may find mistyped links (http://www.abcd instead of http://www.abcd.com) containing interesting information.

Site mapping: More about the 'site' operator

- Mapping the contents of a web server via Google is simple.
- Consider the following query:

```
site:www.microsoft.com microsoft
```

- This query searches for the word 'Microsoft' restricting the search to the www.microsoft.com web site.
- How many pages on the Microsoft web server contain the word 'Microsoft?' According to Google, *all of them!* Remember that Google searches not only the content of a page, but the title and URL as well.

- The word ‘Microsoft’ appears in *the URL* of every page on www.microsoft.com.
- With one single query, an attacker gains a rundown of every web page on a site cached by Google.
- There are some exceptions to this rule.

- If a link on the Microsoft web page points back to the *IP address* of the Microsoft web server, Google will cache that page as belonging to the *IP address*, not the www.micorosft.com web server.
- In this special case, an attacker would simply alter the query, replacing the word 'Microsoft' with the IP address(es) of the Microsoft web server.

How this technique can be used

- This technique makes it very simple for any interested party to get a complete rundown of a website's structure without ever visiting the website directly.
- Since Google searches occur on Google's servers, it stands to reason that only Google has a record of that search.

How this technique can be used

- The process of viewing cached pages from Google can also be safe as long as the Google hacker takes special care not to allow his browser to load linked content such as images from that cached page.
- For a competent attacker, this is a trivial exercise.
- Simply put, Google allows for a great deal of target reconnaissance that results in little or no exposure for the attacker.

Finding Directory listings

- Directory listings provide a list of files and directories in a browser window instead of the typical text-and graphics mix generally associated with web pages.
- Figure 8 shows a typical directory listing.

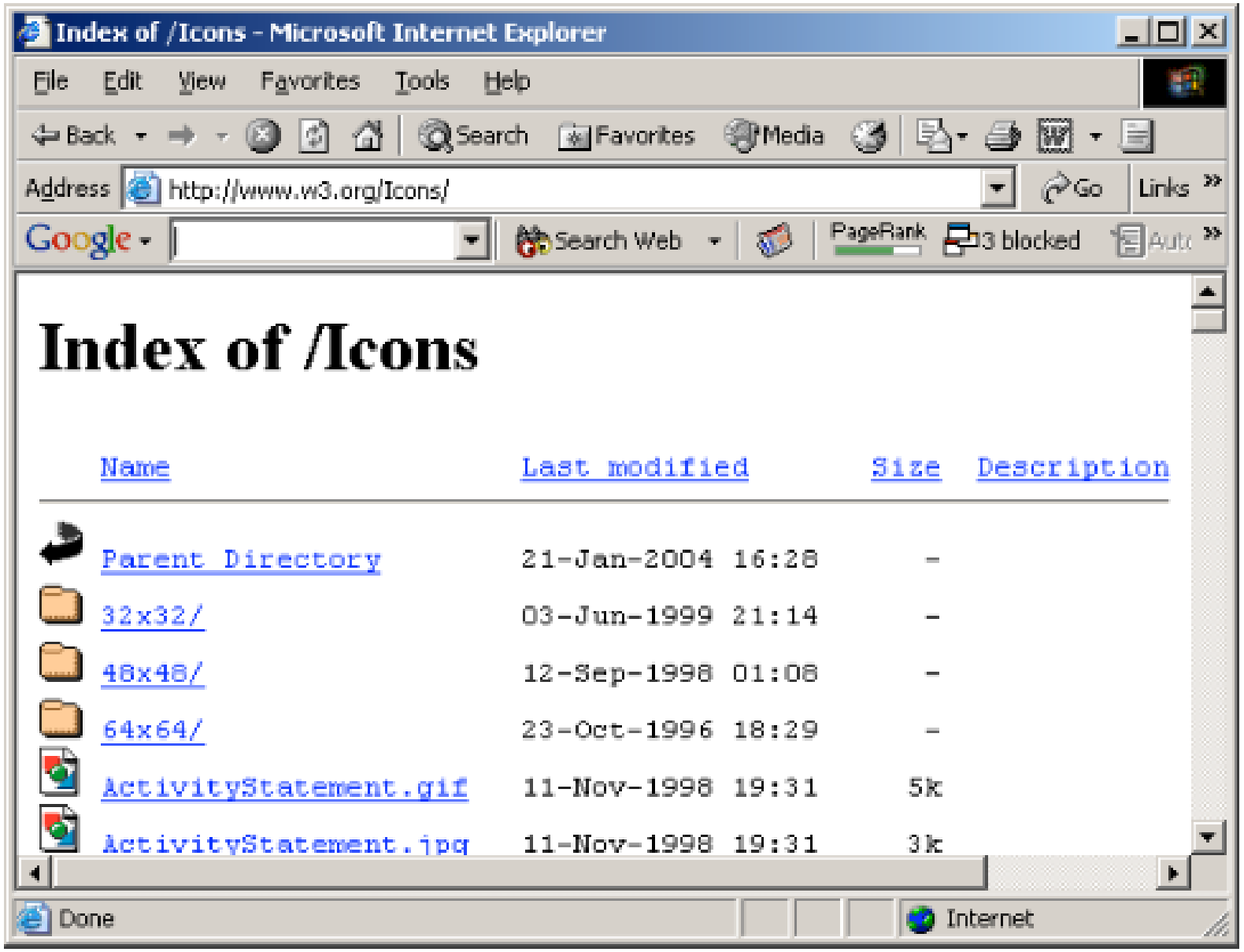

Figure 8: A typical directory listing

- Directory listings are often placed on web servers purposely to allow visitors to browse and download files from a directory tree.
- Many times, however, directory listings are not intentional.
- A misconfigured web server may produce a directory listing if an index, or main web page file is missing.
- In some cases, directory listings are setup as a temporarily storage location for files.
- Either way, there's a good chance that an attacker may find something interesting inside a directory listing

- Locating directory listings with Google is fairly straightforward.
- Figure 8 shows that most directory listings begin with the phrase ‘index of’ which also shows in the title.
- An obvious query to find this type of page might be which may find pages with the term ‘index of’ in the title of the document.

```
“intitle:index.of”,
```

- Unfortunately, this query will return a large number of false-positives such as pages with the following titles:

```
Index of Native American Resources on the Internet
LibDex - Worldwide index of library catalogues
Iowa State Entomology Index of Internet Resources
```

- Several alternate queries provide more accurate results:

```
intitle:index.of "parent directory"
intitle:index.of name size
```

How this technique can be used

- Bear in mind that many directory listings are intentional.
- However, directory listings provide the Google hacker a very handy way to quickly navigate through a site.
- For the purposes of finding sensitive or interesting information, browsing through lists of file and directory names can be much more productive than surfing through the guided content of web pages.
- Directory listings provide a means of exploiting other techniques such as versioning and file searching, explained below.

Versioning: Obtaining the Web Server Software / Version

- via directory listings
  - The exact version of the web server software running on a server is one piece of required information an attacker requires before launching a successful attack against that web server.
  - If an attacker connects directly to that web server, the HTTP (web) headers from that server can provide this information.
  - It is possible, however, to retrieve similar information from Google without ever connecting to the target server under investigation.
  - One method involves the using the information provided in a directory listing.

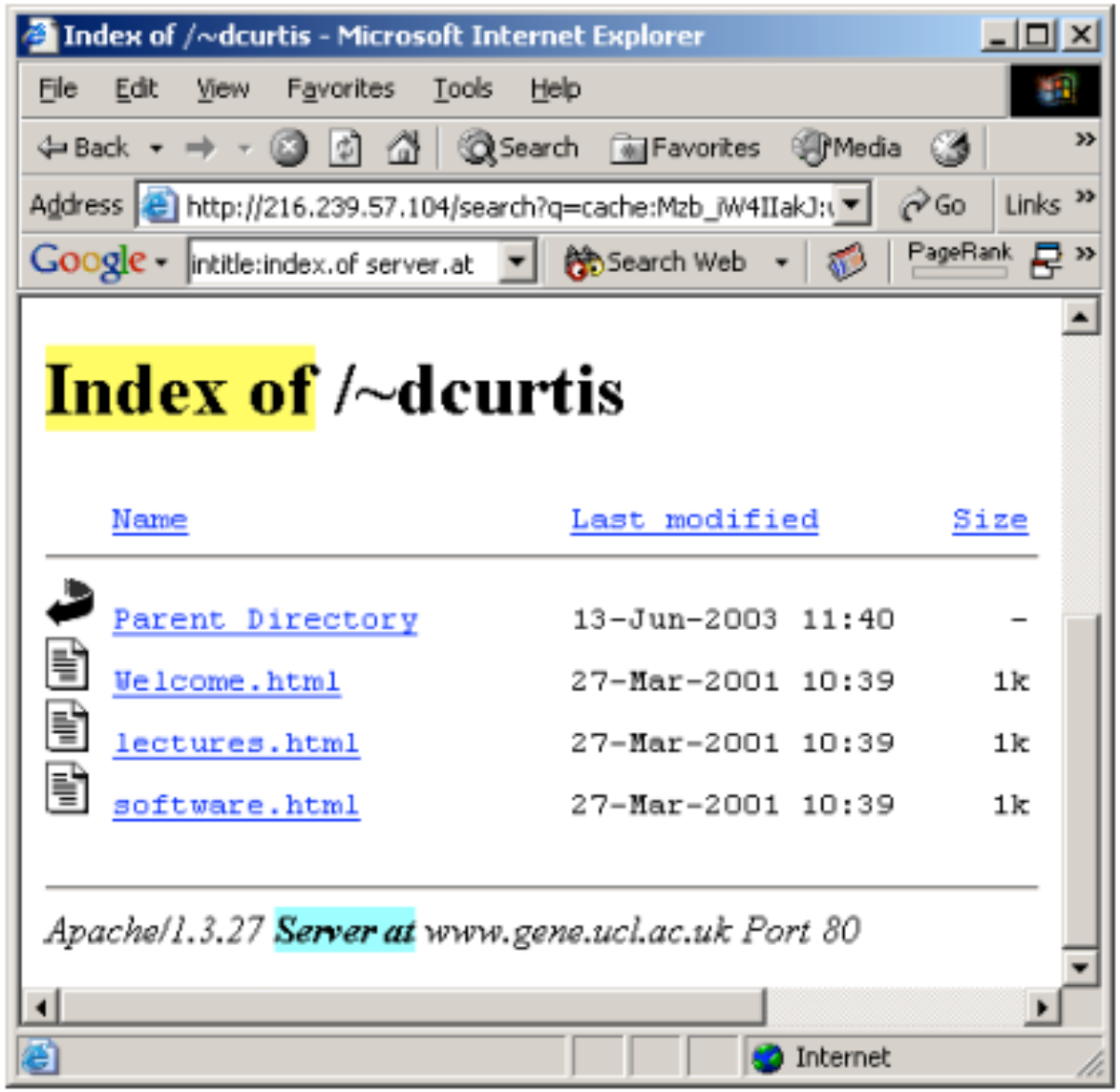

Figure 9: Directory listing "server.at" example

Refferensi

- The Google Hacker's Guide *Understanding and Defending Against the Google Hacker*
- by Johnny Long
- johnny@ihackstuff.com
- http://johnny.ihackstuff.com