

## Internet Search Engines and Site Chaining

### Introduction

The major search engines, Google, Microsoft Live Search, and Yahoo!, can all be used to perform advanced research on web sites. Each of these search engines have some "Advanced Search Options" that can be useful for performing Internet research. This paper outlines a few that may be helpful for identifying websites that are related in some way to other websites of interest.

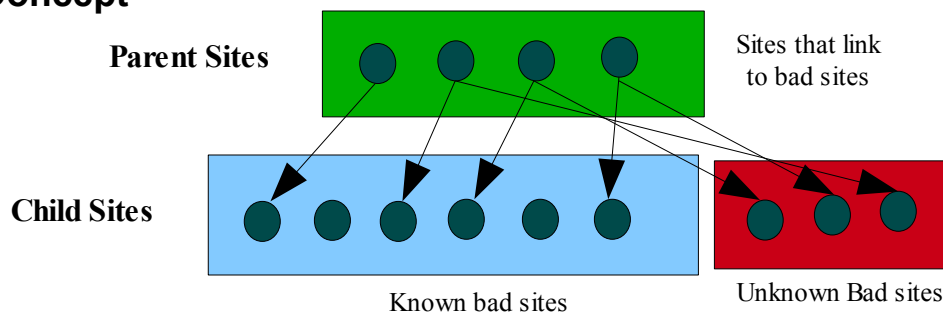
### Site Chaining

To find any sites that link to some suspicious/malicious website. Use the advanced operators supplied by the various search engines.

- Google:
  - "link:" operator – show all websites that link to a specific URL.
- MSN Search:
  - "link:" operator – similar to Google and Yahoo, show all websites that link to a specific URL.
- Yahoo:
  - "link:" operator – show all websites that link to a specific URL.
  - "linkdomain:" operator – shows all websites that link to any pages within specific domain name.

Search operators can be helpful in finding previously unknown sites that may be similar in nature. If you find a site ("parent site") that links to a known bad site ("child site"), the parent site may also link to more child sites that were previously unknown. You can apply the same search technique to the new child sites to see if anymore parent sites can be found.

### Concept

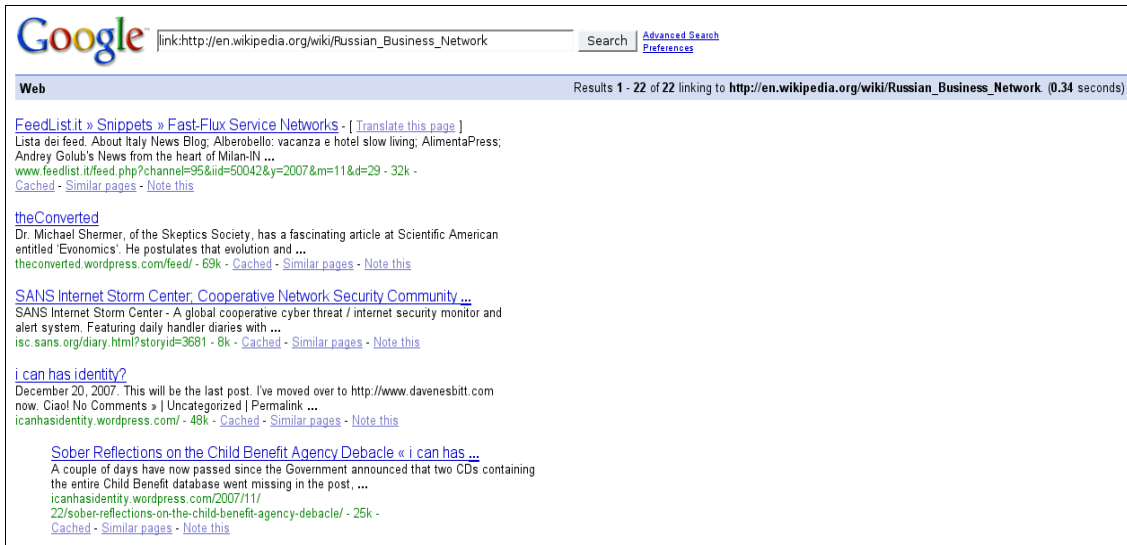


If you apply site chaining to the sites that you know are bad (Child sites on the left), you will identify any site that links to the bad sites (Parent Sites). If you explore the parent sites you may find the new Bad sites (Child sites on the right).

## Example Link Chaining

If I want to see what other websites link to the Wikipedia article about The Russian Business Network ([en.wikipedia.org/wiki/Russian\\_Business\\_Network](http://en.wikipedia.org/wiki/Russian_Business_Network)) I can use Google (and other search engines) to tell me all the websites that link to the Wiki article.

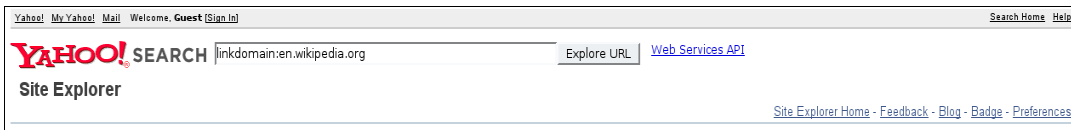
Example using Google (link: operator):



The screenshot shows a Google search for the query `link:http://en.wikipedia.org/wiki/Russian_Business_Network`. The search results are displayed under the 'Web' tab, showing 22 results. The first few results include:

- [FeedList.it » Snippets » Fast-Flux Service Networks - \[ Translate this page \]](#)  
Lista dei feed. About Italy News Blog, Alberobello: vacanza e hotel slow living, AlimentaPress, Andrey Golub's News from the heart of Milan-IN ...  
[www.feedlist.it/feed.php?channel=95&iid=50042&y=2007&m=11&d=29](#) - 32k - [Cached](#) - [Similar pages](#) - [Note this](#)
- [theConverted](#)  
Dr. Michael Shermer, of the Skeptics Society, has a fascinating article at Scientific American entitled 'Economics'. He postulates that evolution and ...  
[theconverted.wordpress.com/feed/](#) - 69k - [Cached](#) - [Similar pages](#) - [Note this](#)
- [SANS Internet Storm Center: Cooperative Network Security Community ...](#)  
SANS Internet Storm Center - A global cooperative cyber threat / internet security monitor and alert system. Featuring daily handler diaries with ...  
[isc.sans.org/diary.html?storyid=3681](#) - 8k - [Cached](#) - [Similar pages](#) - [Note this](#)
- [i can has identity?](#)  
December 20, 2007. This will be the last post. I've moved over to [http://www.davesbitt.com](#) now. Ciao! No Comments » | Uncategorized | Permalink ...  
[icanhasidentity.wordpress.com/](#) - 48k - [Cached](#) - [Similar pages](#) - [Note this](#)
- [Sober Reflections on the Child Benefit Agency Debacle « i can has ...](#)  
A couple of days have now passed since the Government announced that two CDs containing the entire Child Benefit database went missing in the post, ...  
[icanhasidentity.wordpress.com/2007/11/22/sober-reflections-on-the-child-benefit-agency-debacle/](#) - 25k - [Cached](#) - [Similar pages](#) - [Note this](#)

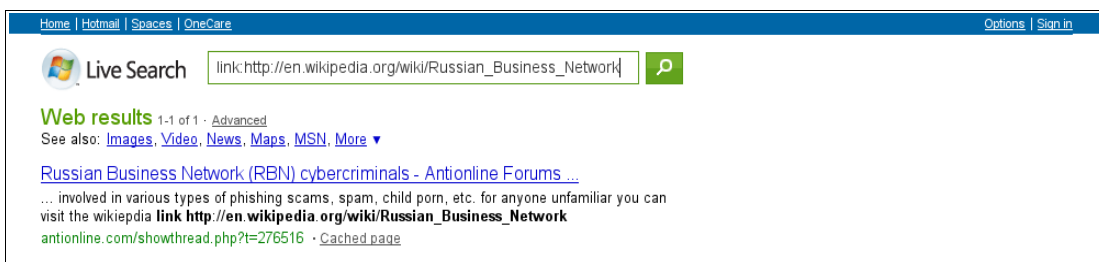
Example using Yahoo (linkdomain: operator):



The screenshot shows a Yahoo Site Explorer search for the query `linkdomain:en.wikipedia.org`. The search results are displayed under the 'Site Explorer' tab, showing 1 result. The result is:

- [Russian Business Network \(RBN\) cybercriminals - Antionline Forums ...](#)  
... involved in various types of phishing scams, spam, child porn, etc. for anyone unfamiliar you can visit the wikipedia link [http://en.wikipedia.org/wiki/Russian\\_Business\\_Network](#)  
[antionline.com/showthread.php?t=276516](#) - [Cached page](#)

Example using MSN (link: operator):



The screenshot shows a MSN Live Search for the query `link:http://en.wikipedia.org/wiki/Russian_Business_Network`. The search results are displayed under the 'Web results' tab, showing 1 result. The result is:

- [Russian Business Network \(RBN\) cybercriminals - Antionline Forums ...](#)  
... involved in various types of phishing scams, spam, child porn, etc. for anyone unfamiliar you can visit the wikipedia link [http://en.wikipedia.org/wiki/Russian\\_Business\\_Network](#)  
[antionline.com/showthread.php?t=276516](#) - [Cached page](#)

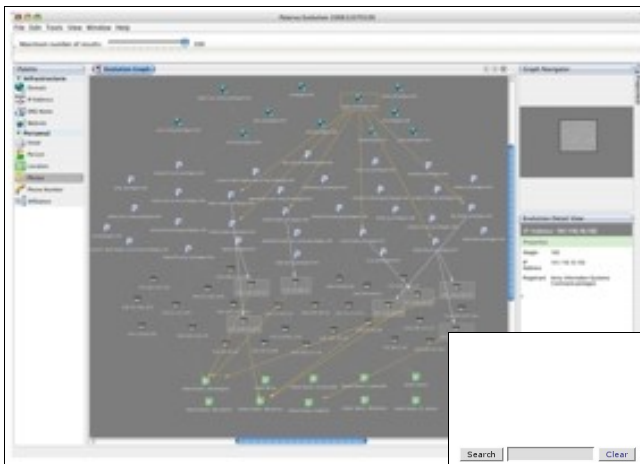
## Automated Analysis

All of the search engines mentioned so far provide a programmatic API for allowing automated tools to use them. This leaves open the possibility of creating an automated system for finding new sites of interest based on some starting set of "interesting" sites.

## Other Tools

Maltego is an advanced search tool that can show associations between sites, email addresses, and people. There are currently two different interfaces for accessing this system.

- Web Interface: <http://www.paterva.com/web2/maltego/maltego-web-interface.html>
- GUI Interface: <http://www.paterva.com/web2/maltego/maltego-gui-1.0-download.html>



Maltego GUI  
Interface

Maltego Web Interface



## Operational Security Concerns

When performing online research where secrecy is a concern operational security practices should be used. Site operators may be able to figure out your search techniques and counter them if you are sloppy.

- **Referrer Field.** Every time a link is clicked using a web browser hidden information is transmitted to the web server that the link takes you to. This information includes the URL of the site that the original link came from ("Referrer" field). This referrer field can give away specific search terms used to find the site. Sites can identify specific search terms used to find their site. This may give away tactics and techniques you are using to find sites of interest.
- To stop sites from knowing how you specifically found them, instead of clicking links found in search engines you can copy the site URL, open a new browser window (or tab), and paste the URL into site box. This will cause the Referrer field to be empty.

**For more information on how search engines and research tools can help secure your network and computing resources, and for more information on Linux and Mac-based security, contact VeriSpect Security Services and Consulting – we're experts on securing your data and can create a custom security plan to fit your needs and your budget. More information is available on the web at [www.Verispect.net](http://www.Verispect.net) or through email at [info @ verispect.net](mailto:info@verispect.net).**