

Copyright Windrose Software 2005

# **OptiSpider User's Guide**

**February 8, 2005**

Introduction.....	3
System Requirements.....	3
Computer.....	3
Proxy Support.....	3
Firewalls.....	3
Tutorial .....	4
Resource Section.....	9
Main Window.....	9
Menus.....	10
Tabs.....	12
Table Operations.....	14
Command View.....	15
Summary View.....	18
Compare View.....	20
Pages View.....	21
Domains View.....	21
Counts View.....	22
Links View.....	22
Robot View.....	23
Log View.....	24
Notes View.....	24

## ***Introduction***

The general purpose of OptiSpider is to present in one place everything you need to know to optimize your site.

- Spiders your entire site to make sure that every page on your site is findable and that there are no spider loops within your site. If OptiSpider can't spider your site correctly, it is likely that the search engines will not be able to either.
- Records the link text of every single internal and external link.
- Records the link text of every single internal and external link.
- Compares the link text that points to each page with the actual topic of that page.
- Provides the inbound and outbound link counts for each page in your site: how many pages on the site link to it, how many pages does it link out to, how many of those outbound links stay onsite, and how many go offsite. These statistics are used to optimize page rank distribution within the site.

Given a target URL on the command pane, OptiSpider spiders all the pages reachable from that page, that are on the same domain as the target. It does not spider off site pages – this would cause it to eventually spider the entire web!

## ***System Requirements***

### ***Computer***

<b>CPU</b>	266 MHZ or faster recommended
<b>Memory</b>	64 Mb RAM minimum
<b>Freespace</b>	70 Mb minimum prior to install
<b>Operating System</b>	Windows 98 or later Linux, any version Macintosh OS-X 10.2.6 or later
<b>Additional System Software</b>	Sun JAVA runtime version 1.4.1 or later

### ***Proxy Support***

OptiSpider supports internet access via a proxy. See the description of the Setup menu for setup instructions.

### ***Firewalls***

These are a common problem with programs, such as OptiSpider, that access the internet. Refer to your firewall vendor documentation for instructions.

# Tutorial

The section gives you a quick overview of OptiSpider. You should also view the 5 Minute Tour which may be seen online by opening the OptiSpider program and clicking on Help/Online Tour. This tutorial does not explain every option in OptiSpider because the purpose is just to get you started.

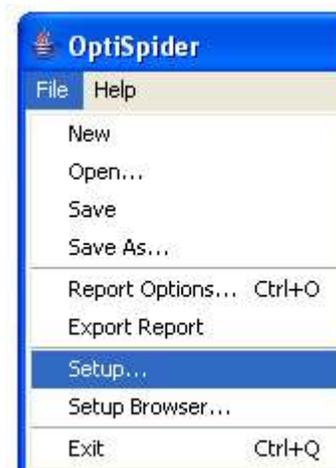
**Step 1:** If you have a dial up connection, connection to the Internet now. If you have a direct connection, you do not need to do anything.

**Step 2:** Open the OptiSpider program by clicking on the Start Menu, then clicking on OptiTools, then on OptiSpider..



**Step 3** Setup OptiSpider with proxy settings.

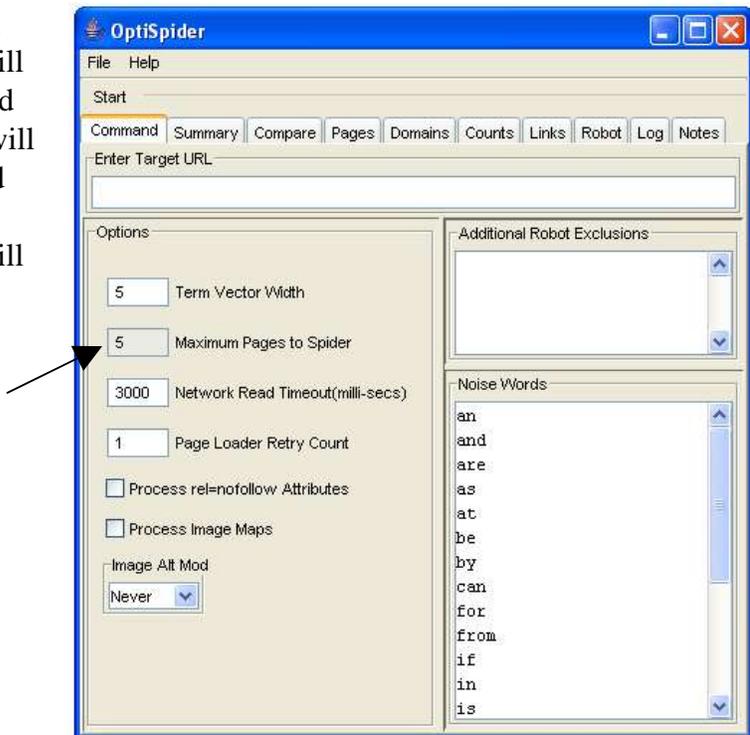
To setup OptiSpider with proxy settings click Setup on the File menu. The Setup window will appear.



If you use a proxy server, enter the proxy server settings here. The browser you are currently using is a good source for how to set these correctly, or ask your system administrator for assistance.



When OptiSpider first comes up, the Maximum Pages to Spider will be set at 5. If you have purchased OptiSpider, the welcome email will contain the authorization key and instructions. If you have not yet purchased OptiTools, you will still be able to spider small sites.

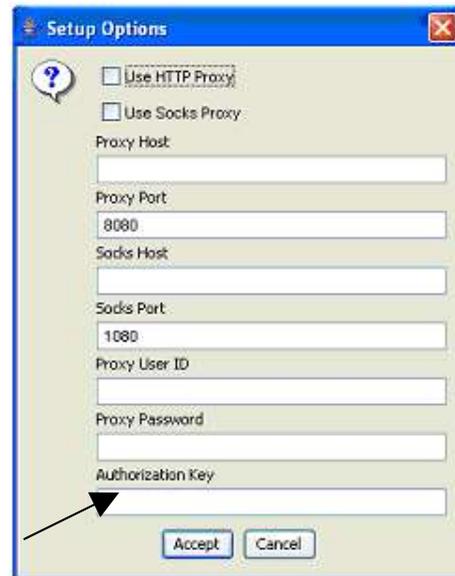


**Step 4:** Enter your authorization key:  
Click Setup on the File menu. The Setup window will appear.

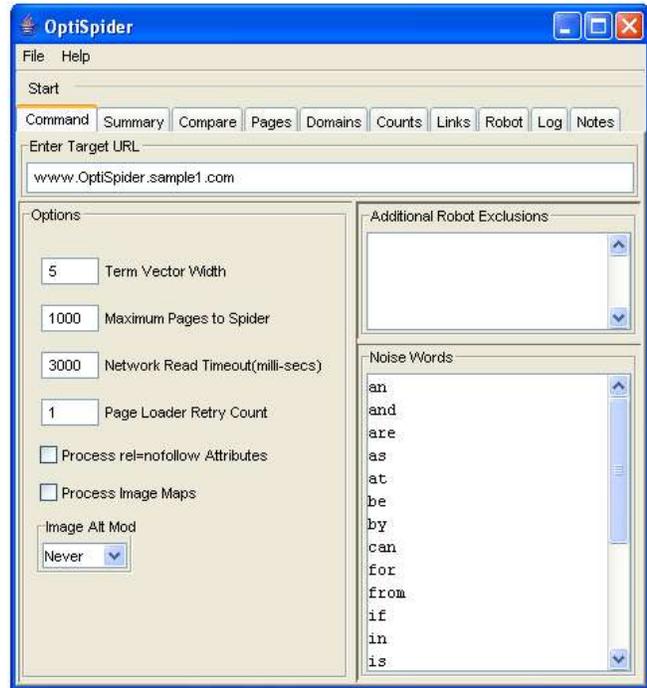


Authorization Key: This is a string that was sent to you if you purchased OptiSpider. This key enables the base features and any optional you purchased. Without an authorization key, the OptiSpider software defaults to allowing only 5 pages to be spidered.

After entering the authorization key, hit "accept", quit the software, and reopen it. The only time OptiSpider looks at the authorization key is when the software starts.



**Step 5:** When you return to the Command View, you will see the Maximum Pages to Spider enabled. Type “www.OptiSpider.sample1.com” in the Target URL box. Leave all the options with their default values.



Here’s a quick description of the list boxes and options you can set for your analysis. For this tutorial, just leave everything at it’s default value. For a complete description of each of these values, see the Resource Section.

**Enter Target URL** Enter the url of the target page you wish to analyze. For this tutorial set this to www.acws.com

**Additional Robot Exclusions** These are url pages or page prefixes that will be blocked from spidering by Optispider. It is as if you added these paths to the robots.txt file for a domain.

**Noise Words** The words that will not be included in the page topic and link reputation for each page.

**Options** – For this tutorial leave all options set to their default vaules

**Term Vector Width** A term vector is a list of terms or words that describe what a document is about.

**Maximum Pages To Spider** OptiSpider will either stop with it comes to the end of the site or it will stop when it gets to the maximum page count.

**Network Read Timeout** Number of millisecs to load a page before a timeout fault is declared.

**PageLoader Retry Count** The number of times OptiSpider will try to load a page.

**Process ref=notfollow  
Attributes**

This will cause OptiTools to honor the rel="nofollow" attribute on links as proposed and being implemented by the major search engines.

**Process Image Maps**

This will cause OptiTools to spider links in the <area> tags that implement HTML image maps.

**Image Alt Mod**

This controls how OptiTools treats text in the alt attribute of images that are contained in links. The default is "never" which means the text is ignored.

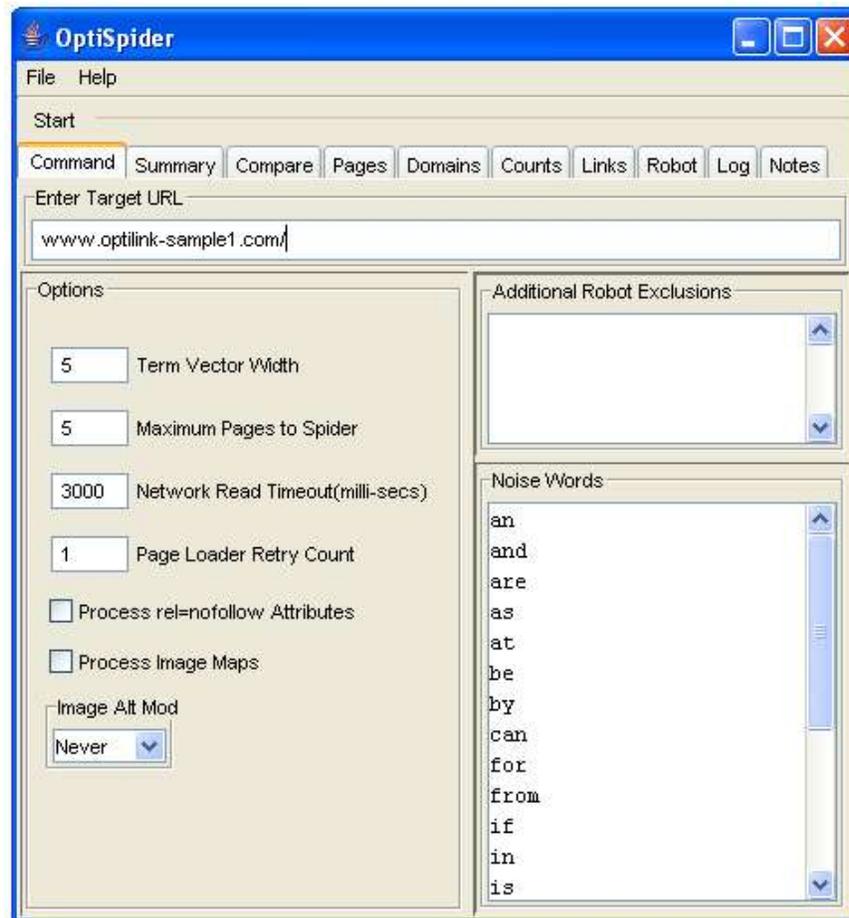
**Step 6:** Press the start button.



**Step 7:** Once OptiLink finishes processing your command, it displays the Summary view, the first of six result views that show different aspects of link reputation for your target page. Here you'll see your target page listed along with its IP address, the number of inbound links to the page reported by the search engine and a few other important particulars. A detailed description of the Summary view along with the other 5 views may be found in the Resource section

# Resource Section

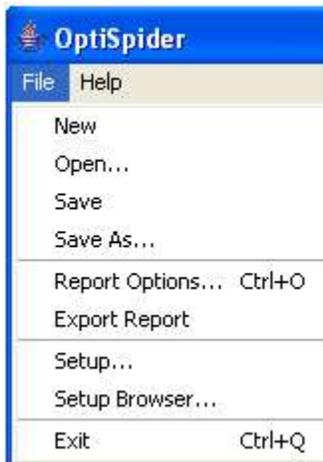
## Main Window



The main window of OptiSpider contains the File and Help menus, and 10 tabs. When a tab is selected it reveals a “view”. The views allow the user to enter information to define the analysis he wants to do or provide different views of the data that has been gathered by OptiSpider about the target site. The view shown here is the Command View.

## Menus

### File Menu



**New** Begins a new report.

**Open** Reads in a previously created OptiSpider analysis.

**Save and Save As** Saves an OptiSpider analysis to a file.

**Report Options** Control which of the results views appear in the report

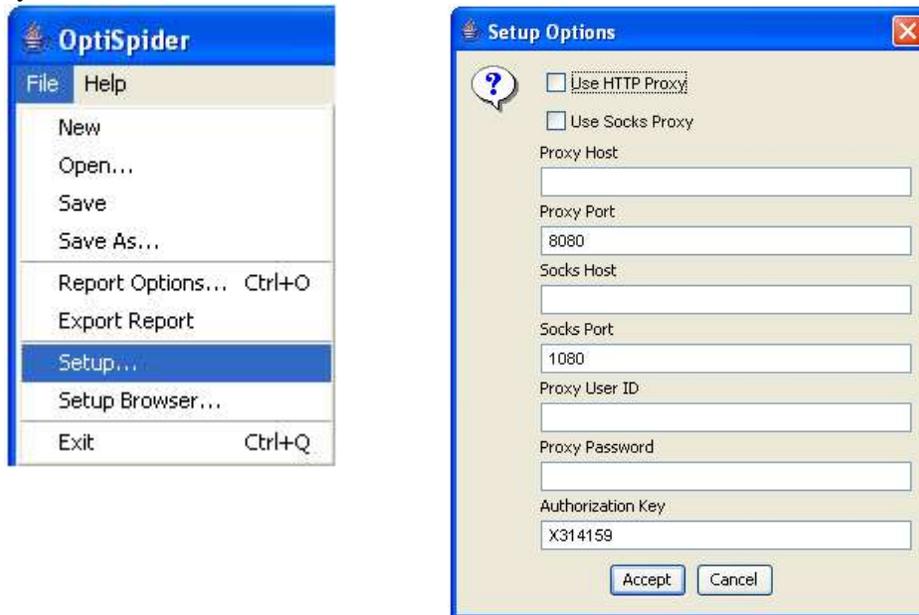
**Export Report** formats the information from an OptiTools program in an html file in a temporary folder on your machine, and launches the default browser to display the user report. The report can then be printed, or saved, or mailed

**Setup** Define internet parameters and authorization key

**Setup Browser** An expert feature for modifying the browser that will be accessed by the OptiTools.

## Setup

Clicking on Setup in the File Menu brings up the Setup Window which allows the user to define parameters for a proxy server as well as enter your authorization key. Whatever browser you are currently using is a good source for how to set these correctly, or your system administrator.



**Authorization Key:** This is a value that was sent to you if you purchased OptiSpider. This key enables the base features and optional features in the program. If this field is blank, you are operating the OptiSpider Viewer which is free software that you can distribute as you like, but it cannot perform an analysis. It can only view an existing analysis built by others. You can do File Open to view results someone else created, but you don't have the ability to create new analyses. If there is an authorization key, then you have purchased a copy of OptiSpider and possibly one or more optional plug-ins.

As soon as OptiSpider is installed, it will come up as an unauthorized viewer with the Start button disabled. If you have purchased OptiSpider, the welcome method will contain the authorization key and instructions. To do file setup enter the authorization key, hit "accept", quit the software, and reopen it. The only time OptiSpider looks at the authorization key is when the software starts. The authorization key will contain information about what options you have purchased. You have the base program, the base program plus some additional search engines, the spider, or any combination. If you want to purchase additional options in the future, you can do that on line and you are immediately enabled for the options you have purchased.

## Report Options

Brings up a window that lists the report views. Click the view or views you wish to include in your report.



## Help Menu



**Online Tour**--Take a 5-minute tour of OptiSpider features.

**Online FAQ**--This page provides answers to some of the most common questions and solutions to the most common problems for the OptiSpider product.

**Reference**--A link to OptiSpider reference material.

### Tabs

Clicking on a tab brings up a “view”. A view is selected by clicking on the tab, which brings the selected view in front of the others.

<b>Command</b>	Defines the target page you want to analyze.
<b>Summary</b>	Provides a brief outline of the results found. Once an analysis is completed, the Summary view is automatically displayed.

<b>Compare</b>	Compares the Topic of the selected page with the Reputation of the page. The Topic is measured by what words, and combinations of words, occur most frequently on the page. The Reputation of a page is a measure of what other pages say the subject page is about. Reputation is measured by the number of incoming links that use various words and combinations of words. The "best" pages in any given category are those that have Topic and Reputation measures that are in close agreement.
<b>Pages</b>	Lists the pages that link to your subject page in terms of Hubs and Authorities. An authority page can thought of as a very popular destination, with many inbound links and relatively view links to other pages. What links there are of course, would be very important. A Hub page is just opposite-- more like a train station with numerous outgoing links. Hubs may have many incoming links as well if they are "very well known" hubs. Use the Pages View to decide which page links are more important than others, and why.
<b>Domains</b>	Shows all the domains that contributed links to your subject page and the IP address of each.
<b>Counts</b>	
<b>Links</b>	Shows the precise text used to link to your page. The text of links to your page determines your "Reputation"-- what others think your page is really about. Maybe they're right? It could be that you should change or enhance your page measures to move your Topic closer to your Reputation.
<b>Log</b>	Provides a blow-by-blow monolog of progress. This can also be reviewed after the fact for error conditions should the results displayed in the other views appear incorrect.
<b>Notes</b>	Make whatever notes you want here and then send the OptiLink report, along with your notes, to the next viewer.

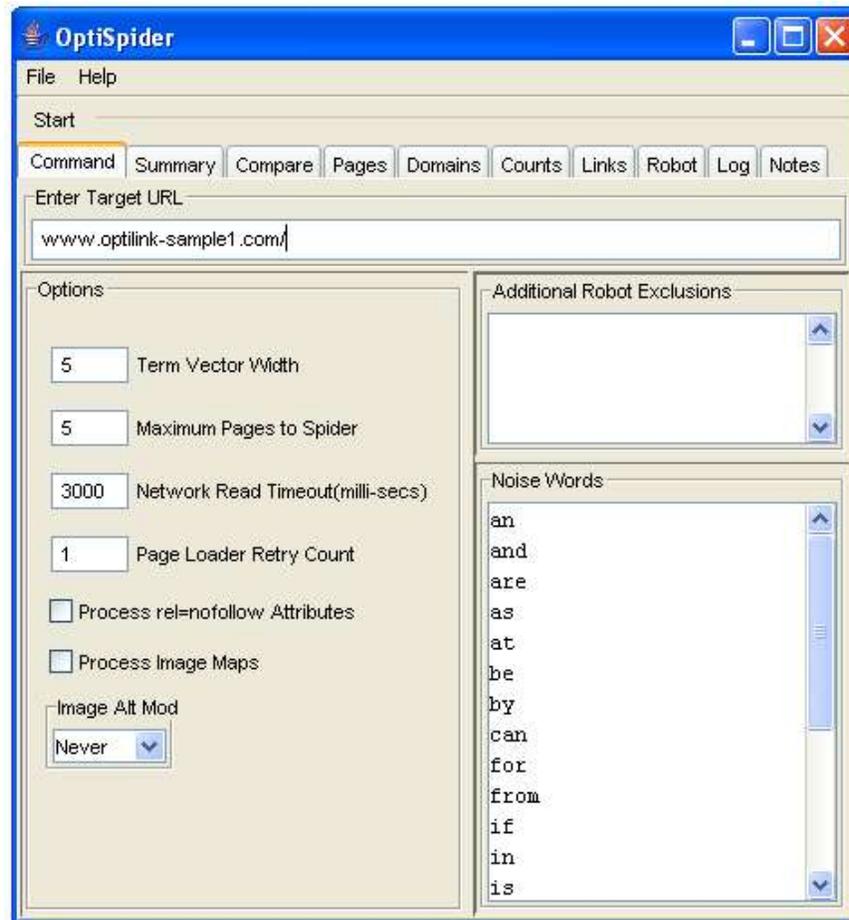
## Table Operations

Each of the views that presents data (Summary, Compare, Domains, Pages, and Links) present that data in the form of tables. Each of the tables can be manipulated to reorder and sort the columns to provide the most helpful format for the data.

<b>Sort Ascending</b>	Right click in the column name cell to sort alphabetically if the column contains strings. If the column is numeric, this will sort in descending numerical value.
<b>Sort Descending</b>	Shift/Right click in the column name cell will sort in reverse alphabetical order. If the column is numeric it will sort in ascending numeric value.
<b>Resize Column</b>	Move the curser over the dividing line between two columns. When the right/left arrow appears, depress the right mouse button, and move the edge of the column to the desired width.
<b>Reorder Columns</b>	Depress the right mouse button in the column name cell and drab the column to the position you wish.

## Command View

The Command View is where you set up the criteria for the analysis you wish to run.



**Enter Target URL:** Given a target URL on the command pane, OptiSpider spiders all the pages reachable from that page, that are on the same domain. It does not spider off site pages – this would cause it to eventually spider the entire web!

**Term Vector Width:** A term vector is a list of terms or words. The terminology “term vector” comes from academic research and is a highly efficient way of recording what a document is about. OptiSpider counts all of the words on the page and counts the text in the links pointing at the page. Then OptiSpider picks out the words that appear most frequently – the most common words on the page. The term vector is the list of the most commonly used words that is shown in the results view. Term vector *width* is the number of words of topic and reputation the OptiSpider will record about a page. . If the term vector width is 5, that means OptiSpider lists the most common 5 words on the page.

**Maximum Pages To Spider:** OptiSpider recursively loads all the pages on your site. When OptiSpider sees a page, it loads all the links on the page. If those pages are on the

same domain, OptiSpider loads all those pages, and so on. There are some sites, which because they are dynamically generated, might never terminate. Or a site might be very large, and you only want to get a sample of how internal linking is done, rather than measuring every single page. Setting “Maximum Pages to Spider”, gives you a way to stop OptiSpider processing. OptiSpider will either stop when it comes to the end of the site or it will stop when it gets to the maximum page count.

**Network Read Timeout (milliseconds):** If the server doesn’t respond within this number of milliseconds it will declare a fault. There are two retries that will be done automatically. If the page times out, you will see IO errors in the log, and OptiSpider will go on.

**Page Loader Retry Count:** The number of times that OptiTools will retry a page load. Default is 1, meaning that OptiTools will make a single attempt following a failed load. Zero means that OptiTools will not retry at all. This is useful in marginal or slow connections or with websites that are heavily loaded, generally slow to respond, or unreliable.

**Process rel=nofollow Attributes:** This will cause OptiTools to honor the rel=“nofollow” attribute on links as proposed and being implemented by the major search engines.

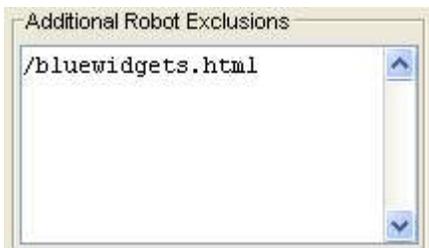
Links to SE announcements:

- [Yahoo Search Blog Notes](#)
- [MSN Search Blog](#)
- [Google Search Blog](#)

**Process Image Maps:** This will cause OptiTools to spider links in the <area> tags that implement HTML image maps.

**Image Alt Mod:** This controls how OptiTools treats text in the alt attribute of images that are contained in links. The default is “never” which means the text is ignored. The “if no text” setting will cause OptiTools to use alt text as link text if there is no plain text in the link. “always” will cause OptiTools to always include alt text as link text. Prior versions of OptiTools operated as “never”.

#### **Additional Robot Exclusions:**



In this example, OptiSpider will ignore the file bluewidgets.html. The syntax is the same as in the robots.txt file without the word “Disallow:”.

These are url pages or page prefixes that will be blocked from spidering by Optispider. It is as if you added these paths to the robots.txt file for a domain.

Web Robots, also called Spiders, are programs that traverse the Web automatically. The robots.txt file is a method that allows you to indicate to visiting robots which parts of your site should not be visited by the robot.. Anything not prohibited is OK.

The Robots META tag allows HTML authors to indicate to visiting robots if a document may be indexed, or used to harvest more links. For example: `<META NAME="OPTISPIDER" CONTENT="NOINDEX, NOFOLLOW">`. The document containing this should not be indexed, or used to gather links. Not all spiders have implemented these metatags.

Optispider honors the robots.txt file and the metatags, “noindex” and “nofollow”. The robots exclusion is there so we can block a certain area of the site from Optispider, so that you can look at only a piece of the site.

For more information on the Robots.txt file:

The Standard For Robot Exclusion                      <http://www.robotstxt.org/wc/norobots.html>

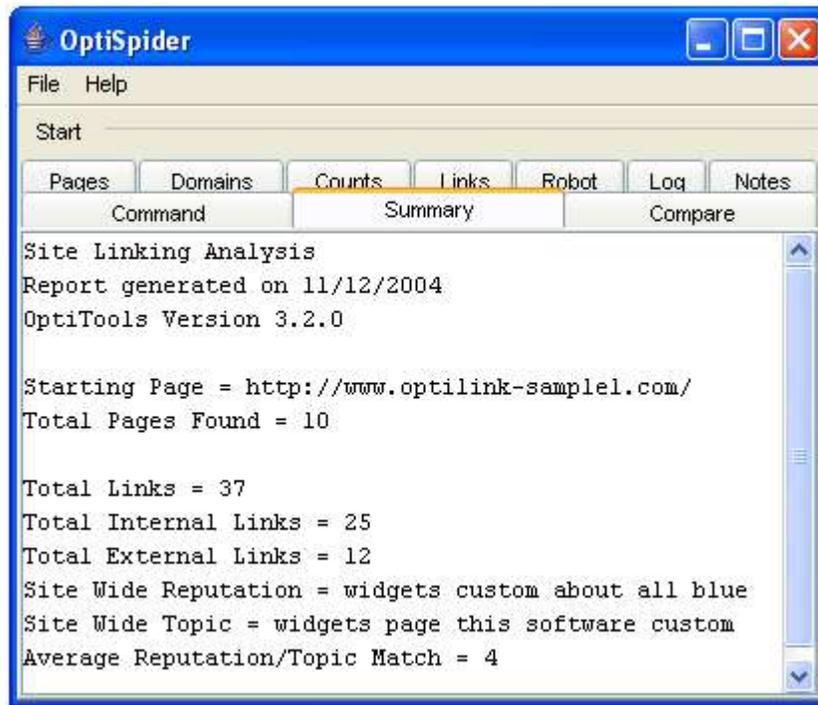
Some Examples    <http://www.searchtools.com/robots/robots-txt.html>

Meta tags    <http://www.searchengineworld.com/metatag/robots.htm>

**Noise Words:** The words that will not be included in the page topic and link reputation for each page. For example, ‘the’ is a common word so that you end up in the situation where the topic of the page will have the word ‘the’ in it. It’s not an important word for optimization purposes, so we skip over it when we do the topic and reputation computations.

**Restore Defaults:** Restores the default words set.

## Summary View



The Summary View provides the following information:

**Total Pages Found:** The total pages in the site

**Total Internal Links:** The number of links provides a summary of the Counts View  
Total Internal Links shows the total internal links in the site.

**Total External Links:** The number of links provides a summary of the Counts View  
The Total External Links shows the total external links in the site.

**Sitewide Reputation.** This does not appear in any other view. Instead it is a weighted average of numbers presented for the topic and reputation for all the pages in the site.

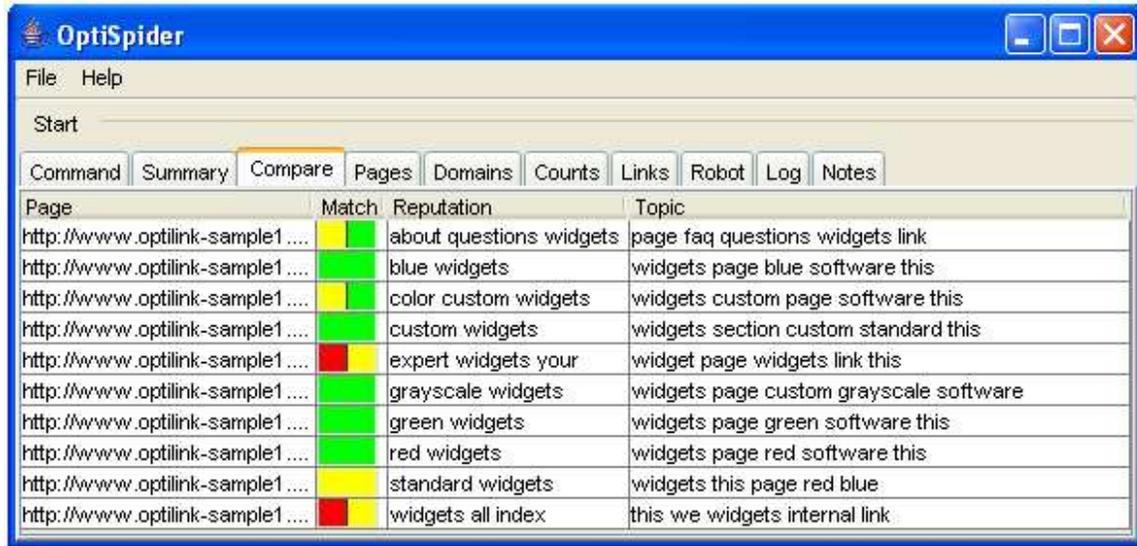
While we do not know of any search engines that calculate sitewide topics or reputation, this information is useful in analyzing the consistency of the site and as an indication of how broad or narrow the site is.

**Site Wide Topic:** a statistical summary of the extent of match in the Compare View. It is a count of the different light bulbs you have turned on., This lists the number of pages with each light bulb glowing. Green corresponds to 5, and Red corresponds to 1. For example, if you have 3 pages that are 5 or Green, and 3 pages that are Yellow or 3, the site average is 4.

**NOTE:** Sitewide measures are NOT used for SEO purposes. When doing SEO analysis, all pages stand alone. The Sitewide matches are only for humans as a guide to how well focused the site is. Focus does not affect search engine ranking. If a person is in a market that has several submarkets, and they have one site devoted to each submarket. The Sitewide measures would tell them how well they have achieved splitting their market into submarkets. These numbers are for humans to evaluate their site design, but they are not useful for search engine optimization.

**Average Reputation/Top Match:** This is the row from the Compare View for the page you entered in the Command View when you executed OptiSpider. This number does not provide any SEO utility, but it does tell you in general how consistently you are doing in your linking across your pages, so it might be a quality factor that you would want to pay attention to, or not.

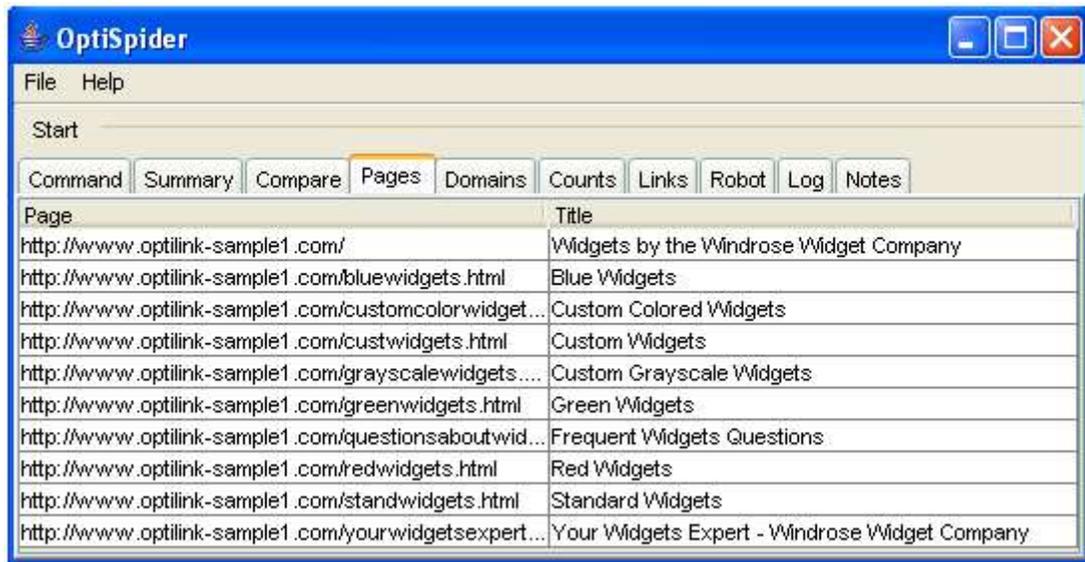
## Compare View



<b>Page</b>	For every page OptiSpider found, it will list the page name (without the host address) in the Page column.
<b>Match</b>	The Match column, also called the Traffic Light column, contains an indicator of how closely the topic and reputation of the page align. If the words in the Reputation column are contained in the Topic column, the light is green, which indicates that the topic and reputation of the page match. If the words in the Topic and Reputation columns don't match at all, then the light is red indicating that the text of the links to that page should be changed. The matching algorithm assigns a number from 1 to 5 indicating how well the topic and reputation match. A green light indicates a 5 – a good match! A red light indicates a 1 – no match! If there are a few matches, but most do not match, then a light that is half yellow and half red is shown. If only half the words match, a yellow light is shown. If there are more matches than non matches, a light that is half green and half yellow is shown
<b>Reputation</b>	lists the top few most used words found in the text of links to that page.
<b>Topic</b>	lists the top few most used words on the page.

Note: The order of the columns can be changed by clicking in the title bar on the column title and dragging it to the position you want.

## Pages View



For each page spidered, this view gives the page title. This page will tell the user exactly what was spidered.

## Domains View



The Domains View provides a list of the domains of all links that were mentioned on the site.

## Counts View

Page	Inbound	Outbound	Onsite	Offsite
http://www.optilink-sample1.com/	9	7	4	3
http://www.optilink-sample1.com/bluewidgets.html	1	3	2	1
http://www.optilink-sample1.com/customcolorwidgets.html	1	3	2	1
http://www.optilink-sample1.com/custwidgets.html	4	5	4	1
http://www.optilink-sample1.com/grayscalewidgets.html	1	3	2	1
http://www.optilink-sample1.com/greenwidgets.html	1	3	2	1
http://www.optilink-sample1.com/questionsaboutwidgets.html	1	2	1	1
http://www.optilink-sample1.com/redwidgets.html	1	3	2	1
http://www.optilink-sample1.com/standwidgets.html	5	6	5	1
http://www.optilink-sample1.com/yourwidgetsexpert.html	1	2	1	1

The Counts View gives, for every page on your site, the number of links to it from within your site, and the number of links out of that page. For the outbound links, this view provides the number of links that are internal, and the number that are external.

## Links View

Page	Target	Link Text
http://www.optilink-sample1.com/	http://www.optilink-sample1.com/questionsaboutwidgets.html	Questions About Widgets
http://www.optilink-sample1.com/	http://www.optitext.com/	Link Reputation Analyzer
http://www.optilink-sample1.com/	http://www.optilink-sample1.com/standwidgets.html	Standard Widgets
http://www.optilink-sample1.com/	http://www.optilink-sample1.com/yourwidgetsexpert.html	Your Widgets Expert
http://www.optilink-sample1.com/	http://www.optilink-sample3.com/	Sample 3
http://www.optilink-sample1.com/	http://www.optilink-sample2.com/	Sample 2
http://www.optilink-sample1.com/	http://www.optilink-sample1.com/custwidgets.html	Custom Widgets
http://www.optilink-sample1.com/bluewidgets.html	http://www.optitext.com/	Link Reputation Analyzer
http://www.optilink-sample1.com/bluewidgets.html	http://www.optilink-sample1.com/standwidgets.html	Standard Widgets

The Links View provides the information that went into the reputation analysis shown in the Compare View. This view shows information on the links that are working. ( You can choose to verify external links, or not).

**Page column** lists the pages that links have occurred on.

**Target column** lists the target of the link (can be internal or external). If you click on this column, the results will be ordered by url, which will group the external links together.

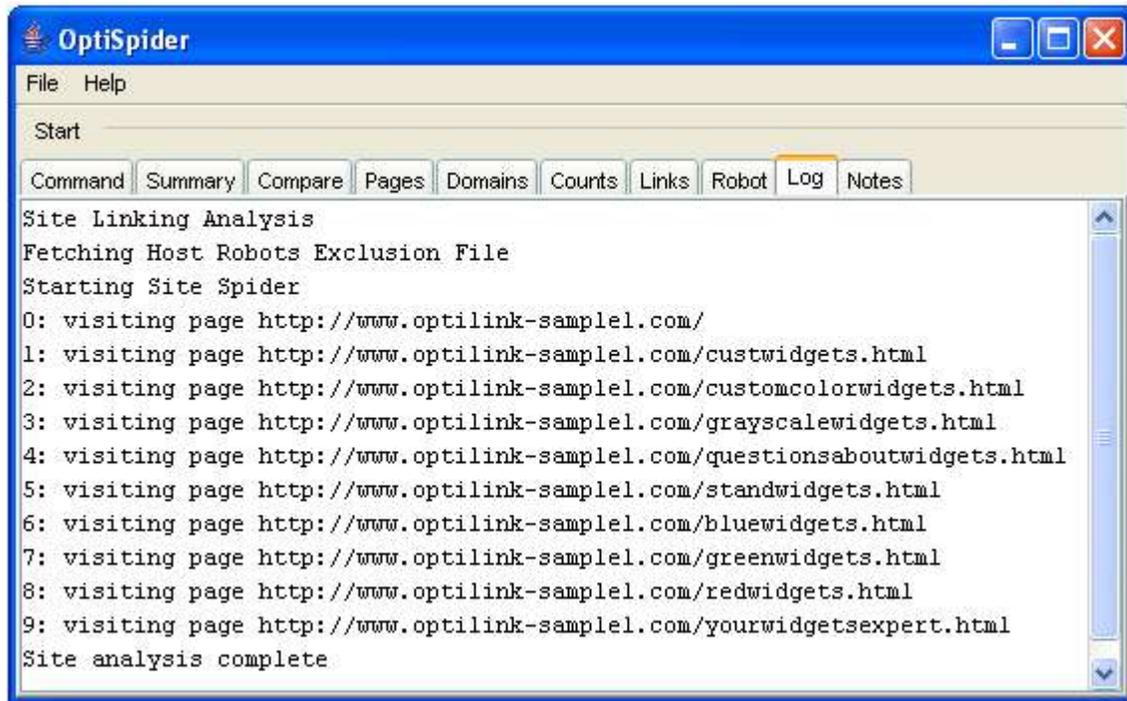
**Text column** gives the text of the link.

## **Robot View**



The Robot.txt file tells a spider which files on the server to ignore. Following an OptiSpider analysis, the Robot View displays the robot.txt file. If you are not getting some pages spidered, or you are getting some pages spidered that you think should be, look at the robot.txt file.

## Log View



As the OptiSpider analysis engine runs it writes informational and error messages to the Log View. Once OptiSpider is finished, a COMPLETE message is written and the Summary View is displayed. You can view the log during execution to make sure the program is accessing the network without problem and to see that links are being discovered and processed. Upon completion, check the log if the results shown in the other views seems incomplete or incorrect.

## Notes View

The Notes View is an edit window where you can write whatever comments you wish about the report that was generated. This provides a convenient way for people who are reviewing the report to record information to pass along to the next reviewer.