

# Table of Contents

|                                      |           |
|--------------------------------------|-----------|
| Foreword                             | 0         |
| <b>Part I Welcome</b>                | <b>3</b>  |
| <b>Part II What's New</b>            | <b>3</b>  |
| <b>Part III Software License</b>     | <b>3</b>  |
| <b>Part IV Purchasing</b>            | <b>5</b>  |
| <b>Part V Product Overview</b>       | <b>6</b>  |
| 1 Overview .....                     | 6         |
| 2 Features .....                     | 8         |
| <b>Part VI Working with Projects</b> | <b>9</b>  |
| 1 Overview .....                     | 9         |
| 2 Step 1 .....                       | 10        |
| Spiders .....                        | 10        |
| Selecting Spiders .....              | 11        |
| Modifying the Spider List .....      | 12        |
| 3 Step 2 .....                       | 13        |
| Disallow .....                       | 13        |
| Grouping Spiders .....               | 14        |
| Setting Disallow Options .....       | 15        |
| Setting Allow Options .....          | 15        |
| 4 Step 3 .....                       | 16        |
| robots.txt .....                     | 16        |
| 5 Step 4 (Professional Only ) .....  | 16        |
| Spider Visits .....                  | 16        |
| <b>Part VII Reference</b>            | <b>17</b> |
| 1 Command Line Options .....         | 17        |
| 2 Select By .....                    | 20        |
| 3 Add/Edit Spider .....              | 20        |
| 4 Mini HTML .....                    | 21        |
| 5 Select Home Page Location .....    | 24        |
| 6 Refresh .....                      | 24        |
| 7 Upload robots.txt .....            | 24        |
| 8 Export robots.txt .....            | 24        |
| 9 Select Access Log Location .....   | 24        |
| 10 Log File Format .....             | 25        |

|                              |  |           |
|------------------------------|--|-----------|
| 11                           | Add Spider to Step 1 .....                         | 26        |
| 12                           | Order Spider Visits By .....                       | 26        |
| 13                           | Export Spider Visits .....                         | 27        |
| 14                           | Export Spider Logs .....                           | 28        |
| 15                           | Export Spider IP Addresses .....                   | 29        |
| 16                           | Options .....                                      | 29        |
| 17                           | Date/Time Formatting Options .....                 | 31        |
| 18                           | FTP Browse Folder .....                            | 31        |
| 19                           | Enter Authorization Code .....                     | 32        |
| 20                           | About Robot-Manager .....                          | 32        |
| <b>Part VIII How Do I...</b> |  | <b>33</b> |
| 1                            | Add robots.txt to GoLive? .....                    | 33        |
| 2                            | Add robots.txt to FrontPage? .....                 | 33        |
| 3                            | Upload robots.txt to my web server? .....          | 34        |
| 4                            | Allow all spiders full access to my web site ..... | 34        |
| 5                            | Exclude all spiders from my web site .....         | 34        |
| 6                            | Direct country specific spiders? .....             | 34        |
| 7                            | Present optimized pages to search engines? .....   | 35        |
| 8                            | Find out where my log files are located? .....     | 37        |
| <b>Index</b>                 |  | <b>38</b> |

## 1 Welcome

# Robot-Manager

The Premier Robot Exclusion File Generator  
and  
Spider Visit Tracking Software

## Version 3.1

©Copyright 2001-2002 Sophtware.com, Inc.  
All rights reserved.

## 2 What's New

The latest release of Robot-Manager 3.1 adds some really great enhancements to the product. Below is a summary of those additions.

### New Features and Benefits:

1. New **Spider Visits** and **Indexed Pages** reports. Both reports are in html format and printable from within the program.
2. New **Whois** dialog box that allows you to look up spiders by domain or ip address.
3. New **Known Spider** editor for entering known spiders that don't request the robots.txt file.
4. New **FTP Browser** for finding files and folders on ftp servers.
5. Improved **FTP Server** support for other types of operating systems. You can now browse reliably on Windows, Linux (Unix), and most Mac systems.
6. Improved **File Permissions** handling when browsing ftp servers. Access permission on Unix style servers are now correctly recognized.
7. Improved **Symbolic Link** handling when browsing ftp servers. Symbolic links to directories and files are now handled correctly.
8. Improved **Log File Parsing** for handling situations in which the log file format has changed or an entry is corrupted. Also made a slight improvement to the matching algorithm which may be noticable on some slower machines.
9. Added **More Spiders** to step one to be used when setting access permissions for spiders.
10. Increased the **Spider Limit** in the Trial edition from 20 to 100 visits per project file.

## 3 Software License

### LICENSE AGREEMENT FOR ROBOT-MANAGER SOFTWARE

IMPORTANT READ CAREFULLY: This agreement is a legal agreement between you (either an individual or a single entity) and Sophtware.com, Inc. (PH) for the Robot-Manager software identified above, which includes computer software and associated

media, and may include online or electronic documentation. By installing, copying, or otherwise using the software, you agree to be bound by the terms of this agreement. If you do not agree to the terms of this agreement, please do not install or otherwise use the software.

The software is protected by copyright laws and international copyright treaties, as well as other intellectual property laws and treaties. The software is licensed, not sold.

PH grants to you a non-exclusive, limited license to use the software, pursuant to the terms hereof. PH provides the license solely for end-user use. You may not reverse engineer, decompile, or disassemble the software. The software is licensed as a single product. Its component parts may not be separated for use on more than one computer. You may not rent or lease the software. You may make copies of the software and transfer your rights under this agreement, provided that each recipient of the software also agrees to the terms of this agreement.

Without prejudice to any other rights, PH may, in its sole discretion and at any time, terminate this license. In such event, you must destroy all copies of the software and all of its component parts.

All title and copyrights in and to the software (including but not limited to any images, photographs, animations, video, audio, music, text, and applets, incorporated into the software), and any copies of the software are owned by PH (or one or more of its licensors). The software is protected by copyright laws and international treaty provisions. Therefore, you must treat the software like any other copyrighted material except that you may make copies of the software, subject to the limitations set forth herein.

#### NO WARRANTY; LIMITATION OF LIABILITY

(a) THE SOFTWARE IS PROVIDED "AS IS," AND NEITHER PH NOR ANY OF ITS LICENSORS MAKES ANY EXPRESS OR IMPLIED REPRESENTATIONS OR WARRANTIES TO YOU REGARDING THE USABILITY, CONDITION OR OPERATION THEREOF. NEITHER PH NOR ANY OF ITS LICENSORS WARRANTS THAT THE SOFTWARE WILL MEET ANY PARTICULAR CRITERIA OF PERFORMANCE OR QUALITY. PH AND ITS LICENSORS EXPRESSLY DISCLAIM ALL IMPLIED WARRANTIES, INCLUDING WITHOUT LIMITATION, WARRANTIES OF MERCHANTABILITY, TITLE, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, COMPATIBILITY, SECURITY OR ACCURACY.

(b) YOUR USE OF THE SOFTWARE IS AT YOUR OWN RISK. YOU ASSUME FULL RESPONSIBILITY AND RISK OF LOSS RESULTING FROM YOUR DOWNLOADING, INSTALLING, AND/OR USE OF FILES OR OTHER MATERIAL. YOU AGREE THAT NEITHER PH NOR ANY OF ITS LICENSORS WILL BE LIABLE FOR DAMAGES (INCLUDING CONSEQUENTIAL OR SPECIAL DAMAGES) ARISING OUT OF YOUR USE OF OR INABILITY TO USE THE SOFTWARE, AND YOU HEREBY WAIVE ANY CLAIMS WITH RESPECT THERETO, WHETHER BASED ON CONTRACTUAL, TORT OR OTHER GROUNDS, EVEN IF PH (OR ANY SUCH LICENSOR) HAS BEEN ADVISED OF THE POSSIBILITY OF DAMAGES. THE ENTIRE LIABILITY OF PH AND ITS LICENSORS, AND YOUR EXCLUSIVE REMEDY WITH RESPECT TO THE USE OF THE SOFTWARE OR ANY BREACH OF THIS AGREEMENT, SHALL BE THE REPLACEMENT OF ANY SOFTWARE FOUND TO BE DEFECTIVE. YOU HEREBY RELEASE PH AND ITS LICENSORS FROM ANY AND ALL OBLIGATIONS, LIABILITIES AND CLAIMS IN EXCESS OF THIS LIMITATION. SOME JURISDICTIONS DO

NOT ALLOW IMPLIED WARRANTIES TO BE EXCLUDED OR MODIFIED OR LIABILITY TO BE LIMITED, SO NOT ALL OF THE ABOVE LIMITATIONS MAY APPLY TO YOU.

(c) NEITHER PH NOR ANY OF ITS LICENSORS SHALL BE HELD RESPONSIBLE IN ANY WAY OR BY ANY MEANS, EITHER DIRECTLY OR INDIRECTLY, FOR ANY COMMUNICATIONS DIFFICULTIES OUTSIDE OF ASI'S CONTROL WHICH COULD LEAD TO THE INTERRUPTION OF DATA DELIVERY SERVICE TO YOUR ELECTRONIC MAIL ADDRESS, PAGER, TELEPHONE OR ANY OTHER RECEIVING DEVICES OR THIRD-PARTY DATA STORAGE AND/OR DELIVERY SERVICES.

(d) IN PARTICULAR, YOU WILL NOT RELY ON ANY REPRESENTATION OR WARRANTY, EXPRESSED OR IMPLIED, MADE BY ANY PERSON OTHER THAN AN AUTHORIZED OFFICER OF PH, IN EVALUATING THE SOFTWARE..

If any part of this agreement is held invalid or unenforceable, that portion shall be construed to reflect the parties' original intent, and the remaining portions remain in full force and effect. The laws of the State of California, excluding its conflicts-of-law rules, govern this agreement and your use of the software, and you expressly agree that exclusive jurisdiction for any claim or dispute arising from the use of the software resides in the courts of the State of Georgia.

## 4 Purchasing

Purchasing Robot-Manager is fast and easy. You can quickly order online using a credit card via our secure order page. We accept VISA, Master Card, American Express, and Discover credit cards. No waiting! Unlock the full potential of your copy of Robot-Manager today.

To Order: <https://www.websitemanagementtools.com/buyonline/>

Sill have questions? Give us a call, we'd be more than happy to answer any questions you might have. We can even process your order for you over the phone. We can be reached at:

(877) 301-0232 toll-free  
(678) 714-9408 outside the US  
(678) 714-9426 fax  
[sales@websitemanagementtools.com](mailto:sales@websitemanagementtools.com)

Robot-Manager comes in both a Standard and Professional edition. You can always upgrade to the next higher edition from our server for just the difference in cost. All are compatible with Windows 95/98/NT/Me/2000/XP and can be completely uninstalled via the Windows Add/Remove Programs control panel if the program does not meet your needs.

Pricing for Robot-Manager is as follows.

### **Robot-Manager Standard Edition**

**\$49.95** US [More Info](#)

### **Robot-Manager Professional Edition**

**\$129.95** US [More Info](#)

## 5 Product Overview

### 5.1 Overview

Robot-Manager™ is an easy-to-learn-and-use, Microsoft© Windows™, software application which generates correctly formatted, industry standard, Robot exclusion files. As part of the Professional edition, Robot-Manager also tracks spider visits and pages indexed by search engine spiders on your web site.

#### **What is robots.txt?**

The file commonly referred to as robots.txt is a robot/spider exclusion file, which resides in the document root directory of web servers and is used to set permissions and control the actions of those robots/spiders. The absence of a robots.txt file on your web site means that spiders have free reign of your web site and can index anything and everything they want to.

With this file it is possible to direct those spiders to relevant parts of your web site, or keep them off totally. In the case of search engine spiders, you can direct them to the areas of your web site that you want indexed and keep them off of the areas that don't need to be indexed. For example, you may have an area on your web site that is just for customers, which spiders do not need to index. You can keep them out of this area using a robots.txt file.

It is also a little known fact that you can use a robots.txt file to point search engine spiders to optimized pages. This is known as cloaking and usually involves expensive software that needs to be ran on the server. It also requires that the cloaking software know the IP addresses for all of the spiders of a particular search engine. With a robots.txt file none of this is required. No software is required on the server, and you don't need to know the IP addresses of the spiders you are presenting optimized pages to.

You can also use the robots.txt file to send email harvesting programs away from your site. Have you ever received spam mail to support@yourdomain.com? If you have, then chances are an email harvesting spider visited your web site and extracted the email address from your support area. With a robots.txt file you can send those spiders away.

You can find out more information on the robots.txt standard at this address <http://www.robotstxt.org>.

#### **How Does Robot-Manager Help?**

Robot-Manager helps by enabling the user to rapidly and easily generate correctly formatted, industry standard, robots.txt files. Whether your web site is large and complex or small and informative, Robot-Manager will always generate a correctly formatted robots.txt file. And because this information is stored in your project, you can come back and make adjustments to your robots.txt file as often as you like.

Furthermore, the Professional edition of Robot-Manager allows the user to actually track spider visits and pages indexed on there web site. Robot-Manager accomplishes this by

analyzing your web site access log file. As long as you have a log file for your web site, Robot-Manager can mine the data from it. No more searching for keywords at each search engine to see if your pages have been indexed. Robot-Manager tells you exactly what pages the search engine spiders have indexed (or at least visited).

### **Should I Have Robot-Manager?**

The simple answer, yes! By promoting your web site to the major US and international Search Engines you are explicitly inviting spiders to roam your web site. These spiders are clever but not very discriminating. Designed to automatically visit entire web sites, examine web pages, extract interesting information, and follow links, Search Engine spiders transmit everything they find back for cataloging by the Search Engine's indexing software. It's not a well known fact that unless you specifically exclude Search Engine spiders from areas where you don't want them to go they will happily wander all over your site arbitrarily recording everything they find.

Although indexing everything might sound superficially attractive, as part of a web site promotion strategy it has a number of disadvantages.

- You don't want a Search Engine spider to visit areas where sensitive information might be stored.
- Indiscriminately indexing everything seriously dilutes the relevancy of your web sites overall theme and produces a sub-optimal rank in Search Engine listings.
- Allowing a Search Engine spider to index everything can even inadvertently lead to the perception by the Search Engine that your web site contains spam, this can lead to your site being blacklisted.
- For multi lingual web sites, it's imperative to focus English language robots onto the relevant pages and to direct robots from international Search Engines, who might be looking for Spanish, German or French language resources, to the localized content areas of your site.
- Some spiders "rapid fire" requests causing severe, server loading problems which can detract from your visitors browsing experience and ultimately cause loss of business.
- Some spiders only index a few pages from your web site, wouldn't you want a chance to tell those spiders which pages to index?

In summary, Robot-Manager can help you by providing a way to read in your web site directory and file structure from either a local machine or a remote server, auto-detecting between Unix and Windows servers. Then allow you to select those documents and directories which are to remain private and upload your newly created robots.txt file back to your web server. Furthermore, Robot-Manager offers critical information on exactly which pages on your web site search engine spiders have visits and subsequently indexed. Robot-Manager also contains a database of most user-agent names of all the major US and international Search Engine robots enabling precise and unambiguous permissions to be set either globally, in groups, or for individual robots. You can also edit this database to include any new spiders that you may have found as a result of tracking spider visits to your web site.



### **More Info**

[Product Features](#)

## 5.2 Features

Listed below is a breakdown of the feature differences between the Trial, Standard, and Professional editions of Robot-Manager.

### **Robot-Manager Trial Edition FREE**

- Contains all of the features of Professional, but limited to...
- A maximum of 5 spiders selected for any project.
- A maximum of 100 spider visits or pages indexed mined from a log file.

### **Robot-Manager Standard Edition**

- Creates correctly formatted, industry standard Robot Exclusion Files.
- Uploads `robots.txt` to your web server.
- Allows manual editing of the `robots.txt` file.
- Allows grouping of spiders with similar rules for smaller exclusion files.
- Downloads web site structures using FTP. Support Windows, Linux, and Mac servers.
- Contains a database of most major US and international spiders.
- Allows Adding, Editing, and Deleting of spiders.
- Can direct search engine spiders to optimized pages (similar to cloaking).
- Easy-to-learn-and-use step-by-step user interface with Quick Help.

### **Robot-Manager Professional Edition**

- Contains all of the same great features of the Standard edition.
- Mines web site access log files for spider and page visits.
- Two detailed reports, Spider Visits and Indexed Pages, that you can print.
- Displays spiders, pages and dates visited in a logical tree format.
- Allows ordering of the spider tree in 3 convenient formats.
- Highlights pages requested in error by the spiders.
- Downloads single or multiple access log files using FTP.
- Telnets to web servers to extract spider visits and pages indexed.
- Allows adding new spiders found while processing log files.
- Exports spider and page visits to both CSV (Comma Separated Values) and XML formats.
- Exports raw log entries of just spider visits.
- Exports all spider IP addresses so they can be used by cloaking software.
- Commandline options for running multiple projects unattended.
- Processes multiple log files at the same time.
- Includes "Whois" so you can investigate unknown spiders.
- Built in list of known spiders that don't request the robots.txt file, that you can edit.
- Support for compressed log files (both Windows and Linux).

### **More Info**

[Product Overview](#)

[How do I purchase Robot-Manager?](#)

## 6 Working with Projects

### 6.1 Overview

In 1996 it was reported that less than 6% of all web sites on the Internet take advantage of a robots.txt file. And I doubt to this date that that number has changed much. This means a lot of spiders out there have been indexing huge amounts of useless information. Most of us already know this. Go to a search engine like AltaVista® and do a search on **collectables**. Our search on AltaVista reveals over half a million (621,210) results! Of the first 10 results, I can report that four had a robots.txt file. This is up from last year, where only two sites had a robots.txt file.

This is where Robot-Manager™ comes to the rescue. Robot-Manager can quickly create industry standard robot exclusion files. A simple four-step process is all that is needed to create simple and very complex exclusion files and to subsequently track spider visits and pages indexed. Robot-Manager deskills this process making it available to everyone with a web site.

With a robots.txt file on your web site, you can direct where the spiders go to index information. Many spiders have limits on the amount of information they collect from a web site. If a spider hits your site, but is only going to index the first 10 pages it comes across, wouldn't you want to be able to tell that spider which pages to index? Of course! A robots.txt file will help you do this. Furthermore, if your site is multi lingual, you can direct country or language specific spiders to those areas of your site.

#### Creating a robots.txt File

1. When you start Robot-Manager for the first time, a new project is automatically created for you. If you have been experimenting around, simply click the **New** button on the toolbar to create a new project.
2. Decide how and what spiders you would like to direct on your site and select those spiders in [Step 1 - Spiders](#). Use the **All Spiders** item for handling exclusion items like cgi-bin and other common directories that all spiders should not visit.
3. Go to [Step 2 - Disallow](#) and select the Home Page of your web site. Robot-Manager allows you to gather the folder tree from your local machine, a machine over your local area network, or straight from your web site via FTP. A list of all Folders in that site will appear in the Folders list below and the root folder will be selected. The Disallow list will display all of the files and folders for the selected folder (the root in this case).
4. Next, group any spiders that will have similar rules. To do this simple drag-and-drop spiders in the list on top of one another to create a parent/child relationship. You can do this to group categories of spiders, or to group language specific spiders.
5. Select each spider in the Spider list one-by-one and set the disallow options. To exclude a spider completely from your site, click the check box above the Folders and Disallow lists. If you click this check box, the Disallow list becomes an Allow list. You can use this option to direct search engine spiders to optimized pages on your web site.
6. Once your disallow rules have been set, you can go to [Step 3 - robots.txt](#) and verify your exclusion file. If all is correct, click the **Upload robots.txt** button on the toolbar and upload your new robots exclusion file to your web site.

#### Tracking Spider Visits (Professional Only)

1. After you have uploaded your robots.txt file you will want to start tracking spider visits to your web site. For Robot-Manager to track spider visits, it needs to know where your access log file is for your web site.
2. From the main menu, select **Tools|Select Access Log Location** and tell Robot-Manager where your log file is located. Robot-Manager gives you three choices of how and where to get your log file. Choose the method most appropriate to your situation.
3. Once the location of your log file has been entered, Robot-Manager will ask if you would like to capture new spider visits now. Click **Yes**, and let Robot-Manager capture the spider visits.
4. If you have had any spider visits, they should now be displayed in the Spider visit details tree.

#### **Notes**

- If you are using a program like [GoLive®](#) or Microsoft [FrontPage®](#) to publish your web site, you can include your new robots.txt file in your web site project and upload it with your web site. If you are unsure on how to do this, click the GoLive or FrontPage link to find out more.

#### **More Info**

[Step 1 - Spiders](#)

[Step 2 - Disallow](#)

[Step 3 - robots.txt](#)

[Step 4 - Spider Visits](#)

AltaVista is a registered trademark of AltaVista Company.  
GoLive is a registered trademark of Adobe Systems Incorporated.  
FrontPage is a registered trademark of Microsoft Corporation.

## 6.2 Step 1

### 6.2.1 Spiders

The first step in creating your robots.txt file is to select the spiders. Your selection will depend a lot on what you want to accomplish. If your requirements are simple, you may only need to select **All Spiders** from the list. For more complicated web sites and requirements, you may want to select additional spiders. Ask your self these few simple questions for determining what spiders your should select for your project.

- What spiders would I like to keep off of my site?
- Does my web site contain multi lingual resources?
- Do I have secure areas that should not be indexed?
- Am I trying to present optimized pages to search engine spiders?
- Do I have pages, like privacy statements, that don't need to be indexed?

The answers to the above question will help get you started in the right direction for making logical choices as to which spiders to include in your project. For instance, if you would like to keep email harvesting spiders off your site, you would select those spiders and disallow access to the entire site. If you are presenting optimized pages to specific search engine spiders, then you would disallow access to the entire site and set allow

rules, pointing those spiders to your optimized pages.

See the How Do I section later in this help file for more examples on selecting spiders for your robots.txt exclusion file.

#### **Caution**

- This step is used to select the spiders that will be included in the **robots.txt** file. Once you have selected the spiders to include, it is not necessary to go back to this step.

#### **Notes**

- This step is used to select which spiders to disallow to certain sections of your web site. If a spider is not disallowed from any section of your web site, don't include it in your selection.
- If you are setting disallow rules for all spiders, select the **All Spiders** item. By default, this spider is selected.
- Some spiders will only visit your site if you request it. These spiders can be safely left out of the selection until you request a visit.
- Right-clicking the spider list will present you with a shortcut menu for common tasks.

#### **More Info**

[Project Overview](#)

[Selecting Spiders](#)

[Modifying the Spider List](#)

[Select By](#)

[Add/Edit Spider](#)

## 6.2.2 Selecting Spiders

Robot-Manager makes selecting spiders easy. There are many different ways to select individual and groups of spiders. Outlined here are those ways.

### **Select All**

To check all of the spiders in the list, select **Edit|Select All** from the main menu. This will place a check next to each spider in the list. You can also use the keyboard shortcut **Ctrl+A**. If you already have a selection of spiders, this will erase that selection. There is no undo for Select All.

### **Invert Selection**

You can select one or more spiders using the standard Windows list selection options.

**Ctrl+left button** click adds spiders to your current selection, while **Shift+left button** selects all items between this and the last selection. Once the spiders are selected, use **Edit|Invert Selection** from the main menu to toggle the checked state of all selected item. You can also use the keyboard shortcut **Ctrl+I**. If a spider was checked, it will be unchecked. If a spider was unchecked, it will be checked.

### **Select By**

To check spiders based on certain criteria, use **Edit|Select By** from the main menu. You can also use the keyboard shortcut **Ctrl+B**. The dialog box allows you to group check spiders based on category, country, and language. See [Select By](#) for more information.

### Sorting By Column

Each column header in the list can be clicked to sort the spider list by that column either descending or ascending. The default sort is ascending on the spider column. Furthermore, the Spider column has an **additional sort** by checked spiders. This column can be clicked three times. The **third click** sorts all checked spiders to the top of the list. This helps when you would like to see what spiders you have checked.

### Spider Description

To help aid in the selection of spiders, each spider has a description with notes about that particular spider. Click a spider in the list to see its description. Many descriptions have links to the web sites of the owners of the spiders. Click those links to potentially find more out about the spider.

### Status bar

The status bar displays the total number of spiders that are available and the number of spiders you currently have selected.

### Notes

- Selecting or highlighting a spider does not change the check state of the spider. You must click the box to the left to check and uncheck the spider.
- The **spacebar** toggles the checked state of the selected or highlighted spider.
- Right-clicking will present you with a shortcut menu for common tasks.

### More Info

[Project Overview](#)

[Spiders](#)

[Modifying the Spider List](#)

[Select By](#)

[Add/Edit Spider](#)

## 6.2.3 Modifying the Spider List

Robot-Manager allows you to add, edit, and delete spiders from Step 1. This comes in handy when using the Professional version to track spider visits to your web site. When new spiders are encountered, you can simply add them, set rules for them, and republish your robots.txt file to your web site.

### Add/Edit Spider

Use this option to add or edit a spider in the spider list. All spider names must be unique and Robot-Manager will not allow you to add a spider that is the same name as a spider that already exists in the list. The Add Spider dialog box will indicate a duplicate name while typing by changing the background of the edit control to red and the text to white and disabling the **OK** button. If you accidentally edit a spider, you can undo the last edit using **Edit|Undo** from the main menu. Once another undo action is performed, you will not be able to undo this action.

### Delete

Use this option to remove a spider that is no longer operational. If the removed spider is checked, it will be removed from Step 2. Furthermore, if this spider is a parent to any spider in step 2, then those spiders will be promoted from children and retain their original disallow options—not the options of the parent spider that was just removed.

### ⚠ Caution

- Modifications to the spider list are stored in `spiders.xml` in the program directory. If you modify the spider list and reinstall Robot-Manager, be sure to answer no to the question that asks you to overwrite this file. Otherwise, your changes will be lost!

### 📁 More Info

[Project Overview](#)

[Spiders](#)

[Selecting Spiders](#)

[Select By](#)

[Add/Edit Spider](#)

## 6.3 Step 2

### 6.3.1 Disallow

The disallow step is where you set the rules for your `robots.txt` file. This is what directs the spiders to certain areas of your web site and keeps them off others. Disallow provides a convenient GUI interface for creating your `robots.txt` file. This step is divided into two sections. On the left is the list of spiders you have chosen for your `robots.txt` file, and on the right is where you set the disallow/allow options for each spider you have chosen.

### 💠 Important

- The first thing to do in this step is to select your home page location. Until you select where your home page is located, no options are available. To enter your home page location, select **Tools|Select Home Page Location** from the main menu. Once your home page is selected, you will be asked if you would like to capture the folder tree now, select **Yes**. You should see the directory tree of your web site in the Folder list with the root folder selected (the folder that contains your home page). You should also see the folders and files of the selected directory in the Disallow list. This is where you disallow or allow access to certain parts of your web site.

Now you are ready to started setting disallow/allow rules for your selected spiders. Select a spider to sets its rules.

### 📝 Notes

- Once a directory has been disallowed, there is no need to set disallow options for any files or directories below it.
- If you have set Disallow options for a spider and then check the disallow access to the entire site check box, your file and directory selections are preserved and become Allow options.
- Right-clicking will present you with a shortcut menu for common tasks.

### 📁 More Info

[Grouping Spiders](#)

[Setting Disallow Options](#)

[Setting Allow Options](#)

[Select Home Page Location](#)

### 6.3.2 Grouping Spiders

To help make setting spider rules easier, you can group spiders with common rules together. This also keeps the size of your robots.txt file down, which is important for spiders that only read robots.txt files up to a certain size.

#### **Drag-and-Drop**

Use your mouse to drag-and-drop spiders on one another, creating parent/child relationships. It does not matter what spider you choose as the parent, as long as the rules for the children should be the same as the parent.

#### **What Should I Group?**

The best way to explain what spiders should be grouped is by example. Following here are some common situations where you would want to group certain spiders together.

##### **Grouping by Category**

Many users will want to stop email harvesting spiders from visit their site and gathering email addresses for spamming campaigns. To stop these spiders, you would select all of the known email spiders in step 1, then group them in this step and set the rules for the parent spider to disallow access to the entire site.

##### **Grouping by Language/Country**

Web sites that are multi lingual will want to group country and/or language specific spiders to point them at the relevant language resources on their web site. If you have both english and German resources, then you would select the German spiders in step 1, then group them in this step and point them at the German resources on your web site.

##### **Grouping by Content Indexed**

There are spiders out their that index only certain types of content (e.g., images, sounds, music, news, etc...). You can group these spiders together and point them to the relevant resources on your web site. For instance, Robot-Manager includes some News collecting spiders that look for news related information. If you have a News or Press Release section on your web site, you would want to point them to these resources.

##### **Grouping for Disallow**

If there is a list of specific spiders that you do not want visiting your web site, you can group these together and disallow access for them all.

#### **Caution**

- Although you can drag-and-drop spiders onto the All Spiders spider, it has no affect except to increase the size of your resulting robots.txt file. Any spiders with the same rules as the All Spiders spider should be left out of your selection in Step 1.

#### **Notes**

- All children of a parent spider inherit the rules of the parent spider. Selecting a child and setting any rules sets the rules for the parent, which will apply to the other children as well.

#### **More Info**

[Disallow](#)  
[Setting Disallow Options](#)  
[Setting Allow Options](#)

### 6.3.3 Setting Disallow Options

Setting disallow rules tells spiders not to visit the disallowed resource. The resource can be a file or a directory. When directories are selected, everything under the directory is disallowed as well. To set disallow options, you need to select the spider in the list first. You will be setting the disallow rules for the selected spider. The selected spider is indicated by a check icon to the left.

Once a spider is selected, you can set all of the disallow options for it. To disallow access to the entire site for the selected spider, use the check box above the Folder and Disallow lists. To disallow access to a directory or file resource, place a check next to the item in the Disallow list. You can verify the contents of the robots.txt file at any time by changing to [Step 3 - robots.txt](#) and examining the file and then returning to this step.

Use the Folder tree to navigate your web site. Alternately, you can double-click a folder in the Disallow list to "drill-down" to that folder.



#### More Info

[Disallow](#)  
[Grouping Spiders](#)  
[Setting Allow Options](#)

### 6.3.4 Setting Allow Options

Setting allow rules tells spiders what they can visit on a web site. This is basically the opposite of using disallow rules. Its purpose is two fold. If your web site contains mostly disallow rules, you may benefit by setting allow rules. This would keep the size of your robots.txt file down. If you are presenting optimized pages to search engines, you can tell them where those pages are at.

Allow rules work by disallowing access to the entire site, followed by rules to allow access to certain areas. Because you are specifically telling spiders what they can visit, you can point them to pages that would not normally be available from spidering your web site. This presents the opportunity to direct search engine spiders to optimized pages on your web site. See [How Do I Present optimized pages to search engines?](#) later in this help file for specific information.

#### ⚠ Caution

- Allow rules are a revision of the original robots.txt spec. Some spiders may not obey, or be able to read, this new feature.



#### More Info

[Disallow](#)  
[Grouping Spiders](#)

[Setting Disallow Options](#)

## 6.4 Step 3

### 6.4.1 robots.txt

This is the last step in creating your robots.txt file for your web site. There is really nothing to do in this step but verify your selections from the previous step and upload or export the robots.txt file.

#### **Upload robots.txt**

To upload robots.txt to your web server, select **Tools|Upload robots.txt** from the main menu. The Upload robots.txt dialog box will display. This is where you enter your FTP server information so Robot-Manager can send the robots.txt file to your server. If you are using FTP to gather your folders and files from the web server, this dialog box will default those FTP settings.

#### **Export robots.txt**

If you make any edits, or just want to save the file to a local area network drive, select **File|Export|robots.txt** from the main menu. A dialog box will display asking where to save the robots.txt file.

#### **Notes**

- If you make any edits to robots.txt, you can only export it to your local drive. You will then need to use an FTP program to send it up to your server.

#### **More Info**

[Upload robots.txt](#)

[Export robots.txt](#)

## 6.5 Step 4 (Professional Only )

### 6.5.1 Spider Visits

This is where you come to track spider visits to your web site. For any web site interested in gaining top position in the major search engines, this is a critical step. You will need access to your web site access log file. If your hosting provider does not give you access to your log file, you might want to consider a hosting package that includes this access. If they don't provide a package with log file access, you may also want to consider moving your web site to a hosting provider that gives access to your log file.

#### **Important**

- The first thing to do in this step is to select your access log location. Until you select where your access log is located, no options are available. To enter your access log location, select **Tools|Select Access Log Location** from the menu. Once your log file is selected, you will be asked if you would like to capture new spider visits, select **Yes**. Depending on the size and location of your log file, this step may take a minute or two to complete.

Once your log file has been processed, and you had spider visits detected, you should see one or more spiders in the spider visit details tree. If you don't actually see the name of the spider, but instead have the IP address, you may need to adjust the

[log file format](#) Robot-Manager detected. It is also possible that you are running on an older web server that doesn't support the format required to extract the name. There are two situation where this can happen:

- You are running an older version of IIS, which doesn't support the User-Agent field required to extract the spider name. Or,
- You are using a custom log file format that is not publishing the User-Agent information.

If either of these cases apply to you, you're in luck because the problem can be remedied. For older versions of IIS, simply upgrade to the latest version. The latest version supports the correct format by default. If you are using a custom log format, simply update the format to include the User-Agent field. Furthermore, if your hosting provider controls the format of your log file, ask them to change it. All modern web servers have a default log format that Robot-Manager can read, which contains all of the required fields.



#### **More Info**

[Select Access Log Location](#)

[Log File Format](#)

[Add Spider to Step 1](#)

[Order Spider Visits By](#)

[Export Spider Visits](#)

[Export Spider Logs](#)

[Export Spider IP Addresses](#)

## 7 Reference

### 7.1 Command Line Options

Robot-Manager can be fully automated with the use of command line options. Single and multiple projects can be ran from the same command line. Used in conjunction with Windows Schedule Tasks program, you can run one or more projects off-hours.

#### **General Options**

There are a couple options not associated with the scheduling and running of project files. They are the following.

##### **`/unregister`**

Use this option to unregister the file extension for Robot-Manager. This is typically used by the uninstall program.

##### **`/no_splash`**

Specify this option when you don't want the splash screen showing when you start Robot-Manager. (This does not affect the program startup time).

#### **Scheduling Options**

These options are used typically in conjunction with the Windows Schedule Tasks program. They are used to automate the running of one or more project files. When using these options, Robot-Manager will run as an icon in the Task tray on the Windows Start bar. The /addr, /log, and /visit can all be used at the same time to export more than one file. Items listed between '[' and ']' are optional, and items between '<' and '>' are required.

**/auto <project file>**

Use this option with a project file to run that project unattended.

**/addr [<export file>] [/auto] <project file>**

Use this option to export the Spider IP Addresses after the project finishes running. If you do not specify an export file name, Robot-Manager will create one for you based on the project name like this: "<project file>.addr.csv".

**/log [<export file>] [/auto] <project file>**

Use this option to export the raw access logs that make up the spider visits to your site. If you do not specify an export file name, Robot-Manager will create one for you based on the project name like this: "<project file>.log.csv".

**/visit [<export file>] [/auto] <project file>**

Use this option to export the spider visit information. If you do not specify an export file name, Robot-Manager will create one for you based on the project name like this: "<project file>.visit.csv".

**/list <list file>**

Use this option to specify more than one project file to run. The <list file> is a file that contains a list of project files, one per line, top run. A full path to each project file should be specified, otherwise Robot-Manager may not know where the file is located. When using this option, the <project file> option is not used.

**⚠ Caution**

- If you exclude the export file name, you will need to include the /auto option before the project file name.

**Running a Single Project**

Here are some examples of running single projects to help get you started.

**Extract new spider information:**

In this example, Robot-Manager starts as an icon in the Task tray, then opens the project file specified on the command line. Next, a Refresh in Step 4 - Spider Visits is performed and the resulting data is saved back to the project. Robot-Manager then exits.

```
RobotManager.exe /auto "c:\My Documents\My Projects\projectfile.rbt"
```

**Extract new spider information and export spider IP addresses:**

In this example, Robot-Manager starts as an icon in the Task tray, then opens the project file specified on the command line. Next, a Refresh in Step 4 - Spider Visits is performed and the resulting data is saved back to the project file. Then Robot-

Manager exports the spider IP addresses to the export file specified and exits. Notice that the "quotes" were not needed around the export file. This is because the path and file name did not contain any spaces. Also, in this case, the /auto option is optional.

```
RobotManager.exe /addr d:\customers\wmt\ips.csv /auto "c:\My Documents\My Projects\wmt.rbt"
```

#### **Extract new spider information and export all available data:**

In this example, Robot-Manager starts as an icon in the Task tray, then opens the project file specified on the command line. Next, a Refresh in Step 4 - Spider Visits is performed and the resulting data is saved back to the project file. Then Robot-Manager exports all of the available data back to the project file's directory. Notice that in this case the /auto option is required. Otherwise, Robot-Manager would mistake the project file for the visit export file.

```
RobotManager.exe /addr /log /visit /auto "c:\My Documents\My Projects\wmt.rbt"
```

#### **Running Multiple Projects**

When running multiple projects from the command line, all of the same options apply above. The only difference is that the project file name is replaced with the /list option and the /auto option is always optional. Furthermore, if you specify a file name for any export options, that file will contain the data exported from all the project files that were ran. In other words, the export file specified is appended to for each project that is run. Here are some sample command lines for running multiple projects.

```
RobotManager.exe /list c:\customers\prjlst.txt
```

```
RobotManager.exe /list "c:\My Documents\My Projects\all_projects.txt"
```

```
RobotManager.exe /addr d:\customers\all_spider_ips.csv /list d:\customers\prjlst.txt
```

Your list file should contain one project file per line. Quotes are not needed around file names and paths that contain spaces. And, there should be no blank lines in the file. All file names must start at the first line and continue down the file. There is no limit on the number of files that can be ran from a list file. Here is an example file.

```
c:\My Documents\My Projects\wmt.rbt  
c:\My Documents\My Projects\sample.rbt  
c:\My Documents\My Projects\websitemanagementtools.rbt  
d:\customers\yahoo.rbt  
d:\customers\overture.rbt
```

#### **Notes**

- To stop Robot-Manager while running in the Task tray, double-click the robot icon to bring the program up and close it.
- Hold your mouse cursor over the robot icon in the Task tray to see the name of the file currently being worked on.

## 7.2 Select By

The **Select By** dialog box helps select spiders based on certain criteria. Currently you can select groups of spiders based on their category, country, and supported language. For each criteria you can select multiple entries, allowing complete flexibility.

### Using the Selection Lists

1. In an item list, click the drop-down arrow to the right of the edit field to display the list of items.
2. For each item you would like to include in the selection, place a check in the box to the left of the item.
3. When you are finished making your selection, click the drop-down arrow again (or anywhere outside the selection list).

### Selection Options

Selection Options present two different ways of modifying the current selection of spiders. The default option erases the current selection and replaces it with a selection matching your selection criteria. The second option leaves your current selection, plus selects any spiders matching your selection criteria.

#### Example:

Select all search engine spiders from the United States, plus any email harvesting spiders.

1. In the Category list, click the drop-down arrow and place a check next to **Email** and **Search**.
2. In the Country list, click the drop-down arrow and place a check next to **United States**.
3. Click **OK** to accept the selection.

#### Notes

- The language shown for each spider is the **base** language that is displayed on the home page (if a search engine) when viewed from the United States on a web browser (unless otherwise noted).
- Clicking **OK** with no selection, clears the selection on all spiders.
- Each selection list displays the items that are available in the spider list. So, for instance, the Country selection list will only display the list of countries that are in the spiders list.

## 7.3 Add/Edit Spider

The **Add/Edit Spider** dialog box allows you to maintain the spider list by adding your own spiders or editing existing ones. Although Robot-Manager comes with a fairly complete list of known spiders, there are still many more out there to be found. When new spiders are found, they will be added to the program. But many times, you may find a spider that isn't included. If you do find a new spider, it is very easy to add it to the list.

### Adding Spiders

- While in Step 1 - Spiders, select **Edit|Add Spider** from the menu.
- While in Step 4 - Spider Visits, select the new spider in the tree, then select

**Tools|Add Spider to Step 1** from the menu.

### Data Fields

- **Name.** This is a unique name for the spider. It can be the User-Agent name, or one that you make up.
- **User-Agent.** This is the user-agent text that the spider will look for in your robots.txt file to identify the rules it should follow. This is usually an abbreviated version from the User-Agent field in your log file. For instance, the spider **FAST-WebCrawler** will recognize the full name and the abbreviated form **fast**. You may have to do a little research to determine if a spider has an abbreviated form. Furthermore, the robots.txt specification states that spiders should look at this field case insensitive. So both upper- and lower-case combination should work.
- **Category.** This allows you to classify each spider that you add. Spiders are categorized to make the selection process easier.
- **Country.** This is typically the country where the spider originates. Again, country is used to make the selection process easier. It also gives user with multi lingual sites an opportunity to select spiders based on country and/or language.
- **Language.** This is typically the language of the resources this spider is looking for. Some search engines spiders index more than one language resource. Acoon, a German search engine spider, is an example of a spider that indexes both English and German pages. Care should be taken when directing these types of spiders.
- **Description.** This can be any text that you want. The motivation here is to give enough information so you can make a logical choice when selecting spiders for your project. This field supports a [mini version of HTML](#).

### ⚠ Caution

- If you add or edit spiders in the list, be sure to answer **no** to the question that asks you to overwrite `spiders.xml` when reinstalling Robot-Manager. Otherwise, you will lose your spider changes.

### 📝 Notes

- Each spider added to the program must have a unique name. Robot-Manager will clue you when you have entered a name that is already in the list.

### 📁 More Info

[Step 1 - Spiders](#)  
[Mini HTML](#)

## 7.4 Mini HTML

### **B : Bold tag**

`<B>` start bold text  
`</B>` end bold text

### **Example**

This is a `<B>test</B>`

### **U : Underline tag**

`<U>` start underlined text  
`</U>` end underlined text

**Example**

This is a `<U>test</U>`

**I : Italic tag**

`<I>` start italic text

`</I>` end italic text

**Example**

This is a `<I>test</I>`

**S : Strikeout tag**

`<S>` start strike-through text

`</S>` end strike-through text

**Example**

This is a `<S>test</S>`

**A : Anchor tag**

- `<A href="value">` the text after the tag is an anchor and should be in the format of "http://www.mydomain.com".

- `</A>` end of anchor

**Examples**

This is a `<A href="mailto:myemail@mail.com">test</A>`

This is a `<A href="http://www.tmssoftware.com">test</A>`

This is a `<A href="somevalue">test</A>`

**FONT : Font specifier tag**

`<FONT face='facevalue' size='sizevalue' color='colorvalue' bgcolor='colorvalue'>` specifies font of text after tag.

- `face` name of the font
- `size` HTML style size if smaller than 5, otherwise point size of the font
- `color` font color in hexadecimal
- `bgcolor` background color in hexadecimal

`</FONT>` ends font setting

**Examples**

This is a `<FONT face="Arial" size="12" color="clred">test</FONT>`

This is a `<FONT face="Arial" size="12" color="#FF0000">test</FONT>`

**P : Paragraph**

`<P align="alignvalue" [bgcolor="colorvalue"]>` starts a new paragraph, with left, right or center alignment. The paragraph background color is set by the optional bgcolor parameter.

`</P>` end of paragraph.

**Examples**

`<P align="right">This is a test</P>`

`<P align="center">This is a test</P>`

`<P align="left" bgcolor="#ff0000">This has a red background</P>`

`<P align="right" bgcolor="clYellow">This has a yellow background</P>`

**HR : Horizontal line**

`<HR>` inserts line break with horizontal line

**BR : Line Break**

`<BR>` inserts a line break

**BODY : Body color / background specifier**

`<BODY bgcolor="colorvalue" background="imagefile specifier">` sets the background color of the HTML text or the background bitmap file

**Example**

`<BODY bgcolor="clYellow">` sets background color to yellow

`<BODY background="file://c:\test.bmp">` sets tiled background to file test.bmp

**IND : Indent tag**

`<IND x="indent">` indents with "indent" pixels

**Example**

This will be `<IND x="75">`indented 75 pixels.

**SUB : Subscript tag**

`<SUB>` start subscript text

`</SUB>` end subscript text

**Example**

`<SUP>9</SUP>/<SUB>16</SUB>`

**SUP : Superscript tag**

`<SUP>` start superscript text

`</SUP>` end superscript text

**UL : List tag**

`<UL>` start unordered list tag

`</UL>` end unordered list

**Example**

`<UL>`

`<LI>`List item 1

`<LI>`List item 2

`</UL>`

**LI : List item**

`<LI>` new list item

**SHAD : Text with shadow**

`<SHAD>` start text with shadow

`</SHAD>` end text with shadow

**Special characters**

`&lt;` less than <

`&gt;` greater than >

`&amp;` ampersand &

&quot; double quote mark "

## 7.5 Select Home Page Location

Robot-Manager can gather the folder tree of your web site in two different ways. The method you use will depend on your situation.

### Local Network

If you publish your web site from your local machine, choose this option. Robot-Manager will gather the folder and file list from your local machine. This is the quickest option. Use the **Browse** button to locate your home page folder.

### FTP Server

If you maintain your web site from the server, or don't have access to the published web site, choose this method. Robot-Manager will download the folder and file list from your server using FTP. Once you have entered your FTP settings, you can browse the remote machine for the home page folder.

### Notes

- If you move your web site to a new server, all is not lost. You can change your home page location and Robot-Manager will attempt to keep all of your prior settings. If the directory structure is the same, nothing will be lost.

## 7.6 Refresh

Use Refresh in Step 2 to update the folder and file list when it changes. In Step 4, use Refresh to capture any new spider visits from your log file.

## 7.7 Upload robots.txt

When you are satisfied with your exclusion selections, you can upload the generated robots.txt file to your web server. By default, Robot-Manager uses the FTP information set in Home Page Location, but this can be changed. If you are not gathering your folder tree from your web server, you will need to enter your FTP settings first.

## 7.8 Export robots.txt

If you make any manual edits to the robots.txt file in Step 3 and want to save them, you need to export the file to your local machine. Robot-Manager does not support uploading a manually edited robots.txt file at this time. To export your changes, select **File|Export|robots.txt** from the menu.

## 7.9 Select Access Log Location

Robot-Manager can capture the data from your log file in three different ways. The method you use will depend on your situation and connection speed. In all capture methods the full path and file name of the log file must be supplied.

### Local Network

This is the first, and preferred method of capture. If you are hosting your web site

locally and have access to the log file over your local area network, choose this capture method. Robot-Manager will capture the spider visits directly from this file. If your log file is compressed, select which compression method matches the server type you are using.

### FTP Server

This is the slowest method of capture, but must be used on Windows systems running IIS, or Unix/Linux servers where you don't have Telnet (Shell) access and your web server is not being hosted locally. Using FTP, Robot-Manager downloads your current log file and then processes it locally. Using FTP Server method, Robot-Manager can also work with multiple log files. If your log file is archived daily to a separate directory, you can use wildcards in the file name to have Robot-Manager gather all of the files for processing. FTP Server method also supports compressed log files.

### Telnet Capture

This is usually the fastest method of capture. Currently, Telnet Capture will only work on Unix/Linux servers that give access to 'grep' and 'gunzip' in the shell. All of the processing of the log file happens on the server. Only the raw spider visit data is transferred back to the local machine. Telnet capture will work on compressed log files too.

### ⚠ Caution

- At this time Telnet capture does not support Windows systems running IIS.
- Robot-Manager uses a standard Telnet shell and does not support SSH. If you need a secure connection, please use the FTP Server method instead.

### 📝 Notes

- You can change the location and file name of your log file as often as you like, as long as the log file format remains the same.
- The '\*' and '?' wildcard characters can be used in the file name for the FTP Server method above.

## 7.10 Log File Format

Although Robot-Manager can auto detect most log file formats, it may be necessary to adjust the detected format for some log files. Log File Format gives you the option to adjust that format. In most cases of an incorrectly detected log file, Robot-Manager does not identify the User-Agent field and defaults to the Remote IP. Change the index and type of the Spider Name to correct this.

### Sample

The sample list displays the first log entry from the spider visits found. Use the double-down arrow to move to the next log entry if the current entry does not display enough information to determine the fields. Each field in the list is numbered to indicate its position in the log entry.

### Data Fields

- **Date**. This is the index of the item in the log entry that is the date field.
- **Date Type**. This is the type of date field in the log. Robot-Manager supports four types, which can be '-' or '/' separated.
  - Apache (standard) — 03/Sep/2001:18:49:03 -0400

- IIS (standard) — 2001/09/03
  - U.S. (m/d/yy) — 9/3/01 or 09/03/2001
  - Euro (d/m/yy) — 3/9/01 or 03/09/2001
- **Spider Name.** This is the index of the item in the log entry that is the name of the spider.
  - **Spider Type.** This is the type of the spider name. Robot-Manager supports three types of entries that identify the spider.
    - Remote IP — 216.64.94.102
    - Remote Host — http://www.host.com or any other valid host name
    - User-Agent — Spider/1.0 (info@spider.com; http://www.spider.com) or any text
  - **Page Request.** This is the index of the item in the log entry that is the page request field.
  - **Page Request Type.** This is the type of page request. Robot-Manager supports two types of page request.
    - Request Line — GET /index.html HTTP/1.0
    - Filename — /index.html
  - **Spider IP.** This is the index of the item in the log entry that identifies the spider's IP address.
  - **Error Code.** This is the index of the item in the log entry that is the error code for the page requested.

#### **Caution**

- If your log file format changes, you might need to create a new project.
- If you simply add one or more fields to your log file, Robot-Manager will still work with your logs.

#### **Notes**

- If your log file format is detected correctly, you do not need to change any setting in here.

## 7.11 Add Spider to Step 1

When you find new spiders that are not included in Robot-Manager and you would like to be able to set rules for that spider, use the Add Spider to Step 1 option. When you use this option, you must first select the spider you would like to add. Once the spider is selected, select **Tools|Add Spider to Step 1** from the menu. Add Spider will display with most of the fields filled in for you already. You should give the spider a unique name and update the description to reflect the spiders purpose or origin.

## 7.12 Order Spider Visits By

When viewing the spider visit data, it may help to see it in different orders. Robot-Manager supports three ways of ordering the spider tree.

- **By Date.** Which displays the Date followed by the Spider name and then the Page requested.

- **By Spider.** Which displays the Spider name followed by the Page requested and then the Date.
- **By Page.** Which displays the Page requested followed by the Spider name and then the Date.

In all cases, the Date information is always displayed in descending order. This means that newer requests always sort to the top of the tree. In all sort orders, the last nodes on the tree can be selected to display the raw log entry that generated the spider visit.

When displaying the tree in **By Date** order, you can mouse over the page to see the error code in the local status bar. If a page request generated any error code other than a "200 OK", Robot-Manager will mark that page with a red X. This indicates a potential problem with that page or the server when that page was requested.

#### Notes

- By Date is the only ordering of the tree that will indicate error page requests.
- The local status bar displays the total count of the order by item.

## 7.13 Export Spider Visits

If you would like to look at the spider visit data in other ways that Robot-Manager does not allow, use the Export Spider Visits option. You can export the data in either comma separated values (CSV) or XML format.

#### Example CSV Format

Below is an example of the CSV format export. The first column is the robot's name (if name isn't available, then the IP address is used instead), the page requested, and the date/time of the request (Microsoft IIS users will not have the time in this field).

```
"ia_archiver", "/robots.txt", "09/03/01 18:49:37"
"ia_archiver", "/products/topdog/", "09/03/01 20:37:19"
"ia_archiver", "/legal/", "09/03/01 05:27:44"
"ia_archiver", "/robots.txt", "09/03/01 05:27:44"
"ia_archiver", "/robots.txt", "09/26/01 14:40:03"
"ia_archiver", "/index.php", "09/01/01 11:02:48"
"ia_archiver", "/news/", "09/01/01 16:37:36"
"ia_archiver", "/robots.txt", "09/03/01 05:18:16"
"ia_archiver", "/", "09/25/01 09:17:17"
"Scooter-W3-1.0", "/", "09/03/01 23:39:38"
"Scooter-W3-1.0", "/robots.txt", "09/03/01 23:39:38"
"Openfind", "/robots.txt", "09/23/01 15:53:45"
"Googlebot/2.1", "/robots.txt", "09/02/01 18:19:56"
"Googlebot/2.1", "/products/", "09/02/01 18:19:58"
"Googlebot/2.1", "/company/", "09/03/01 14:27:36"
"Googlebot/2.1", "/robots.txt", "09/07/01 03:56:12"
"Googlebot/2.1", "/products/", "09/07/01 08:45:45"
"Googlebot/2.1", "/products/robot_manager/", "09/13/01 17:59:10"
"Googlebot/2.1", "/robots.txt", "09/13/01 17:59:10"
"Googlebot/2.1", "/", "09/07/01 00:38:56"
"Googlebot/2.1", "/", "09/27/01 02:54:33"
"Googlebot/2.1", "/index.php", "09/03/01 01:58:10"
"Googlebot/2.1", "/news/articles/wp20010610.txt", "09/08/01 23:18:05"
"Googlebot/2.1", "/news/", "09/27/01 10:14:54"
"Googlebot/2.1", "/news/articles/wp20010613.pdf", "09/10/01 12:03:21"
```

```
"Googlebot/2.1", "/news/articles/wp20010610.doc", "09/10/01 10:35:34"
"Googlebot/2.1", "/robots.txt", "09/10/01 10:35:34"
"Googlebot/2.1", "/awards/", "09/03/01 20:08:20"
```

### Example XML Format

Below is an example of the XML format export. The thing to note here is that the name tag has an extra attribute named type. Type can be any one of the following values: user-agent, remote-host, or remote-ip. The value depends on your log file format.

```
<?xml version='1.0' encoding='us-ascii' ?>
<!-- asi3.xml -->
<spider_visits>
  <spider_visit>
    <name type="user-agent">ia_archiver</name>
    <page>/robots.txt</page>
    <date>09/03/01 18:49:37</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">ia_archiver</name>
    <page>/products/topdog/</page>
    <date>09/03/01 20:37:19</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">Scooter-W3-1.0</name>
    <page>/</page>
    <date>09/03/01 23:39:38</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">Scooter-W3-1.0</name>
    <page>/robots.txt</page>
    <date>09/03/01 23:39:38</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">Openfind</name>
    <page>/robots.txt</page>
    <date>09/23/01 15:53:45</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">Googlebot/2.1</name>
    <page>/robots.txt</page>
    <date>09/02/01 18:19:56</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">Googlebot/2.1</name>
    <page>/products/</page>
    <date>09/02/01 18:19:58</date>
  </spider_visit>
  <spider_visit>
    <name type="user-agent">Googlebot/2.1</name>
    <page>/products/robot_manager/</page>
    <date>09/13/01 17:59:10</date>
  </spider_visit>
</spider_visits>
```

## 7.14 Export Spider Logs

Use Export Spider Logs to archive the raw log data that generated the spider visits. The format of this file is exactly the same as your access log file format.

## 7.15 Export Spider IP Addresses

If you need to extract the IP addresses for all of the spiders that have visited your site, select **File|Export|Spider IP Addresses** from the menu. Robot-Manager will export all of the spider names and IP addresses to a comma separated values list. Spiders with more than one IP address are included as well. By default the name `spdr_ips.csv` is given to the file.

### Example Output

```
211.100.25.200, "BaiDuSpider"
66.77.74.243, "FAST-WebCrawler/2.2.10"
66.77.74.160, "FAST-WebCrawler/2.2.11"
66.77.74.160, "FAST-WebCrawler/2.2.7"
66.77.74.160, "FAST-WebCrawler/2.2.8"
216.239.46.100, "Googlebot/2.1"
216.239.46.102, "Googlebot/2.1"
216.239.46.105, "Googlebot/2.1"
216.239.46.112, "Googlebot/2.1"
216.239.46.12, "Googlebot/2.1"
216.239.46.124, "Googlebot/2.1"
216.239.46.129, "Googlebot/2.1"
216.239.46.130, "Googlebot/2.1"
216.239.46.148, "Googlebot/2.1"
216.239.46.19, "Googlebot/2.1"
216.239.46.20, "Googlebot/2.1"
216.239.46.26, "Googlebot/2.1"
216.239.46.39, "Googlebot/2.1"
216.239.46.96, "Googlebot/2.1"
207.253.45.199, "Java1.3.0"
211.72.252.150, "Openfind"
66.7.131.163, "Openfind"
209.73.162.22, "Scooter-W3-1.0"
209.73.162.41, "Scooter_trk0-3.0.3"
```

### Notes

- If the same IP address is used by more than one spider (User-Agent), that IP address will show up more than once in the list.
- If you are using POSIX style regular expression to help detect spiders, those expression may show up instead of the spider's ip address.

## 7.16 Options


Program Options allow you to set the look-and-feel for the application and any proxy server parameters that may be needed for an Internet connection. If you are not sure what a proxy is, then you probably don't need to enter any information for either proxy. Otherwise, you may need to consult with your network administrator for the correct settings. Here is a list of the options that are currently available.

### Application

- **Look-and-Feel.** Sets the display style for the application. The default is Flat, but you can change it to Standard or Windows XP to suite your style.
- **Show Quick Help at Startup.** Check this option to have Robot-Manager start with Quick Help showing. Quick Help is used to help guide new users through the program

for the first time.

### Spider Visits

- **Spider Visit Request Contains.** This is the text that Robot-Manager will look for in your log file that will indicate that a spider has visited your web site. Generally, this text should not be changed.
- **Ignore Requests Containing.** This field contains a list of text that Robot-Manager should use in determining if a log entry should be ignored as a spider visits. Use the  button to the right to enter more ignore text. This setting only applies on log entries that are a request for the robots.txt file.
- **Ignore "Unknown Spiders".** Check this option to have Robot-Manager ignore spiders that don't identify themselves. This option is only valid when spiders are identified by User-Agent and not IP address.
- **Display Date/Time Format.** Enter the format used to display the date and time in Step 4. This is helpful when sorting the list by date or time. [Formatting Specifiers](#)
- **Default Web Page.** This is the page name of the file that is displayed by default by your web server. This option is used to group similar requests. (i.e., www.domain.com and www.domain.com/index.html are the same request)

### HTTP Proxy

- **Server.** Specifies the host name or numeric IP address for the HTTP Proxy.
- **Port.** Specifies the port number for the HTTP Proxy.
- **User Name.** Specifies the user name, sometimes referred to as UserID, for the HTTP Proxy. User Name is used to perform HTTP Proxy authentication when required.
- **Password.** Specifies the password for HTTP Proxy authentication. Password is not needed if the proxy does not require authentication.

### SOCKS Proxy

- **Version.** The Socks protocol version required for the client connection to the Socks proxy server. The default is **Not Used** which indicates that a Socks server is not used. Version contains the following supported values: Socks 4, Socks 4a, and Socks 5.
- **Server.** Specifies the host name or numeric IP address for a Socks proxy server.
- **Port.** Specifies the port number used to connect to the Socks proxy.
- **User Name.** Specifies the user name for authenticating a connection to the Socks proxy server.
- **Password.** Specifies the authentication password for a connection to the Socks proxy server.

### Caution

- HTTP Proxy is not used for FTP or Telnet. Only SOCKS proxies are supported when using FTP or Telnet.

### Notes

- MS Proxy™ server normally runs as a client on your PC that you must log in to when you connect to the network. Settings for this proxy don't normally need to be entered here.
- You only need to fill in a User Name and Password when your proxy requires authentication. Otherwise, these fields should be left blank.
- When a host name doesn't seem to work for either proxy server, you can try and substitute the IP address to see if that helps.

- Most proxy servers are supported. If you are still having trouble connecting, you may have a configuration issue with your proxy server.

MS Proxy is a trademark or registered trademark of Microsoft Corporation.

## 7.17 Date/Time Formatting Options

These are standard ANSI-defined Format Specifiers that are used by Robot-Manager when displaying the Date and/or Time for a spider visit.

### Date and Time Specifier

- `%c` Date and time

### Date Specifiers

- `%x` Date
- `%a` Abbreviated weekday name
- `%A` Full weekday name
- `%m` Two-digit month as a decimal number (1 - 12)
- `%b` Abbreviated month name
- `%B` Full month name
- `%d` Two-digit day of month (01 - 31)
- `%y` Two-digit year without century (00 to 99)
- `%Y` Year with century

### Time Specifiers

- `%X` Time
- `%H` Hour of the day, 24 hour day
- `%I` Two-digit hour, 12 hour day (01 - 12)
- `%M` 2-digit minute (00 - 59)
- `%p` AM or PM
- `%S` Two-digit second (00 - 59)

### ◆ Important

- Time formatting specifiers are only available from Apache server logs using the standard log Time format. Using the Time formatting options with other log file formats results in the time not being displayed.
- The format of dates displayed on reports come from this setting.

## 7.18 FTP Browse Folder

The FTP browser dialog box is used to locate folders and/or files on your ftp server.

### Browse for Folder

The Browse for Folder dialog box is used when locating your home page directory on the web server. Files in the selected folder are displayed on the right to make finding this folder easier. Click **OK** when the selected folder is your web site's home page directory (root).

### Browse for File

The Browse for File dialog box is used when locating your web site access log file. Files in the selected directory are displayed on the right with approximate file sizes. Select the file that is your log file and click **OK**.

## 7.19 Enter Authorization Code

After you purchase Robot-Manager, you will need to enter the authorization code assigned to you on your Order form. This authorization code can unlock up to two copies of Robot-Manager. Typically, one for your work computer and one for your notebook or home computer. Once the code has been used twice, it will not work on any other computer but the two it was installed and registered to.

The registration process contacts our registration server over the web to validate your license. For this reason, you will need an Internet connection. (This shouldn't be a problem since the software was most likely downloaded over the Internet.) If the registration process stalls or doesn't seem to be working, there could be two problems.

### Potential Registration Problems

1. You may not be connected to the Internet and Robot-Manager is having trouble getting Windows to initiate a dial-up connection. In this case you may need to establish the dial-up connection first, then attempt the registration.
2. You may be behind a proxy or firewall. In this case you will need to enter those settings using the **Proxy Settings** button. If you are unsure of those settings, you may need to consult with your network administrator first. There are two types of proxies, HTTP Proxy and SOCKS Proxy. Robot-Manager supports them both.

### Entering Your Authorization Code

1. Find the authorization code on your Order form. It is usually listed below the Robot-Manager purchase item.
2. Enter the authorization code in the Authorization Code field. Authorization codes are **not** case sensitive and may be entered in either upper or lower case.
3. Click the **OK** button to register. This will contact our registration server and validate your license.

### Caution

- Robot-Manager is copyrighted software. Please do not make illegal copies or give out your authorization code.

### Notes

- If you have already registered your copy of Robot-Manager, this dialog box does nothing.
- Once registered, the About Robot-Manager dialog box will display your license information.

## 7.20 About Robot-Manager

The About box displays the version number, license information, and copyright notice.

### Version Number

Use the version number to compare your version with the version on our web site. If your copy is older than the current version on the web site, you may want to update. You can update by downloading the latest version and installing it on top of the your current version. You can download the latest version here:

<http://www.websitemanagementtools.com/downloads/>

### License Information

This area displays the registered user's full name and company. If the authorization code was for a full-trial, the number of days left for the trial is displayed.

### Copyright Notice

You can click this with your mouse to view the entire license agreement from our web site.

## 8 How Do I...

### 8.1 Add robots.txt to GoLive?

To add your new robots.txt file to your Site Document in GoLive, do the following:

#### Add robots.txt

1. Start GoLive and open your Site Document for your web site.
2. Make sure you have saved your robots.txt file to the root directory of your web site. If you have not done this yet, go back to [Step 3 - robots.txt](#) and export your robots.txt file to this directory.
3. Now with your Site window open and the Files tab showing, right-click in the tree view and choose **Update** from the popup menu. This will rescan your root folder and add your new robots.txt file to your web site project.
4. You should now see your new robots.txt file in the tree view (usually at the bottom, depending on your sort option).

#### Notes

- • Once the file has been added to your site, use the FTP Server commands to connect to your site and upload the new file.

### 8.2 Add robots.txt to FrontPage?

To add your new robots.txt file to your FrontPage web site, do the following:

#### Add robots.txt

1. Make sure you have saved your robots.txt file to the root directory of your web site. If you have not done this yet, go back to [Step 3 - robots.txt](#) and export your robots.txt file to this directory.
2. Start FrontPage and open your web site. FrontPage will automatically add your new robots.txt file to the web site.

#### Notes

- • Once the file has been added to your site, use the **Publish Web** command to connect to your site and upload the new file.

### 8.3 Upload robots.txt to my web server?

Uploading robots.txt to your web server is very simple in Robot-Manager. Once you have created an acceptable robots.txt file, click the **Upload robots.txt** button in the toolbar. You are presented with a dialog box that allows you to enter your FTP settings for uploading. Once those settings have been entered, click the **Upload** button. Because of the typical size of a robots.txt file, it will probably take more time to connect to your server than it will to upload this file.

#### Notes

- • If you are downloading your directory tree from your web server in Step 2, the Upload dialog box will default to the information you entered when setting your Home Page location.

### 8.4 Allow all spiders full access to my web site

There are two ways to allow all spiders full access to your web site. You can either not include a robots.txt file in your web site. Or, you can do the following:

#### **Allow All Spiders**

1. If not already open, Start Robot-Manager and create a new project.
2. Click **Upload robots.txt** on the toolbar.

#### **Resulting robots.txt File**

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

### 8.5 Exclude all spiders from my web site

To exclude all spiders from your web site, do the following:

#### **Exclude All Spiders**

1. If not already open, Start Robot-Manager and create a new project.
2. Click **Disallow** on the Project Steps sidebar.
3. Check the option "Disallow access to this site for the selected spider".
4. Click **Upload robots.txt** on the toolbar.

#### **Resulting robots.txt File**

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

### 8.6 Direct country specific spiders?

Many sites today are multi lingual and need to manage where spiders visit their site. If you have both German and English content on your web site, you don't want a German spider visiting and attempting to index your English content. They may give up and leave your site even though you provide German content. Here's how you can direct

those spiders to the relevant content of your site.

This example assumes a web site directory structure like the following:

```
http://www.mysite.com/index.html
http://www.mysite.com/english/index.html
http://www.mysite.com/german/index.html
```

This will be a simple site. The home page in the root directory gives the user a choice of either English or German. Depending on their choice, they are redirected to the home page in either the english or german directory. The home pages in these directories are language specific.

Again, for simplicity, we will only choose three spiders. The standard All Spiders and an English and German spider. Here are the spiders we chose:

```
All Spiders
Acoon      (German)
AltaVista  (English)
```

We'll jump ahead for a moment and show you what the resulting robots.txt file should look like. We have stripped out comments for brevity sake.

```
User-agent: Acoon Robot
Disallow: /english/
Disallow: /index.html

User-agent: Scooter
Disallow: /german/
Disallow: /index.html

User-agent: *
Disallow:
```

To explain, the first section for Acoon tells its spider to ignore the english directory and the home page. There is no need for the spider to index the home page since we are merely using it to redirect users to the appropriate language specific home page. Same goes for Scooter. It doesn't need to index the home page or any content under the german directory. All other spiders are given full access to the web site.

We could take it a step further and disallow access to the home page for all spiders. After all this page really has no content and merely redirects users to the appropriate language specific home page.

This was a simplified example, but does build the foundation for a more complex robots.txt exclusion file. This may even give you ideas on how to organize your web to be multi lingual.

## 8.7 Present optimized pages to search engines?

Many users will want to direct search engine spiders to optimized pages on their web site. Robot-Manager gives you the ability to do this very easily. As an example, we'll create a robots.txt to direct both AltaVista and Google's spiders to optimized home pages.

Our web site directory structure looks like the following:

```
http://www.mysite.com/index.html
http://www.mysite.com/scooter.html
http://www.mysite.com/google.html
```

This is a simple site, but serves the purpose for this example. The home page in the root directory is the one that all users will see when entering your web site. Both the scooter and google pages are optimized home pages for those engines. They should mimic the home page functionality, or provide a way for the user to navigate to your default home page quickly. Do not include any redirect statements in the header because the spider will normally skip the page.

Here are the spiders we choose for this example:

```
All Spiders
Scooter
Googlebot
```

We'll jump ahead for a moment and show you what the resulting robots.txt file should look like. We have stripped out comments for brevity sake.

```
User-agent: scooter
Disallow: /
Allow: /scooter.html

User-agent: googlebot
Disallow: /
Allow: /google.html

User-agent: *
Disallow:
```

To explain, for both Scooter and Googlebot we have disallowed access to the entire site, while we let all other spider through. After disallowing access, we set Allow rules for both spiders. In Step 2, this is easy, once you have checked the disallow access check box, the Disallow list becomes and Allow list. We allow Scooter to access the optimized scooter.html page, while we let Google access the optimized google.html page.

Of course, this is a simplified example. For a bigger site, you would also want to allow access to other relevant areas of your web site. This example allow Scooter and Google to only index the one page. You would probably want the spiders to continue indexing the rest of the site. You may even have more optimized pages for other things as well.

#### Notes

- • The Allow rule is a revision to the robots.txt standard. Some spiders may not be able to read allow rules. When using this feature, you will want to monitor your spider visits to make sure the spiders you are allowing are actually visiting the pages you have told them to.

## 8.8 Find out where my log files are located?

This is a tricky question since there is no standard in place as to where log files should be kept. Following here are some typical examples for some common types of web servers and hosting services.

### Shared Hosting Accounts

Web sites hosted by companies like EarthLink® are usually somewhat standard. When you log on using FTP, you will typically see four or five directories like the following.

```
<your domain>-logs
<your domain>-mail
<your domain>-secure
<your domain>-www
www
```

The directory you are interested in will be the one ending in "logs". This directory usually contains your current log file and an additional directory named "old", although it could be named different. Your log file will be named something like `access-log` or `web.log`, or something very similar. This is your current log file, and depending on how your account was set up, it will typically contain the current day, week, or month of log entries. Most shared hosting companies will archive this file once a day in your "old" directory to save space on the server.

Unless you have Telnet (Shell) access to your web server, the log files archived in the "old" directory will be of no use to you. For log files that are archived on a daily basis, you will need to extract the spider visits once a day before this log file is archived. Most hosting companies will archive the current log file sometime after midnight their time. If your log file is archived less frequently, you're in luck. You will not need to extract the spider visits every day. You can safely wait just before the log file is archived to extract the latest spider visits.

### Dedicated Hosting Accounts

If your web site is hosted on a dedicated server through a company like Interland®, you will have easier access, and more control, over your log file. For dedicated accounts, the easiest way to find your log file is to ask your service provider where it is located. Otherwise, consult the documentation that came with your server for the particular location of your web site access log file.

### Notes

- • If you are on a dedicated server with Telnet access, this is the fastest way to collect spider visits. Currently, Robot-Manager can use Telnet to extract spider visits from both uncompressed and compressed log files when ran from a Unix or Linux machine running Apache.

# Index

## - A -

about robot-manager 32  
 access 27  
 access log file 16, 25  
 access log location 24  
 allow 15  
 allow all spiders 34  
 authentication 29  
 automate 17

## - B -

browse 24, 31  
   ftp 24, 31  
   local 24  
 buy now 5

## - C -

category 14, 20  
 change 25  
   log file format 25  
 cloaking 29, 35  
 code 32  
   find 32  
 column sorting 11  
 command line options 17  
 contacting us 5  
 copyright notice 32  
 cost 5  
 country 14, 20  
 creating robots.txt 9  
 csv 27

## - D -

data mining 9  
 date 25  
 date order 26  
 date type 25  
 date/time formatting 31

description 20  
 description 21  
 disallow 15  
 disallow step 13, 14, 15  
 drag-and-drop 14

## - E -

enter authorization code 32  
 error code 25  
 excel 27  
 exclude all spiders 34  
 export 27  
   logs 28  
   spider ip addresses 29  
   visits 27

## - F -

features 8  
   professional edition 8  
   standard edition 8  
 find code 32  
 firewall 29  
 folder 24  
   access log 24  
   home page 24  
 folders 24  
 format 16, 27  
   date 25  
   date type 25  
   error code 25  
   log file 25  
   page request 25  
   page request type 25  
   spider ip address 25  
   spider name 25  
   spider type 25  
 frontpage 9, 33  
 ftp 6  
 ftp browse 31  
 ftp server 24

## - G -

golive 9, 33  
 grep 24

grouping 14  
gunzip 24

## - H -

home page 24  
host 29  
html 20, 21  
http proxy 29

## - I -

invert selection 11

## - L -

language 14, 20  
license 32  
    agreement 32  
    information 32  
license agreement 3  
local network 24  
locate log file 37  
log file 16  
    date 25  
    date type 25  
    error code 25  
    format 16  
    location 37  
    page request 25  
    page request type 25  
    spider ip address 25  
    spider name 25  
    spider type 25  
look up code 32  
lost authorization code 32

## - M -

mine log file 9  
mini html 20, 21  
ms proxy 29  
multi lingual 34

## - N -

name 20

## - O -

optimized pages 35  
options 29  
    application 29  
    http proxy 29  
    sock proxy 29  
    spider visits 29  
order by 26  
    date 26  
    page 26  
    spider 26  
order spiders 26  
overview 6  
    product 6

## - P -

page order 26  
page request 25  
page request type 25  
password 29  
phone numbers 5  
pricing 5  
product overview 6  
professional 8  
program options 29  
project overview 9  
properties 29  
proxy 29  
    authentication 29  
    host 29  
    password 29  
    port 29  
    server 29  
    user id 29  
    user name 29  
purchase 5  
purchase robot-manager 32

**- Q -**

quick start 9

**- R -**

raw logs 28  
refresh 24  
register program 32  
robots 12  
robots.txt 24  
    add to frontpage 33  
    add to golive 33  
    export 24  
    exporting 16  
    upload 34  
    uploading 16  
robots.txt step 16  
root directory 24  
rules 14

**- S -**

scheduling 17  
select 24  
    access log 24  
    home page 24  
select all 11  
select by 20  
selecting 11  
    spiders 11  
selecting spiders 20  
send robots.txt 24  
server 24, 29  
socks 29  
socks proxy 29  
spiders 20  
    country 20  
spider ip address 25  
spider name 25  
spider order 26  
spider selection 20  
spider type 25  
spider visit options 29  
spider visits 24  
spider visits step 16

spiders 12  
    add 12  
    adding 20, 26  
    allow all 34  
    category 20  
    cloaking 35  
    delete 12  
    description 20, 21  
    direct 34  
    drag-and-drop 14  
    edit 12  
    editing 20  
    exclude all 34  
    export 27  
    grouping 14  
    language 20  
    multi lingual 34  
    name 20  
    optimized pages 35  
    ordering 26  
    tracking 16  
    update 24  
    user-agent 20  
    visits 27  
spiders step 10, 11, 12  
standard 8  
step 1 10, 11, 12  
step 2 13, 14, 15  
step 3 16  
step 4 16

**- T -**

tasks 17  
telnet capture 24  
tracking 16  
tracking spider visits 9

**- U -**

unattended 17  
unlock program 32  
update 24  
upload 6, 16  
    robots.txt 24, 34  
user name 29  
user-agent 20

---

**- V -**

version 29, 32  
  information 32  
  number 32

**- W -**

web folder 24  
where to buy 5  
windows xp 29

**- X -**

xml 27