



**Inspyder Sitemap Creator  
User's Reference Manual**

Copyright © 2007 Inspyder Software Inc.

# Inspyder Sitemap Creator User's Reference Manual

Copyright © 2007 Inspyder Software Inc.

All rights reserved. No parts of this work may be reproduced in any form or by any means - graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems - without the written permission of Inspyder Software Inc.

Products that are referred to in this document may be either trademarks and/or registered trademarks of the respective owners. The publisher and the author make no claim to these trademarks.

While every precaution has been taken in the preparation of this document, the publisher and the author assume no responsibility for errors or omissions, or for damages resulting from the use of information contained in this document or from the use of programs and source code that may accompany it. In no event shall the publisher and the author be liable for any loss of profit or any other commercial damage caused or alleged to have been caused directly or indirectly by this document.

Printed October 2008 in Canada.

# Table of Contents

## Part 1 Introduction

## Part 2 Quick Start Guide

## Part 3 Usage

3.1	Button Bar & Menu .....	3
3.2	Project .....	4
3.2.1	Advanced Project Settings .....	7
3.2.2	Exporting .....	11
3.2.3	Import Robots.txt .....	13
3.3	Settings .....	14
3.3.1	Ping URLs .....	14
3.3.2	Email Configuration .....	15
3.3.3	Scheduler .....	16
3.3.4	Master Project Settings .....	19
3.4	Scan Results .....	19
3.4.1	Last Updated .....	21

## Part 4 About Inspyder Software

## Part 5 Product License Agreement

## Index

## 1 Introduction

Sitemap Creator from [Inspyder Software Inc.](#) is a Windows based utility that helps you create and keep up-to-date sitemaps for your websites. Sitemap Creator makes creating and uploading a sitemap easy. In addition to mapping your site, Sitemap Creator also has built-in FTP capabilities and can notify (or 'ping') the search engines automatically when your sitemap is updated. Sitemap Creator provides an end-to-end solution for creating and maintaining your sitemap files.

Sitemap Creator is a web crawler so it maps the pages that are served up by your webserver, regardless of your configuration. Sitemap Creator is compatible with Apache, IIS and other web server software. It can map pages created by PHP, ASP, JSP, or other technologies. There's no need to enter in hundreds of URLs. Just give it your homepage and let Sitemap Creator do the rest.

### Features and Benefits

- Ease of use - Just enter your homepage URL and go!
- Upload Directly from Sitemap Creator - Once your FTP settings are configured you can export and upload in one easy step.
- Ping Search Engines - Notify (or Ping) the search engines to let them know your sitemap has been updated from within Sitemap Creator.
- Bulk Configuration - Specify priority and update frequency information for many pages at once using partial URLs or wildcards to save time.

Sitemap Creator requires Windows 2000/XP/2003. Please refer to the [web site](#) for current licensing and pricing information.

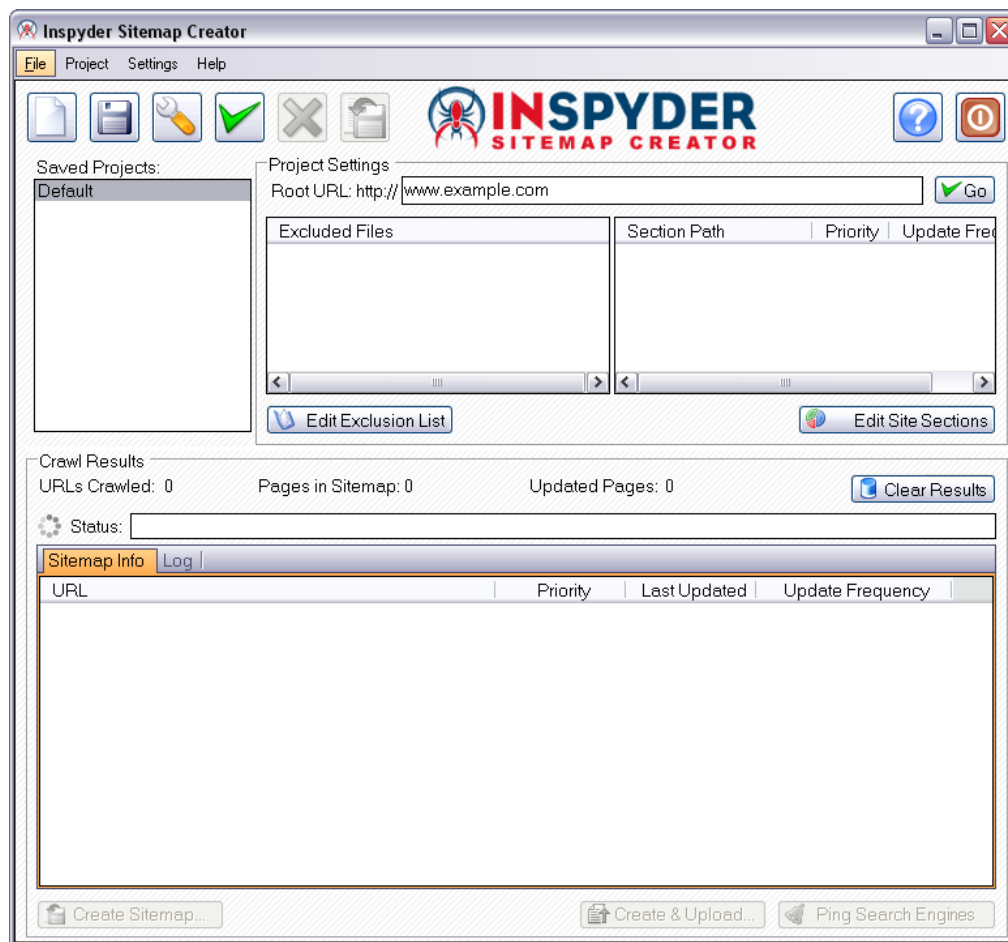
## 2 Quick Start Guide

**Step 1** The default project is now loaded but does not contain any settings. Enter the URL of your homepage in the 'Root URL' field.

**Step 2** Click the "**Go**" button.

**Step 3** When crawling is finished, click 'Export Sitemap' to save your sitemap to disk.

**Step 4** Optionally configure your FTP settings and upload your sitemap directly to your server.

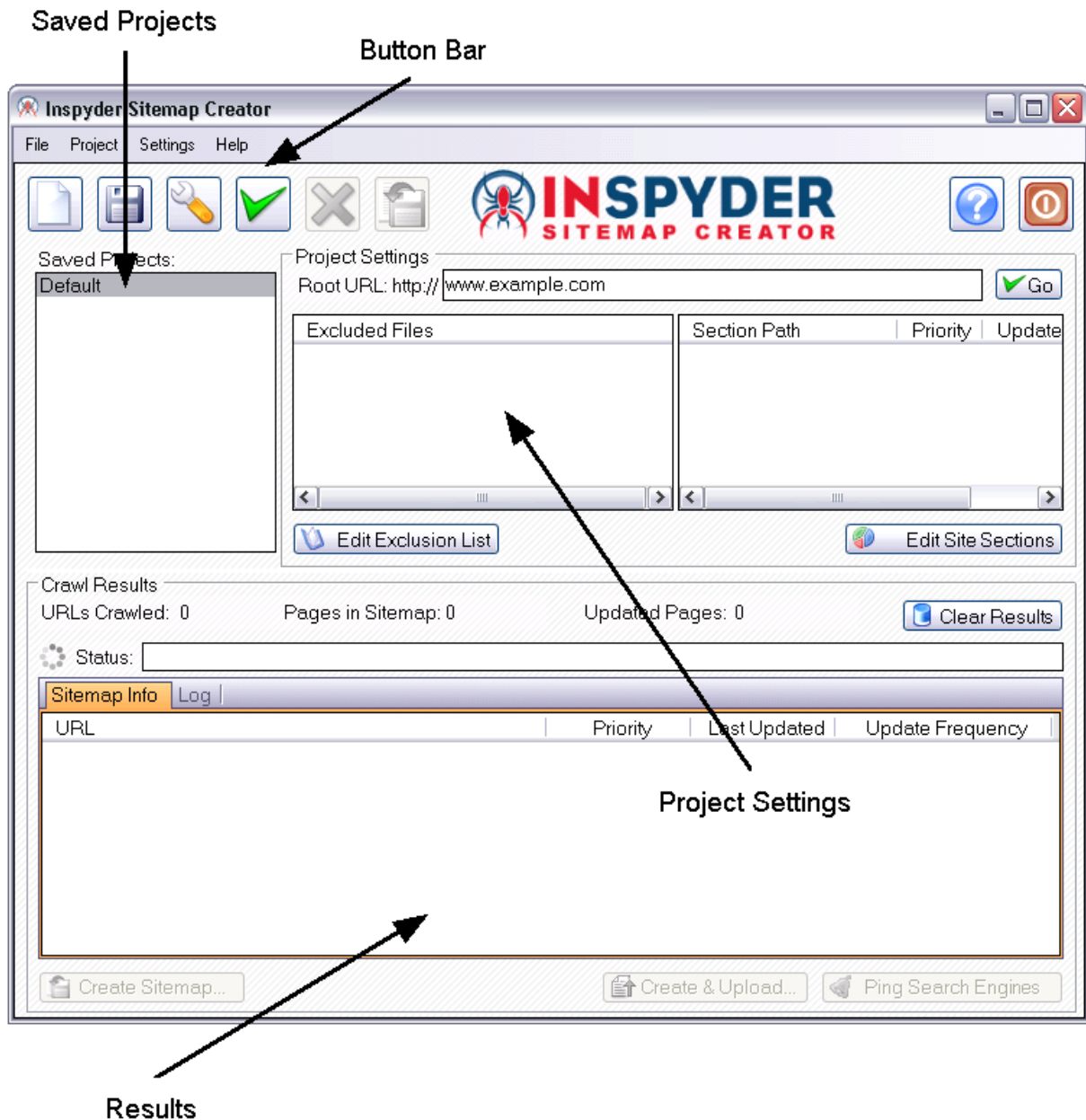


**Hint:** Configure your server's FTP information in the [Advanced Project Settings](#) screen. The Advanced Project Settings are accessible from the "Wrench" icon found on the button bar at the top of the screen.

## 3 Usage

The Sitemap Creator user interface area is broken down into four main areas:

- Button Bar** The actions performed by these buttons, with the exception of the 'wrench', affect the entire program or all projects.
- Saved Projects** This is a list of the currently saved projects. When you create a new project, it will appear in this list. Projects can be deleted or renamed by right clicking the project name in the project list.
- Project Settings** This is where the settings for the currently select project will appear.
- Scan Results** This is where the output from Sitemap Creator will be presented.



The following sections describe the individual fields and controls of Inspyder Sitemap Creator's user interface.

### 3.1 Button Bar & Menu

This section describes the functions of the button bar controls. These are the most often used features of Sitemap Creator and are placed on the button bar for convenience. Some of this functionality is available by right-clicking on controls in other parts of the user interface.



The **New Project** button is used to establish a new web site that you wish to analyse. You will be prompted for a shortcut name, which will appear in the Projects list, and the root URL of that web site.



The **Save Project** button is used to save the current project settings. The main menu contains a "**Save As**" command if you want to replicate the current project. As well, you can right-click on any project name in the Saved Projects list to delete or rename the project.



This button allows you to quickly access the Advanced Project Settings. For more information please refer to the [Advanced Project Settings](#) section of this document.



This button starts the crawling processing. The crawling can be interrupted at any time by clicking on the **Stop** button.



Use this button to stop the crawling process.



This button provides a shortcut to create (or export) your sitemap file. For more information please refer to the [Sitemap Export](#) section of this document.



Provide access to the help file and About screen. The About screen contains version and your registration information.



Closes the program.

The main menu contains most of the functionality of the button bar, but extends that capability with a few additional commands. The most notable menu additions are the "**File | Save As**" and "**Settings | Scheduler**" commands. The "Project" menu contains features and options that are specific to the current project. The "Settings" menu contains features that are program wide or affect all projects.


## 3.2 Project

This section describes the options, fields and features of a Project (and the Project menu).

### Root URL

Specify the address of the homepage for the website to be analysed. You can enter an IP address or the URL of the web page. The URL does not need to begin with the "http://" prefix. (If your site requires SSL, include the "https://" prefix.)

### Go Button

Use this control to begin the mapping from the Root URL. This button has the same functionality as the  control on the button bar. It's included in this location for convenience.

### Excluded Paths

The Excluded Paths listbox contains URLs that should not be part of the analysis. This is important if you have pages or scripts that generate a large amount of unimportant content (such as a message board or article pages that are formatted for printing). You might not want these included in your sitemap.

When a directory (like "/blog") is specified as an Excluded Path, then all of its sub-directories will be excluded (such as: [www.mysite.com/blog/index.html](http://www.mysite.com/blog/index.html) and [www.mysite.com/blog/archives/january2005.html](http://www.mysite.com/blog/archives/january2005.html)).

Wildcard matching is also supported through the use of the '\*' (asterisk) character. The '\*' will replace zero or more characters as shown in the examples below. The '\*' can be used multiple times in a string.

*inventory/warehouse\*index*

will exclude the following:

*inventory/warehouse1index*  
*inventory/warehouse125index*  
*inventory/warehouseindex*

*userreports/\*.txt*

will exclude all files with an extension of ".txt"

Click the **Edit Exclusion List** button to add URLs to the exclusion list. Internal URLs should be relative to the root URL. URLs beginning with '/' are taken from the domain root (this is not recommended). URLs **not** beginning with '/' are taken from the start of the crawl. For example, if the root URL is "www.mysite.com/somepage/":

*'apage.html'* becomes *'www.mysite.com/somepage/apage.html'*

*'/apage.html'* becomes *'www.mysite.com/apage.html'*

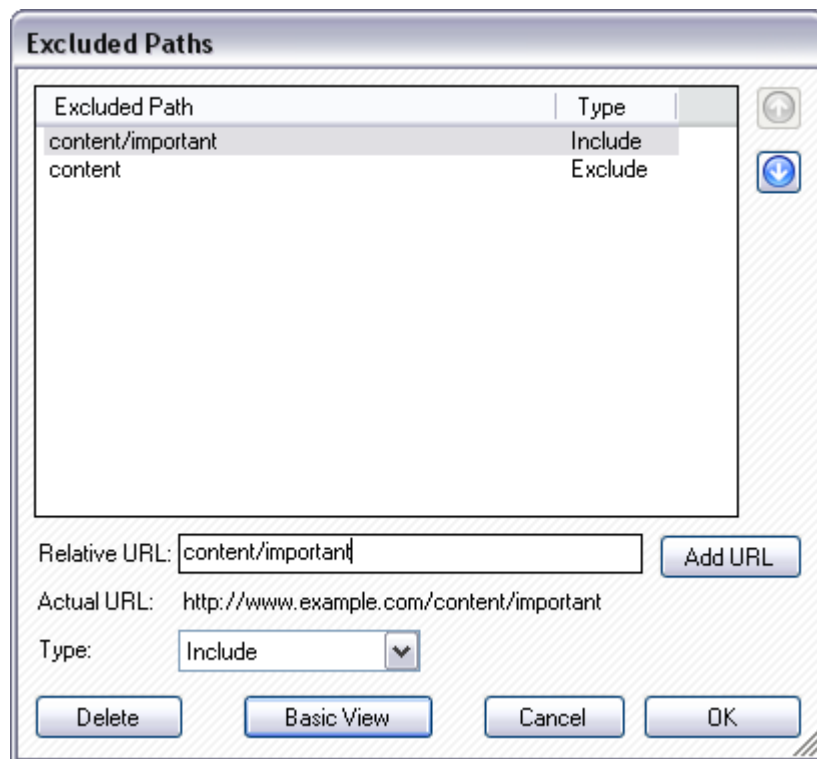
URLs beginning with 'http://' are treated as full URLs, and can be used to exclude links to external sites. For example:

*'http://www.anothersite.com'* **remains** *'http://www.anothersite.com'*

A page can also be added to the exclusion list by right clicking the URL in the Results tab.

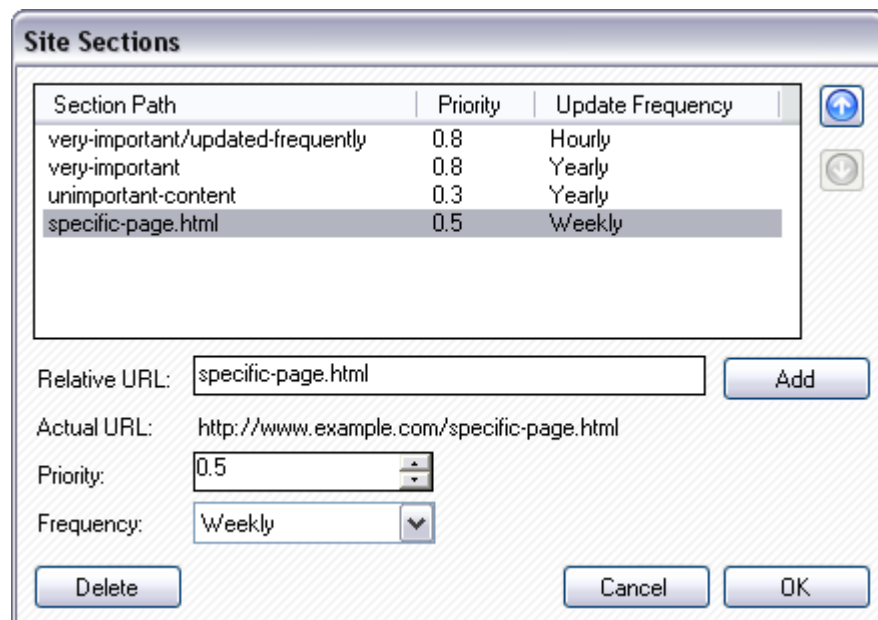
By clicking the "Advanced View" button on the exclusion dialog you are able to mark a page or section as specifically **included**. The list is matched from top to bottom. If you want to exclude all pages under "www.example.com/content/" except "www.example.com/content/important" then you would create an entry that includes "/content/important" above the rule that excludes "/content". You can move the rules up and down with the arrows on the right.





### Site Sections

Site Sections allow you to override the default Update Frequency and Priority values or a page or group of pages (section) on your site. A site section is defined the same way an excluded path is: with a Relative URL. To access the Site Sections dialog, click 'Edit Site Sections' from the main window.



To create a new Site Section, specify a Relative URL (you can use the wildcard (\*) character in your relative URL), choose the Priority and Update Frequency values you wish to give any files matching this section. Finally, click 'Add'. These rules

are matched from top to bottom, so it's necessary to move more specific paths (or paths you want matched first) to the top of the list.

You can convert any page in the "[Sitemap Info](#)" list to a Site Section by double clicking it. This will open the Site Sections dialog with the Relative URL field already populated.

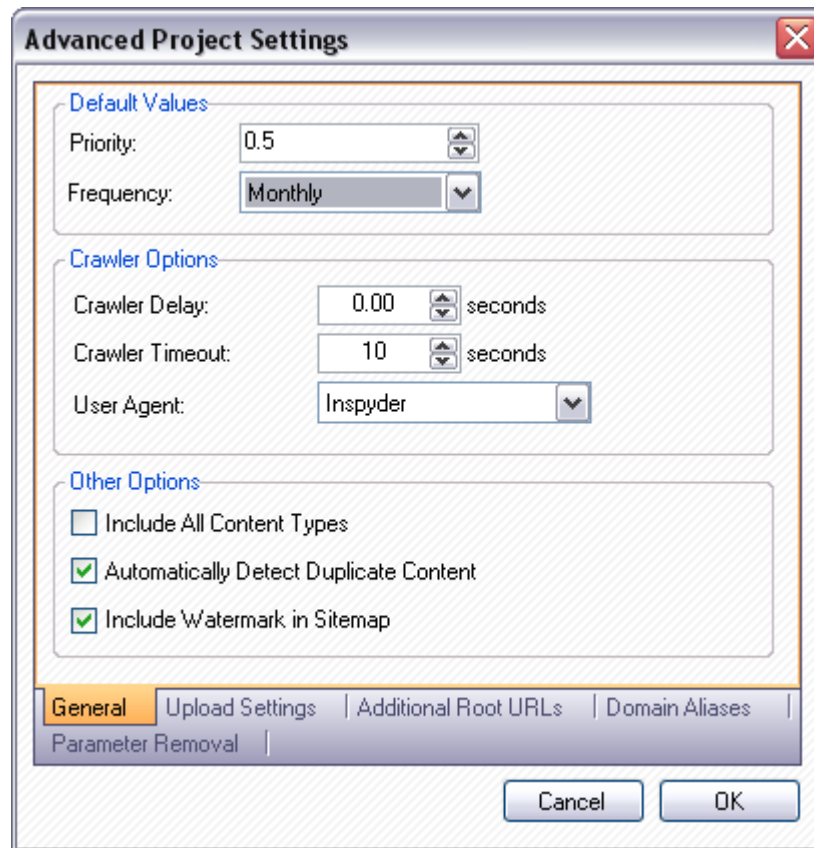
Any changes made to the Site Sections will be retroactively applied to your results when you click the 'OK' button.

For advanced users it may be of value to note that Sitemap Creator Projects and settings are stored in folder:

C:\Documents and Settings\username\Application Data\Sitemap Creator\

### 3.2.1 Advanced Project Settings

The Advanced Project Settings provide you with a way to further refine how Sitemap Creator crawls your site. These settings are project specific.



#### Default Values

These are the default values that are applied to your pages when no matching 'Site Section' rule is found.

#### Crawler Timeout

The Crawler Timeout value specifies the number of seconds that the crawler should wait for a response from your webserver. If your server is slow to respond or has long running

scripts, increase this value.

### **Crawler Delay**

The Crawler Delay value specifies how long the crawler should pause between HTTP requests to your server. If Sitemap Creator places too high a load on your server, increase this value to slow the crawl down.

### **User Agent**

The User Agent value indicates how Sitemap Creator should identify itself to your webserver. By default, Sitemap Creator identifies as the Inspyder Crawler. If your site uses browser detection scripts, you may wish to have Sitemap Creator masquerade as Internet Explorer or FireFox.

### **Include All Content Types**

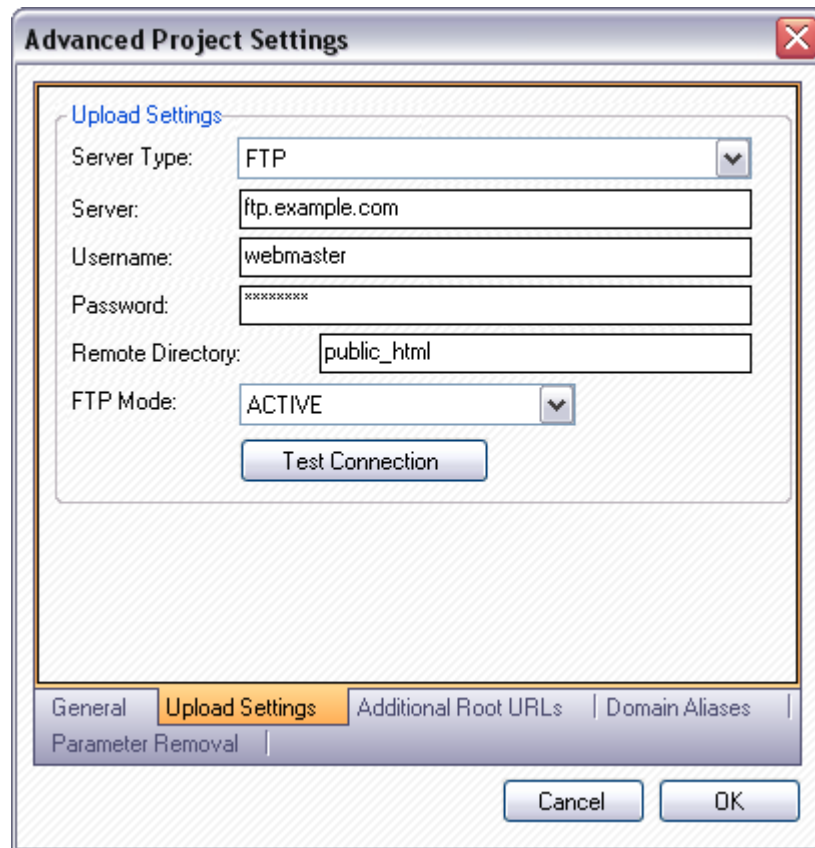
By default, Sitemap Creator only includes HTML and PDF content in your XML sitemap file. If you want Sitemap Creator to include every single file from your website in your sitemap, turn this option on.

### **Automatically Detect Duplicate Content**

Enabling this option will cause Sitemap Creator to check to see if each page's content is already included in the sitemap. This feature effectively weeds out any pages that have more than one possible URL. For example, if "index.html" is your server's default page, Sitemap Creator might find links to both "www.example.com" and "www.example.com/index.html". Enabling this option will prevent one of those URLs from being entered in the sitemap.

### **Include Watermark in Sitemap**

When this option is checked, Sitemap Creator includes an XML comment in your sitemap file indicating Sitemap Creator as the software used to create the sitemap and the version number. Typically this information is only used during troubleshooting. It is recommended that you leave this option turned on.



### Upload Settings

These settings allow you to configure Sitemap Creator to upload your sitemap directly to your server.

<b>Server Type</b>	The protocol that Sitemap Creator should use when communicating with your server. Most servers use FTP. If you are in doubt, try FTP first.
<b>Server</b>	The hostname or IP address of your FTP server. If your server uses a non-standard port number (not port 21), append it to the hostname, separated with a colon. For example: ftp.example.com:2121
<b>Username</b>	The FTP username you use to log into your FTP server
<b>Password</b>	The password you use to log into your FTP server
<b>Remote Directory</b>	The directory on the FTP server where your website's files are stored. This is typically 'public_html' on Linux/Unix servers. This field can be left empty if you do not require changing directories before uploading files to your site.
<b>Mode</b>	Specifies whether Sitemap Creator should use an Active or Passive FTP connection when communicating with your server. If you encounter problems, try toggling this option. (This option applies to FTP transfer modes only.)

### Additional Root URLs

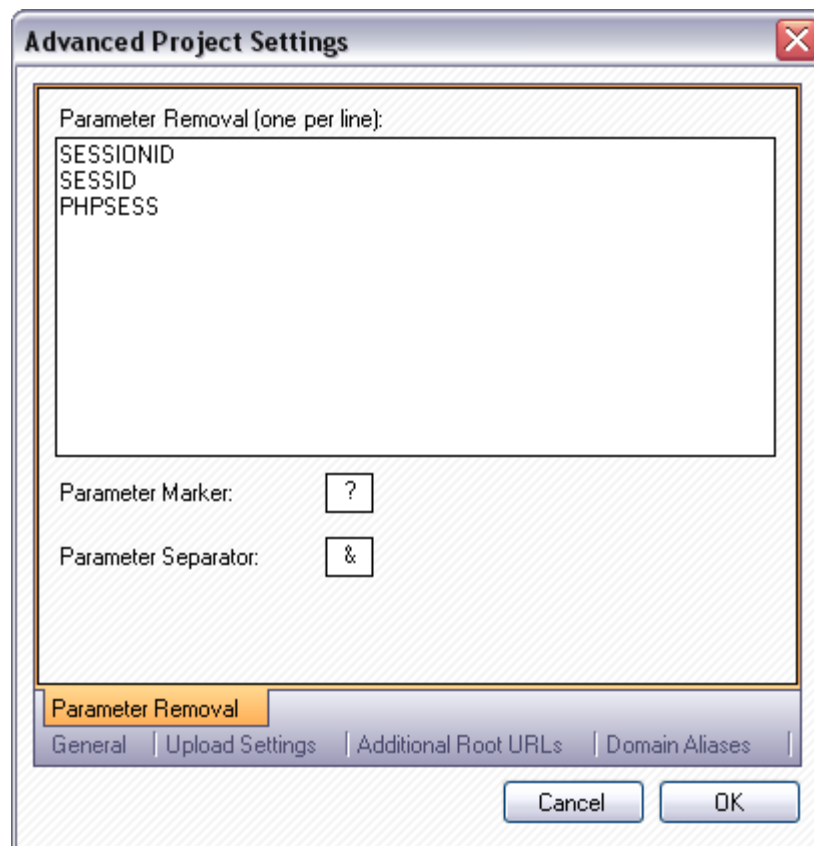
Additional root URLs are used to 'seed' the crawler if your site is not fully cross-linked (for example, if a particular page on your site has no inbound links, you may want to enter that URL here). **Important:** These URLs **MUST** begin with the same domain/hostname as your root URL otherwise they will be treated as external links and ignored.

To remove an unwanted Root URL from the list, select it with the mouse and press the delete ('Del') key on your keyboard to remove it.

### Domain Aliases

Domain aliases are used to tell the crawler that your site has more than one domain name. This feature is useful if your site is referenced internally by more than one domain. For example, many sites use "www.example.com" and "example.com" interchangeably. If the root URL has the "www" prefix, links using the other domain name will be treated as external to the site (and ignored) unless it is specified as a domain alias.

This feature should **NOT** be used to crawl multiple sites at once, as it could lead to an invalid sitemap file.



### Parameter Removal

Parameter Removal allows you to automatically strip off any unwanted URL parameters from the URLs in your sitemap. The XML sitemap specification defined by [sitemaps.org](http://sitemaps.org) indicates that your sitemap should not contain any session variables or other parameters that should not be used by a search engine when indexing your site.

### Parameter Marker

The Parameter Marker is the character that your web server uses to separate the parameters from the URL file.

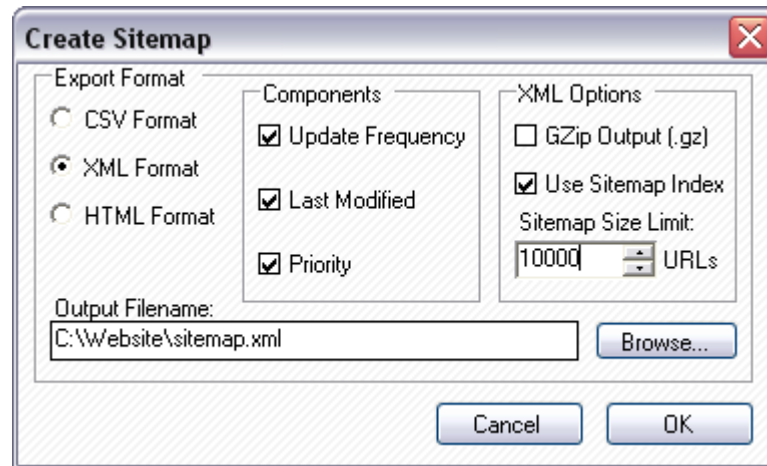
### Parameter Separator

The Parameter Separator is the character that your scripting environment uses to separate parameter/value pairs. (The '=' character is always used to separate parameters from values.)

### 3.2.2 Exporting

Once your site has been successfully crawled, the results can be exported in simple text (CSV) format, HTML, or XML.

The Export Sitemap dialog is used to specify which fields from the [Scan Results](#) section of the screen are to be included in the sitemap as well as the format and destination.



**CSV Format:** If the CSV Format is selected, the sitemap results will be exported to a simple comma separated text file that can be opened with Microsoft Excel, Wordpad, Notepad or any other text editor (or spreadsheet application). The file will optionally contain priority, update frequency and/or last modified data depending on which "Components" are selected.

**XML Format:** If the XML Format is selected, an XML file conforming to the standards listed on [www.sitemaps.org](http://www.sitemaps.org) will be generated. The output will always contain the "loc" tags, and depending which XML Options are selected, the "changefreq", "lastmod" and "priority" tags can also be included.

A sample XML output file is provided here:

```
<?xml version='1.0' encoding='UTF-8'?>
<!-- generator="Inspyder Sitemap Creator" -->
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.example.com</loc>
    <lastmod>2007-01-01</lastmod>
    <changefreq>monthly</changefreq>
    <priority>0.6</priority>
  </url>
</urlset>
```

**HTML Format:** The HTML format creates a sitemap that can be browsed by end-users and non-XML compatible web spiders/robots/crawlers. If this option is selected, a list of templates will appear (hiding the XML/CSV specific settings). The default templates provided with Inspyder Sitemap Creator can be customized to suit individual needs. Alternatively, the HTML output can also be modified in an HTML editor before publishing.

For information on creating and customizing the available HTML templates, refer to the [HTML Report Templates](#) section.

**Output Filename:** This is the filename of the sitemap (either in XML or CSV). If exporting for FTP upload, this file will be sent to the server. If saving the file to disk you may want to enter an absolute path so that the file can be easily located after exporting.

**Components (XML and CSV Formats Only):**

**Update Frequency:** If checked the exported file will contain the "changefreq" field (monthly, daily, etc.) for each URL.

**Last Modified:** If checked the exported file will contain the "lastmod" (the day when the file was last modified) field for each URL.

**Priority:** If checked the exported file will contain the "priority" field (1-10) for each URL.

**XML Options (XML Format Only):**

**GZip Output:** If the XML Format is selected, the option to "GZip" the output file will become active. If this option is checked, a ".gz" file will be created upon export that will contain a single "sitemap.xml" file. The gzip format is compatible with the standards outlined on [www.sitemaps.org](http://www.sitemaps.org) and can greatly reduce the size of the sitemap file.

**Use Sitemap Index:** If this option is checked, Inspyder Sitemap Creator will automatically split the outputted XML file into multiple parts. The sitemap will be split based on the Sitemap Size Limit field. Inspyder Sitemap Creator will automatically insert a number into the filename. The sitemap index will be in whatever filename is specified as the output file.

### 3.2.2.1 HTML Report Templates

Sitemap Creator uses customizable HTML templates for creating HTML sitemap files. The templates are simple HTML files with some special tags that tells Sitemap Creator how to format the template.

The template document structure is as follows:

Pre-Sitemap Content

Sitemap Content:

    Path Group (<isc\_group>  
        Item <isc\_item></isc\_item>  
    End of Path Group (</isc\_group>)

Post-Sitemap Content

Each time the path of the next URL in the sitemap changes, a new Path Group is created. The deeper into the path the URL is, the more Path Groups it will be in. Path Groups provide a way to give your HTML sitemap files a simple hierarchy.

Consider the following files:

<http://www.example.com/index.html>

<http://www.example.com/products/list.html>

<http://www.example.com/products/widget.html>

The "index.html" file would be in the base Path Group, while the "list.html" file and "widget.html" file would appear in the "products" Path Group.

The Item section is where the individual files/URLs are printed out in the report.

A simple template looks like this:

```
<HTML>  
<HEAD>
```

```
<TITLE>Sitemap</TITLE>
</HEAD>
<BODY>
<H1>Basic HTML Sitemap</H1>

  <isc_group>
    #PATH#<BR>
    <isc_item><A HREF="#URL#">#TITLE#</A><BR></isc_item>
    <BR> <!-- Insert a blank line for spacing !-->
  </isc_group>
</BODY>
</HTML>
```

For more examples, open the templates that were supplied with Sitemap Creator in Notepad or your favorite HTML editor. All templates MUST include the <isc\_group> and <isc\_item> tags (open and close) to be interpreted correctly.

The following markers can be embedded in the template file:

#PATH# The current URL Path of the Path Group (for example: http://www.example.com/products/)  
#URL# The current URL (used within an Item section)  
#TITLE# The page title (taken from the HTML)

The HTML report templates must be placed in the application's "Templates" directory. By default, it's located at:

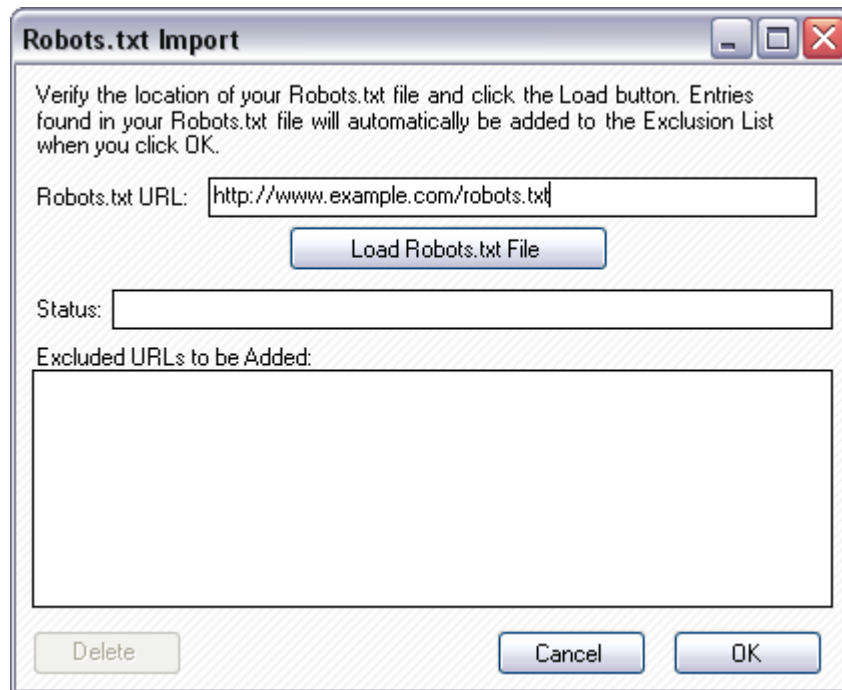
C:\Program Files\Inspyder Software Inc\Sitemap Creator\Templates\

Templates can have any file extension and can include other script languages if necessary. For example, it is possible to create a PHP template that includes PHP script to be run by your web server after being uploaded.

### 3.2.3 Import Robots.txt

If your website uses a "robots.txt" file which contains URLs you wish to exclude Sitemap Creator from crawling, you can import them with the Robots.txt Import tool. To access this window, select "Project | Import Robots.txt..." from the main menu bar.





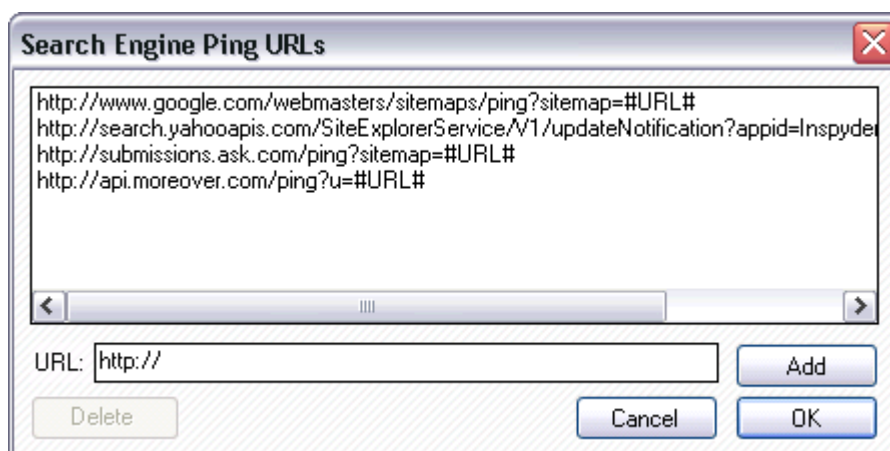
The "Robots.txt URL" field should contain the location of your website's robots.txt file. Sitemap Creator will try and guess this URL based on the Root URL for the current Project. If the URL that Sitemap Creator guesses is not correct, you can correct it in place. To load the file, click "Load Robots.txt File". Sitemap Creator will retrieve the file from your server and include any rules that match the wildcard User Agent ("User-agent: \*"). Other rules defined explicitly for certain robots/crawlers will be ignored.

Any rules that are found in the Robots.txt file will appear in the "Excluded URLs to be Added" list. You can remove any rules you don't want to import by selecting them and clicking the 'Delete' button.

## 3.3 Settings

### 3.3.1 Ping URLs

To access this option, click 'Settings | Edit Ping URLs' from the main menubar. This option applies to every project configured in Sitemap Creator.



### Search Engine Ping URLs

These are the URLs that Sitemap Creator will 'ping' each search engine with. To add an additional URL, enter it in the textbox and click 'Add'. Use '#URL#' to indicate where you want your sitemap's URL to be entered in the Ping URL. Sitemap Creator will automatically apply 'URL Encoding' the '#URL#' value.

Your URL is automatically generated by Sitemap Creator. Sitemap Creator applies the export/upload filename to the root URL and then uses the result as the '#URL#' value. Sitemap Creator is smart enough to strip any trailing filename from your root URL (if you've entered one).

For example: If your root URL is 'http://www.example.com/index.shtml' and you export 'mysitemap.xml' Sitemap Creator will create the #URL# parameter as: 'http://www.example.com/mysitemap.xml'

### 3.3.2 Email Configuration

ISC has provisions for customizing the emails that can be sent when scheduled scanning has completed. The customization screen is accessed by clicking on the "**Settings | Email Settings**" menu item.

Enter the **SMTP Server** (mail server) that is used for your outgoing email, and fill out the **Sender** field with the email address to be used as the sender. If your SMTP server uses a non-standard port, it can be appended to the server address with a colon (example: smtp.example.com:25). These settings will be used when automatically exporting the results of a [scheduled sitemap](#). The recipient's address is set up when setting Sitemap Creator to run on a scheduled basis.

Simply modify the **Subject** line or the email **Body** as desired and click the **OK** button to save your changes. Special tags can be used which are replaced with specific values by ISC before the email is sent.

Tag	Value
#URL#	The project Root URL
#SITEMAP_URL#	The URL of the exported (and uploaded) sitemap
#TOTAL_FILES#	The total number of files crawled on your site
#TOTAL_PAGES#	The total number of pages included in the sitemap
#TOTAL_UPDATE D#	The total number of pages in the sitemap that have changed since the last crawl
#PING_RESULTS#	A list of results from pinging the search engines
#UPDATED_URLS#	A list of URLs that were updated since the last crawl

The sample screen illustrates the usage of the tags in customizing the email message **Subject** and **Body**.

Also keep in mind that the report with the full scan results listing is included with the email as an attachment. The sections to include in the email and the format of the attachment are specified in the [Scheduler](#) set up.

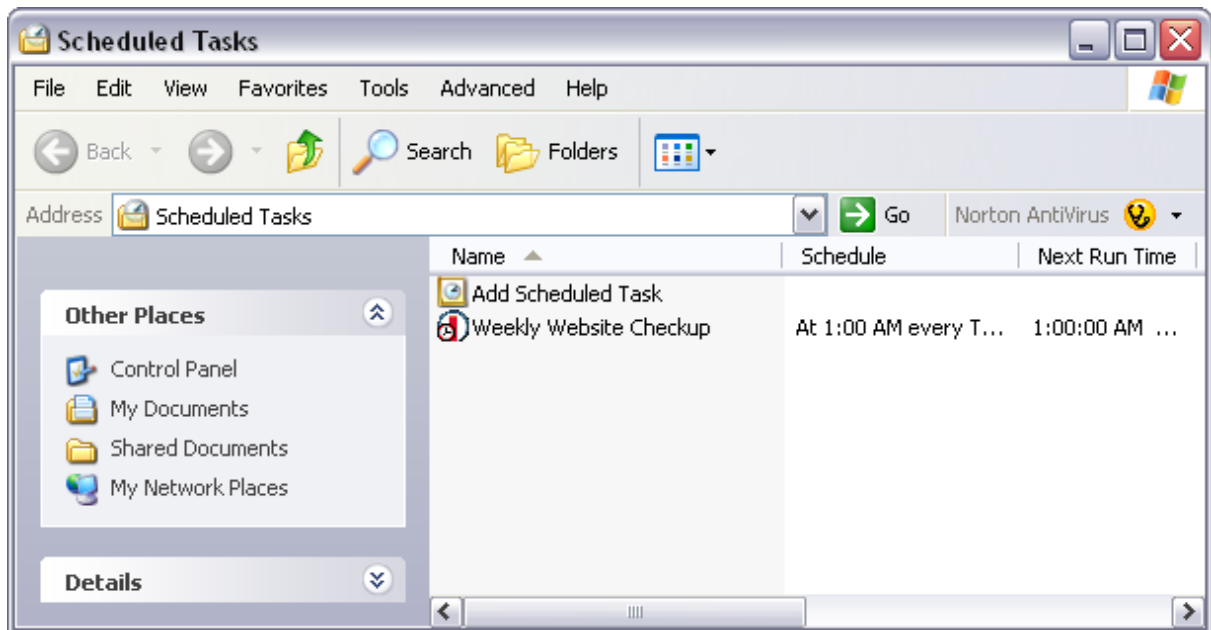
### 3.3.3 Scheduler

Sitemap Creator has the ability to run in an unattended mode and automatically export and upload a sitemap file. To configure Sitemap Creator for scheduled operation simply select "**Tools | Scheduler**" from the main bar. The dialog below contains the choices from the [Export Sitemap](#) dialog as well as those necessary for timing the execution of Sitemap Creator. Once the Schedule Options are filled in and the **Add Task** button is clicked the task will be scheduled using the Windows Task Scheduler.

The **Export Options** section is used to configure what actions to take when running in unattended mode. If "Only Upload if Sitemap has Changed" is checked then Sitemap Creator will do nothing (neither ping nor upload) if no changes to your website are detected since the last execution. If "Notify Search Engines after Export/Upload" is checked, then Sitemap Creator will 'Ping' the search engines after a successful export/upload operation. The "Export to Disk Only" option tells Sitemap Creator not to attempt an FTP upload (this is useful if Sitemap Creator is running on the web server itself).

If an email address is specified in the "Email Notification" field, then after crawling an email will be sent to that address. More information about configuring email notification can be found in the [Email Configuration](#) section.

Use the **View Schedule** button to view all the tasks submitted to the Windows Task Scheduler as well as **to modify the scheduling parameters**. If you double-click the scheduled entry in the Windows Scheduled Tasks dialog a new dialog will be displayed which contains all of the parameters for that task.



The actual command used to invoke Sitemap Creator is found in the 'Run:' field. You can modify the parameters used by Windows when it starts Sitemap Creator by **carefully** editing the contents of the Run field. The following are the command line arguments:

<b>-frequency</b>	(Optional) Includes 'Update Frequency' in XML sitemap
<b>-priority</b>	(Optional) Includes 'Priority' in XML sitemap
<b>-modified</b>	(Optional) Includes 'Last Modified' in XML sitemap
<b>-ping</b>	(Optional) Indicates that Sitemap Creator should ping/notify the search engines
<b>-checkforchanges</b>	(Optional) Indicates that Sitemap Creator should only export/upload/ping if the sitemap has changed since the last run
<b>-noupload</b>	(Optional) Indicates that Sitemap Creator should not perform an FTP upload after exporting
<b>-outputfile=</b>	Specifies the output file for the results (e.g. C:\sitemap.xml)
<b>-outputtype=</b>	Specifies output format; Text, XML, Gzip or HTML (e.g. -outputtype=XML)
<b>-project=</b>	The project name to be analyzed (e.g. -project=default)
<b>-email=</b>	(Optional) This specifies the email address of the email notification recipient (do not use this option if you do not wish to receive an email notification)
<b>-template=</b>	(Optional) This specified the name of the HTML template to use (this applies to HTML output format only)
<b>-xmlsizelimit=</b>	(Optional) Specifies the number of URLs per sitemap before rolling over to the next file (if this option is specified, a sitemap index file is automatically created, if not specified, a single sitemap file is created)

To satisfy the Task Scheduler's parsing algorithm please ensure that each argument is placed within double-quotes (e.g. "-outputfile=sitemap.xml"). This is usually only required if the argument contains spaces.

Note to Windows XP and 2003 users: The scheduled tasks will not run on an account with a blank password. For a workaround to this please see the Sitemap Creator FAQ.

#### Scheduled Execution Return Codes:

These are the codes returned to the Windows Task Scheduler after the Sitemap Creator task has been executed. You can view them in the Scheduler window to help troubleshoot any problems.

- 0 Success (no errors)
- 1 Not registered  
**Solution:** Run the program interactively and enter your registration information when prompted.
- 2 Report Generation Error  
**Solution:** Verify that you have permission to write to the folder where the sitemap is being exported to. Try running the report in interactive mode, and check for errors.
- 3 Crawler Error  
**Solution:** The crawler detected some type of unhandled error during crawling. Run this project in interactive mode and check for errors (corrupt HTML, etc.).

### 3.3.4 Master Project Settings

In addition to the regular user defined Projects, Inspyder Sitemap Creator contains a single Master Project. The Master Project contains the default values for any newly created projects. Options that are unique to every Project (such as Root URL) cannot be defined in the Master Project. To edit the Master Project click 'Tools | Edit Master Project...'. The settings in the Master Project Settings window are identical to those of a regular project.

The screenshot shows the 'Master Project Settings' dialog box. It is divided into two main sections: 'FTP Settings' and 'Default Values'.  
The 'FTP Settings' section contains:

- Server: [text box]
- Username: [text box]
- Password: [text box]
- Remote Directory: [text box]
- Mode: [dropdown menu] set to 'ACTIVE' with a 'Test' button next to it.

The 'Default Values' section contains:

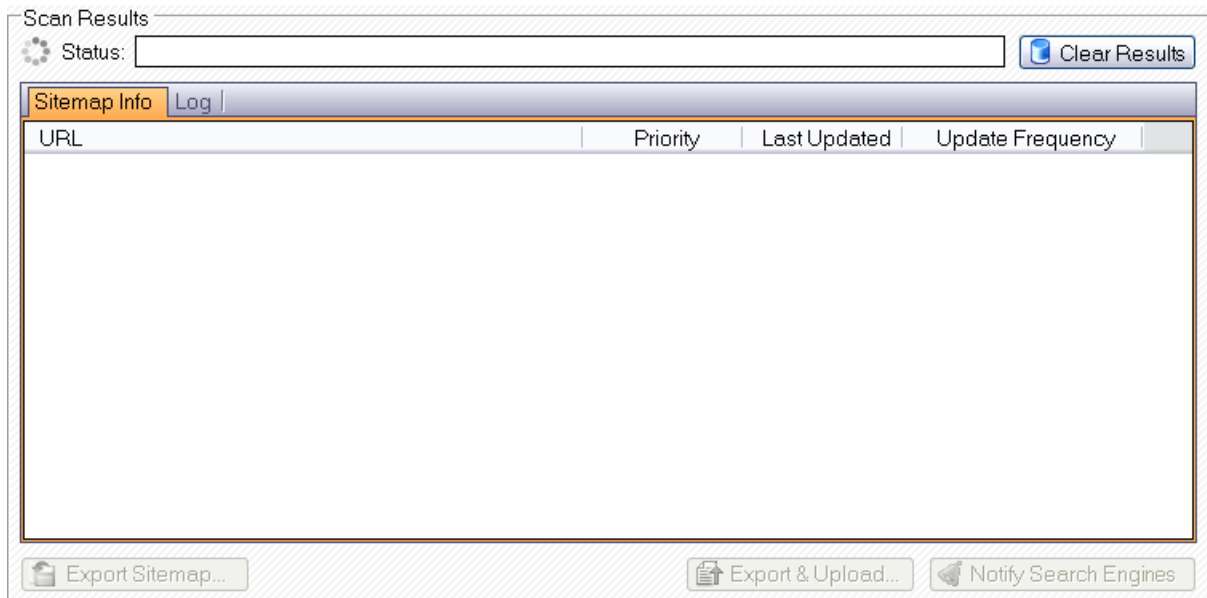
- Priority: [spinner box] set to '0.5'
- Frequency: [dropdown menu] set to 'Monthly'

At the bottom of the dialog, there are three tabs: 'FTP and Defaults' (which is selected and highlighted in orange), 'Parameter Removal', and 'Other'. Below the tabs are 'Cancel' and 'OK' buttons.

For details on what each specific options means, please review the settings in the ['Advanced Project Settings'](#) section of the manual.

## 3.4 Scan Results

This section describes the options, fields and features of the Scan Results group.



### Clear Results

Click to reset the contents of the **Scan Results** section, including all of the tabbed dividers.

### Sitemap Info

The contents of this tab provide information on what your sitemap export will contain. Each crawled (and mappable) page will appear here along with the frequency, last updated and priority values. Any pages matching a 'Site Section' rule will have their new priority and update frequency values applied here. Any changes you make to the Site Sections after crawling will also be applied to these results.

You can right click on any entry in the Sitemap Info tab to access a convenient menu which allows you to:

- **Create a Site Section from URL** - This option will automatically create a 'Site Section' based on the current URL. Choosing this option lets you specify a new Priority or Update Frequency value for this exact URL.
- **Open URL in Browser** - This option allows you to quickly open the selected page in your default browser
- **Exclude this Page from Future Crawls** - This option allows you to add this file to the Excluded Paths list

### Log

The Log provides a history of all the events and errors that occur while crawling your site. If you are experiencing difficulty with Sitemap Creator, the log can be helpful in trouble shooting your problem.

### Export Sitemap

This button provides a shortcut to exporting your sitemap to a local file.

### Export and Upload

This button provides a shortcut to exporting and uploading your sitemap in one step. For the upload to work, you must configure your FTP settings by clicking the 'Wrench' icon on the button bar. The status of the upload will be stored in the Log.

### Notify Search Engines

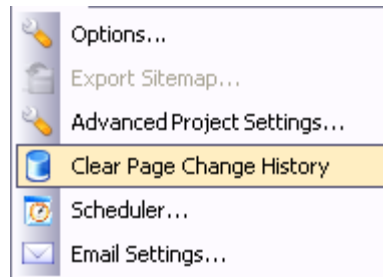
Clicking this button will cause Sitemap Creator to access each of the Search Engine Ping URLs (see [Options](#) for more details), informing the search engines that you've updated your sitemap file. After a successful upload or export this button will become enabled. The status of each ping result will be stored in the Log.

## 3.4.1 Last Updated

To make creating an accurate sitemap easy, Sitemap Creator automatically determines the "Last Updated" value for each of your pages. Essentially, Sitemap Creator remembers what each of your pages looks like when it crawls and compares it to the last time you updated your sitemap. If the page has changed, then Sitemap Creator sets the "Last Updated" value to the current date and updates its internal snapshot of that page.

It is important to update your sitemap when you make changes to your site so that Sitemap Creator is able to produce an accurate view of which pages were updated and when. This also helps the search engines (like Google and Yahoo) know which pages need to be re-indexed when they crawl your site.

To reset the "Last Updated" values for the current project, use the 'Clear Page Change History' option from the 'Tools' menu. (Note: If you release the Page Change History then the next sitemap you create will use the current date as the "Last Updated" value for all pages.)



To exclude a section of a page from Sitemap Creator's "Last Updated" check, the content can be enclosed in special SPAN tags where the class attribute is set to "isc\_exclude" (shorthand for "Inspyder Sitemap Creator Exclude"). This is particularly useful if a page contains content that changes every time the page is loaded, even though the content is not central to the page (such as a hit-counter or the current date).

For example, if your page contained the following PHP code to produced a simple hit counter:

```
We've had <?php print(getHits()); ?> visitors so far.
```

It could be modified as follows:

```
We've had <SPAN CLASS="isc_exclude"><?php print(getHits()); ?></SPAN> visitors so far.
```

Anything within the span tags will be ignored, so when the hit count changes, it won't trigger Sitemap Creator into thinking that the page has changed.

## 4 About Inspyder Software

Founded in 2004, [Inspyder Software](#) is a leading provider of web crawling technologies for content analysis, filtering and web access management. These technologies have formed the foundation of our software products and made 'Inspyder' a recognizable brand in the software industry.

As the World Wide Web continues to grow, organizations are looking for ways to improve their online presence, increase the quality of their sites, and maximize overall user experience. At Inspyder



Software, our focus is on the development of technology that enables organizations to do just that.

For more information regarding future products, custom development or consulting services, please contact us at [sales@inspyder.com](mailto:sales@inspyder.com).

## 5 Product License Agreement

**This legal document is an agreement between you, as licensee and Inspyder Software Inc. ("Inspyder"), as licensor. You should carefully read the following terms and conditions before using this product. Using this product indicates your acceptance of these terms and conditions.**

**DEFINITIONS:** "Product" means (a) all of the contents of the files, disk(s), CD-ROM(s) or other media with which this Agreement is provided, including but not limited to Inspyder or third party computer information or software; written documentation or documentation files; and (b) upgrades, modified versions, updates, and additions. "Use" or "Using" means to access, install, download, copy or otherwise benefit from using the functionality of the Product in accordance with the documentation. "You" and "Your" means the purchaser of the Product.

**GRANT OF LICENSE:** The Licensor grants to You a non-transferable non-exclusive license to use the product on a single computer. You may physically transfer the product to more than one computer provided the product is used only on a single computer. You may not modify, adapt, translate, reverse engineer, decompile, disassemble or create derivative work based on: (a) the product; (b) written material associated with the product; (c) the concepts or technology utilized in this product.

**COPY RESTRICTIONS:** You may not copy the product including any product that has been modified, merged or included with other products except as specified in this Agreement, nor may You copy any written materials associated with the product. You shall be held legally responsible for any copyright infringement that is caused or encouraged by Your failure to abide by the terms of this license.

**TRANSFER RESTRICTIONS:** The product is licensed only for You, and You may not transfer the product or the license to use the product without Licensor's prior written consent. Any authorized transferee of the product shall be bound by the terms of this Agreement. In no event may You transfer, assign, rent, lease, sell or otherwise dispose of the product on a temporary or permanent basis except as expressly provided for herein. The product cannot be resold.

**USAGE RESTRICTIONS:** The product is licensed only for Your personal use. You cannot use the product to provide a service to others. To be clear, you can use the product as part of your web development activities, including web development for others. However, you cannot use Sitemap Creator to provide an exclusive service such as using Sitemap Creator as part of a web checking or analysis service.

**LIMITED WARRANTY:** Except as expressly set forth herein, the product is provided 'AS IS' without warranty of any kind, either express or implied, including, but not limited to the implied warranties of merchantability and fitness for particular purpose. Inspyder Software Inc. does not warrant that the function of this product will be error free. However, Inspyder Software Inc. does warrant the media on which the software is furnished to be free from defects in material and workmanship under normal use for a period of 30 days from the date of delivery to You.

**LIMITATIONS OF REMEDIES:** The Licensor's entire liability and Your exclusive remedy shall be: (a) the replacement of any media not meeting Licensor's 'Limited Warranty' and which is returned to Licensor or an authorized representative of Inspyder Software Inc. with a copy of

Your receipt; or (b) if Inspyder Software Inc. is unable to deliver replacement media which is free of defects in materials or workmanship, You may terminate this Agreement by returning the product and Your money will be refunded. In no event will Licensor nor anyone involved in the creation, production or distribution of the product be liable to You for any direct, indirect, consequential or incidental damages including any lost profits, lost savings, lost business revenue or other commercial or economic loss arising out of the use or inability to use the product even if Licensor or any authorized representative of Licensor has been advised of the possibility of such damages or for any claim by any other party.

**JURISDICTION:** The laws of the Province of Ontario, Canada shall govern this Agreement.

All rights are reserved; Copyright © 2007 Inspyder Software Inc.

# Index

## - A -

Advanced Project Options 1

## - B -

Button Bar 2, 3

Buttons

Exit Program 3

Export 3

Help 3

New Project 3

Save Project 3

Start Analysis 3

Stop Analysis 3

## - D -

Default Values 7

## - E -

Excluded Pages 4

Exit Program Button 3

Export Button 3

Export Options 11

## - F -

Features and Benefits 1

FTP Options 7

## - H -

Help Button 3

## - I -

Inspyder Software Inc. 21

Internet Usage Management 21

Introduction 1

## - M -

Multiple Root URLs 7

## - N -

New Project Button 3

## - O -

Options 19

Overview 2

## - P -

Product License

Copy Restrictions 22

Definitions 22

Grant of License 22

Jurisdiction 22

Limitations of Remedies 22

Limited Warranty 22

Transfer Restrictions 22

Project Settings 2

Exclusion List 4

Root URL 4

## - R -

Root URL 4, 7

## - S -

Save As 3

Save Project Button 3

Saved Projects 2

Scan Results 2

Log 19

Sitemap Info 19

Scheduled Execution Return Codes 16

Site Sections 4

Start Analysis Button 3

Stop Analysis Button 3

**- T -**

Text Output 11  
TXT Output 11

**- X -**

XML 11  
XML Output 11



Inspyder Software Inc.  
698 Holt Drive,  
Burlington, Ontario,  
Canada, L7T 3N5

Tel: 888.732.6134

Web: <http://www.inspyder.com>  
Email: [support@inspyder.com](mailto:support@inspyder.com)