

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

Copyright © 2014 Inspyder Software Inc.

# **Inspyder Sitemap Creator (Sitemap Creator) User's Reference Manual**

**Copyright © 2014 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 2014 in Canada.

# Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
<b>2</b>	<b>Quick Start Guide.....</b>	<b>2</b>
<b>3</b>	<b>Tool Bar .....</b>	<b>3</b>
<b>4</b>	<b>Project .....</b>	<b>4</b>
4.1	<b>Advanced Project Settings .....</b>	<b>5</b>
4.1.1	<b>Image Sitemap .....</b>	<b>11</b>
4.1.2	<b>Soft 404 Detection .....</b>	<b>13</b>
4.2	<b>Priority Rules .....</b>	<b>14</b>
4.3	<b>Excluded Pages .....</b>	<b>15</b>
4.3.1	<b>Importing Robots.txt .....</b>	<b>17</b>
4.3.2	<b>Exporting Robots.txt .....</b>	<b>18</b>
4.4	<b>Passwords and Forms .....</b>	<b>19</b>
<b>5</b>	<b>Crawl Results.....</b>	<b>21</b>
<b>6</b>	<b>Creating a Sitemap.....</b>	<b>22</b>
6.1	<b>HTML Report Templates .....</b>	<b>24</b>
<b>7</b>	<b>Ping URLs.....</b>	<b>26</b>
<b>8</b>	<b>Scheduler .....</b>	<b>27</b>
8.1	<b>Scheduling Basics .....</b>	<b>27</b>
8.2	<b>Adding/Editing a Task .....</b>	<b>28</b>
8.3	<b>Deleting a Task .....</b>	<b>28</b>
8.4	<b>Scheduler Log .....</b>	<b>28</b>
8.5	<b>Command Line Interface .....</b>	<b>29</b>
8.6	<b>Result Codes .....</b>	<b>30</b>
<b>9</b>	<b>Email Settings.....</b>	<b>31</b>
9.1	<b>Server Settings .....</b>	<b>31</b>
9.2	<b>Message Body .....</b>	<b>32</b>
<b>10</b>	<b>Application Settings.....</b>	<b>34</b>
<b>11</b>	<b>Project Import and Export.....</b>	<b>35</b>
11.1	<b>Exporting Project Files .....</b>	<b>35</b>

11.2	Importing Project Files .....	36
12	Default Project Settings.....	37
13	About Inspyder Software.....	38
14	License Agreement.....	39

# 1 Introduction

Sitemap Creator from [Inspyder Software Inc.](#) is a Windows based utility for generating XML sitemaps. An XML sitemap is like a roadmap for search engine crawlers (such as Google, Yahoo and Bing). It helps crawlers quickly learn about all the pages on your site and when they were last modified. This helps your site get indexed more quickly and accurately.

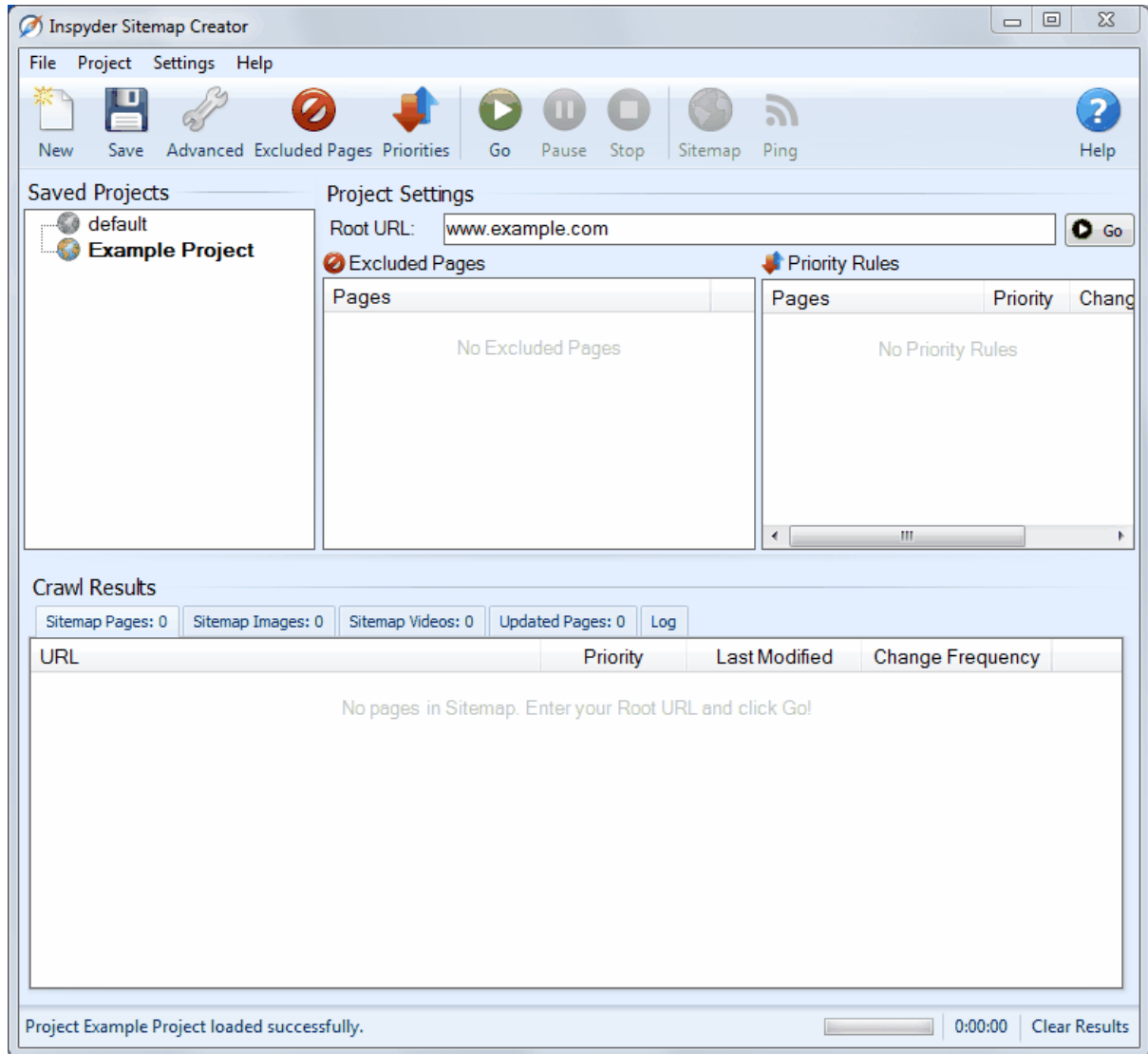
Sitemap Creator features a number of innovative features, such as:

- An integrated FTP client to upload your sitemap to your server.
- "Last Modified" value is automatically determined for each page by detecting content changes.
- Scheduler for automatically updating your sitemaps on a regular basis.
- Priority Rules so that your pages can easily be assigned a "Priority" or "Update Frequency" value when your sitemap is created.
- An easy to use Windows interface. Just enter your Root URL and click Go!

The details on how to get the most out of these features, and more, is explained on the following pages. If you have a question that is not addressed in this manual, please check our website for answers to other common questions and usage scenarios. If you need direct assistance you can also contact our support staff through our website.

## 2 Quick Start Guide

When you first run Sitemap Creator the default project is automatically loaded but contains no settings. Follow the steps below to configure Sitemap Creator for the first time.

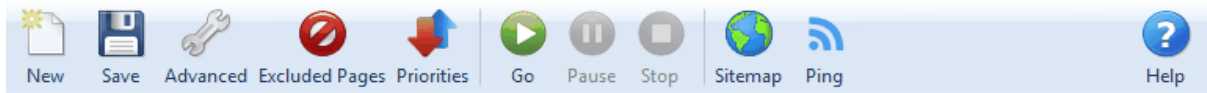


- Step 1** Enter the URL of your homepage in the 'Root URL' field.
- Step 2** Click the "Go" button. Sitemap Creator will crawl your website and list all the pages it finds in the "Sitemap Info" list.
- Step 3** When crawling is finished, click 'Sitemap' to save your sitemap to disk.
- Step 4** Upload your sitemap file to your server, then click the "Ping" button to notify the search engines that your sitemap has been updated.

**Hint:** Configure your FTP server settings in [Advanced Project Settings](#) and you can automatically upload your sitemap from Sitemap Creator.

### 3 Tool Bar

This section describes the functions of the tool bar shortcuts. These are the most often used features of Sitemap Creator and are placed on the tool bar for convenience. Some of this functionality is available by right-clicking on controls in other parts of the user interface, or from the main menu bar at the top of the screen (File, Project, etc.).

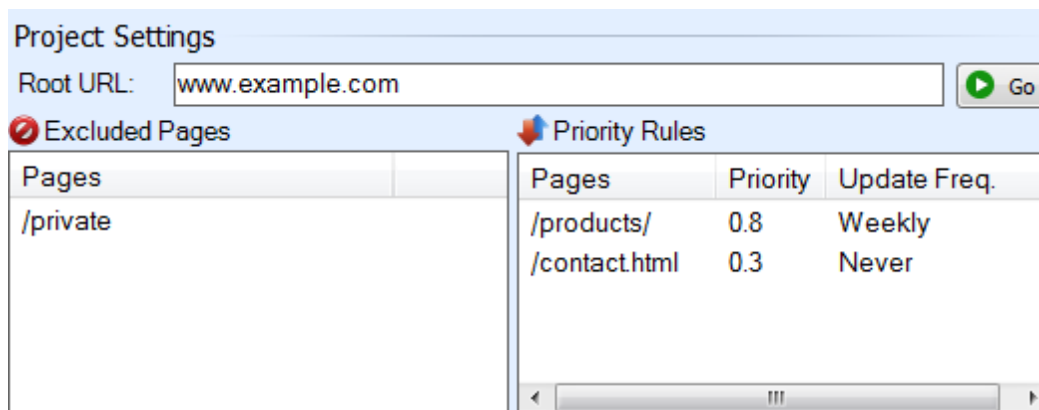


- **New** - The **New Project** button is used to establish a new website that you wish to create a sitemap for.
- **Save** - The **Save Project** button is used to save the current project settings. The main menu contains a "**Save As**" command if you want to replicate the current project. As well, you can right-click on any project name in the Saved Projects list to delete or rename the project.
- **Advanced** - This button allows you to access the Advanced Project Settings for the current project. For more information please refer to the [Advanced Project Settings](#) section.
- **Excluded Pages** - Opens the [Excluded Pages](#) window which can be used to control which pages the crawler visits on a website.
- **Priorities** - Opens the Priority Rules window which is used for configuring the sitemap Priority and Update Frequency values for specific pages or groups of pages.
- **Go** - This button starts the crawling process on the current Root URL. The crawl can be interrupted at any time by clicking on the **Stop** button.
- **Stop** - Use this button to stop the crawling process.
- **Sitemap** - Use this button to create your sitemap file. This button is disabled until you have crawled your website (or a portion of it).
- **Ping** - This button "Pings" the search engines with your sitemap URL to indicate that your sitemap has been updated. This button becomes active after your generate your sitemap using the "**Sitemap**" or "**Sitemap & FTP**" buttons.
- **Help** - Opens the help file.

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.

## 4 Project

A project contains settings that are typically associated with a single website. This section describes the options, fields and features of the Project Settings area.



**Root URL** - Specify the address of the homepage for the website to be crawled. 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.) It is important to enter your Root URL the same way it should appear in your sitemap file. If your website address begins with "www", then be sure to include it in the Root URL.

Sitemap Creator will only crawl from your Root URL and deeper into your site. Unless you only want to create a website for a subsection of your website, use your domain name of your website as the Root URL.

**Excluded Pages** - Excluded pages are used to explicitly exclude (or include) files or paths from your sitemap. More information on this is available in the Excluded Pages section of the manual.

**Priority Rules** - Priority Rules are used to specify explicit "**Priority**" and "**Update Frequency**" values for pages and paths on your website. Creating rules allows Sitemap Creator to automatically assign these values the next time your website is crawled. More information is available in the Priority Rules section of the manual.

If no Priority Rules are defined, the default Priority and Update Frequency values are used. To change the default values, see [Advanced Project Settings](#).

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

**Windows 2000/XP:**

C:\Documents and Settings\[username]\Application Data\Inspyder Sitemap Creator\

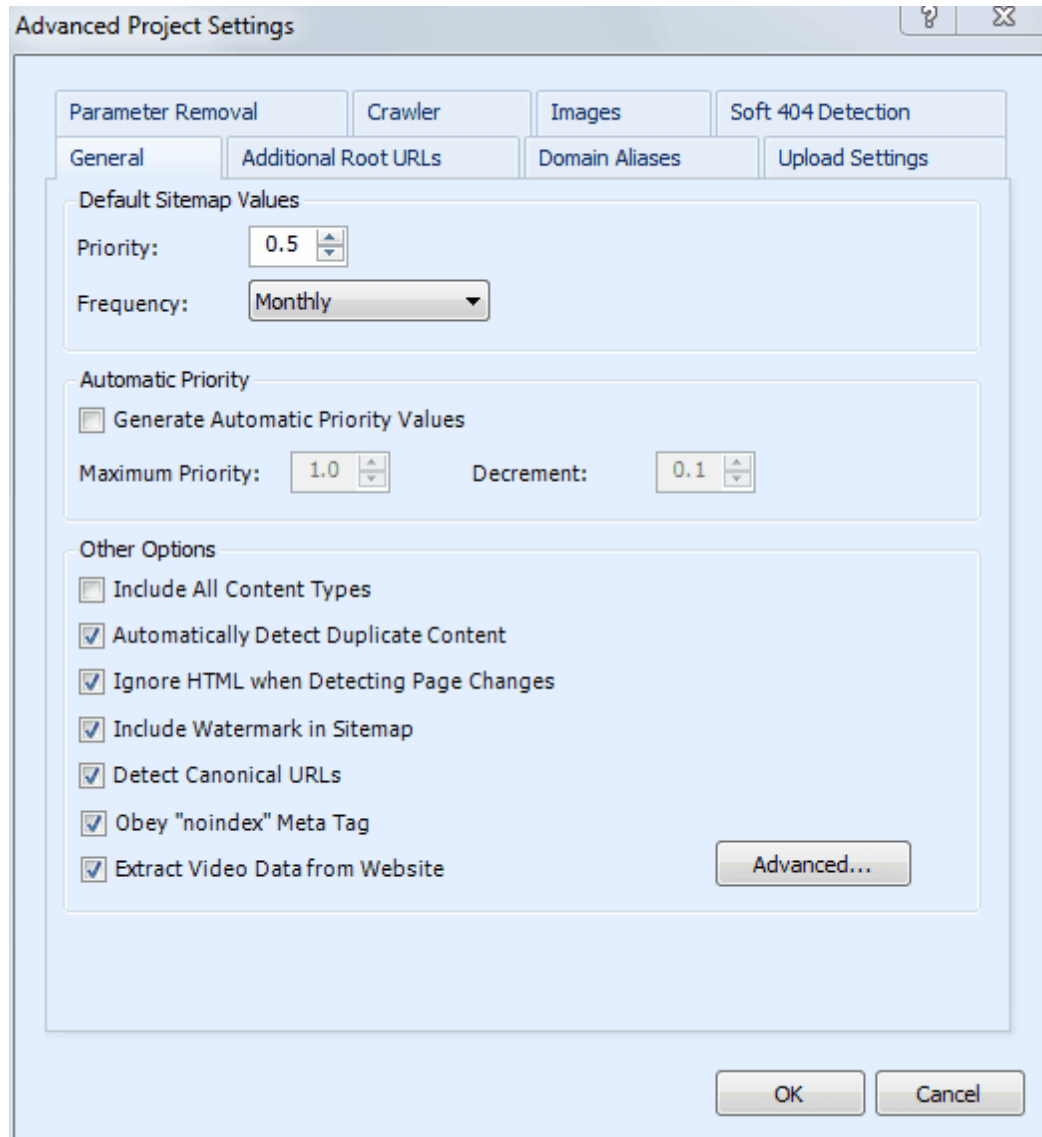
**Windows Vista/7:**

C:\Users\[username]\appdata\roaming\Inspyder Sitemap Creator\



## 4.1 Advanced Project Settings

To access the advanced project settings, click the 'Wrench' icon from the button bar, or choose "**Project | Advanced Project Settings**" from the main menu bar.



### General Tab

**Default Sitemap Values** - The Default Sitemap values are used to specify which values will be used in your sitemap when no Priority Rules match a particular URL. If you have no Priority Rules defined, then the Priority and Frequency values specified here will be used for all the pages in your sitemap.

**Automatic Priority** - Automatic Priority is used to dynamically assign a priority value for each page in your sitemap based on the link depth from the Root URL. This feature is useful if you do not wish to configure Priority Rules, but do not wish to use the same priority value for all pages. If [Priority Rules](#) are defined, only pages with no matching Priority Rule will be assigned an automatic priority value.

**Generate Automatic Priority Values** - Check this box to enable Sitemap Creator's Automatic Priority feature.

**Maximum Priority** - This is the starting priority value used for the Root URL. The default is 1.0.

**Decrement** - This is the amount that the priority will be decreased for each link depth level into the website. If this value were set to 0.1 and the Maximum Priority value were set to 1.0, a page 2 "clicks" deep into the site would receive an automatic priority value of 0.8.

### Other Options

**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 removes duplicate 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 ensure only one of those URLs is entered in the sitemap.

**Ignore HTML when Detecting Page Changes** - When this option is checked, Sitemap Creator will strip the HTML markup from your pages before looking for any changes to webpage content. Sitemap Creator looks for content changes to help determine the Last Modified date. By default this option is enabled.

It is also possible to exclude sections of text when detecting page changes. To do so, click the "**Advanced...**" button and the Advanced Sitemap Options window will open. Here you can add a CSS class name or ID to the list of Excluded Text that will be removed before detecting page changes. For example, assume the following HTML code were used on your website:

```
<div class="example">This is some sample text.</div>
```

If "example" were added to the list of classes for Excluded Text, then all the text contained within this "div" block would be ignored. ("This is some sample text.")

**Note to version 1.x/2.x users:** *Sitemap Creator no longer excludes classes named "isc\_exclude" automatically. You can reproduce this behaviour by adding "isc\_exclude" to the list of excluded CSS classes.*

**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.

**Detect Canonical URLs** - This option tells Sitemap Creator to watch for and obey Google canonical URL meta tags. These tags tell the crawler the preferred URL to use when accessing a particular page. If this tag is found by Sitemap Creator, the preferred URL will be used in the sitemap. This option is enabled by default.

**Obey "noindex" Meta Tag** - This option tells Sitemap Creator omit any pages with the "noindex" robots meta tag from your sitemap. This option is enabled by default since most websites using the "noindex" tag will not want these pages to appear in the sitemap.

**Extract Video Data from Website** - This option tells Sitemap Creator to detect embedded videos for inclusion in the sitemap.

## Crawler

**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 itself as the Inspyder Crawler. If your site uses browser detection scripts, you may wish to have Sitemap Creator masquerade as Internet Explorer or FireFox. The "Custom" option allows you to enter a free form User Agent string. This feature is useful if you have a specific browser you want Sitemap Creator to mimic (such as a mobile browser or iPhone).

**Maximum File Size** - This option limits the amount of data the crawler will download from a single URL. This is useful for websites that have either very large HTML pages (1 MB or more) or script errors that cause pages to generate infinite amounts of HTML. If you receive "Out of Memory" errors while crawling your website with Sitemap Creator, try setting a limit of 1024 KB. For most websites setting this value is not required (use 0 for no limit).

**Maximum Link Depth** - This option limits the depth of the crawl for the current project. The depth of the crawl is the number of links (or 'clicks' for a user browsing your website) from the root URL. For example, if the link depth limit was 1, then all the pages that could be reached in one mouse-click from the root URL would be crawled. If the link depth is 0, then there is no limit.

To limit the crawl by path depth (such as "no pages deeper than 3 subdirectories") you can create an entry in the Exclusion List that looks like: `"/**/**"`. This would exclude any URLs that had 3 or more '/' characters.

**Maximum File Count** - This option limits the maximum of URLs to crawl before stopping. This feature is useful if you wish to crawl a certain number of pages on your site before stopping completely. Note that the number of URLs crawled includes images, CSS files and other content that may not be included in your sitemap.

**Maximum URL Length** - This option limits the maximum URL length the crawler will follow. This provides a mechanism to limit the crawler from crawling endlessly on websites that have infinitely recursive links. If you notice that Sitemap Creator seems to be crawling similar URLs over and over again, try setting this value to limit the problem. For most websites setting this value is not required (use 0 for no limit).

**Crawler Threads** - The crawler threads slider is used to adjust the number of simultaneous HTTP requests sent to your webserver. By default, Sitemap Creator is set to use 5 threads. Increasing this value can make Sitemap Creator crawl your site more quickly, but puts more load on your local computer and on your web server. If you find that your website is running slowly while crawling with many threads, you may need to decrease this value.

**Process Links in JavaScript** - If this option is checked Sitemap Creator will attempt to locate URLs

within JavaScript code.

**Use JavaScript Heuristic** - If this option is checked Sitemap Creator will use a more aggressive technique for locating links within JavaScript. If a site makes heavy use of JavaScript (such as for the menu system or other navigation) it may be required to enable this feature for Sitemap Creator to automatically discover all your pages.

**Process Links in PDF** - If this option is checked Sitemap Creator will attempt to locate URLs within PDF documents.

**Process Links in XML** - If this option is checked Sitemap Creator will attempt to locate URLs within XML content. Note that because XML documents generally contain text attributes, Sitemap Creator will utilize a heuristic (similar to the JavaScript heuristic) to determine which text within the XML document are URLs.

## Advanced Crawler Options

To access the Advanced Crawler Options window, click the "Advanced..." button on the Crawler Options tab page.

**Use HTTP Compression** - This option tells Sitemap Creator to use compression when communicating with your webserver. If you see error messages that indicate problem with the "Zip Header", try disabling this feature.

**Use HTTP 1.0** - This tells Sitemap Creator to use HTTP version 1.0 instead of version 1.1. This option should be disabled unless you are experiencing difficulty crawling your website.

**Use HTTP Keep-Alive** - This option tells Sitemap Creator to keep the HTTP connection (if possible) between multiple requests. Disabling this feature can have a negative impact on performance, but may be necessary with some older servers.

**Obey "nofollow" Meta Tag** - If this option is enabled links found on pages with the "nofollow" meta tag will be ignored.

**Ignore Protocol Prefix** - This option tells Sitemap Creator to ignore the http/https protocol prefix when comparing URLs. Disable this option if your website's secure pages (https) are different than the normal, non-secure pages (http). An easy way to check this is visit <http://www.yoursite.com> and <https://www.yoursite.com> (notice the different protocol prefix). If these pages appear the same leave this option checked. If these pages appear different then uncheck this option. Most sites will leave this checked (default).

**Ignore HTTP Content-Type Encoding** - This option tells Sitemap Creator to ignore the HTTP "Content-Type" header when determining the page encoding of your website. This option only needs to be checked if your website's HTTP Content-Type headers conflict with any "Content-Type" meta tags used within your site. For most websites this should be left disabled.

**Ignore HTTP Content-Type Encoding** - This option tells Sitemap Creator to ignore the HTTP "Content-Type" header when determining the page encoding of your website. This option only needs to be checked if your website's HTTP Content-Type headers conflict with any "Content-Type" meta tags used within your site. For most websites this should be left disabled.

**Parse Form "Action" Attribute** - This option tells Sitemap Creator to parse URLs from the HTML "action" attribute on <form> tags. This option is disabled by default.

**Accept Cookies** - This option tells the crawler to accept cookies from the website. The default is enabled.

## Images

See [Image Sitemap](#).

## 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, some 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.

## Additional Root URLs

Additional Root URLs are used to 'seed' the crawler if your site is not fully cross-linked. If you have doorway pages, or a disconnected section of your site (to which no links point), you can enter these pages as Additional Root URLs. The crawler will process these pages while crawling your site.

To add a new Root URL, simply enter it in the URL field and click 'Add'. To remove a URL from the list, select it and press the delete key ('Del') on your keyboard.

## 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](#) 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.

For example, if your website used a parameter named "jssid" that should be stripped from all URLs in the sitemap, you would add "jssid" on a line by itself. Do not include a trailing "=" or the parameter values.

**Correct:**

```
jssid
```

**Incorrect:**

```
jssid=  
jssid=324234234
```

**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 parameter names from values.)

**Secondary Markers and Separators** - (Advanced use only, most websites these two fields should be left blank.) The Secondary Marker and Separator settings are for use with sites that use multiple parameter separator formats on a single URL. For example:

```
http://www.example.com/catalog/PartSelector.do;jsessionid=17AA7B17650BFC888A308999392F03FF?mod
```

In this URL example, the ";" and "?" characters are both Parameter Marker values. Since "?" is the preferred marker, we'd specify "?" as the Parameter Marker, and ";" as the Secondary marker. If secondary values are found within the URL, they will be replaced by the primary values when the URL is reconstructed by Sitemap Creator.

## 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

**FTP Mode**

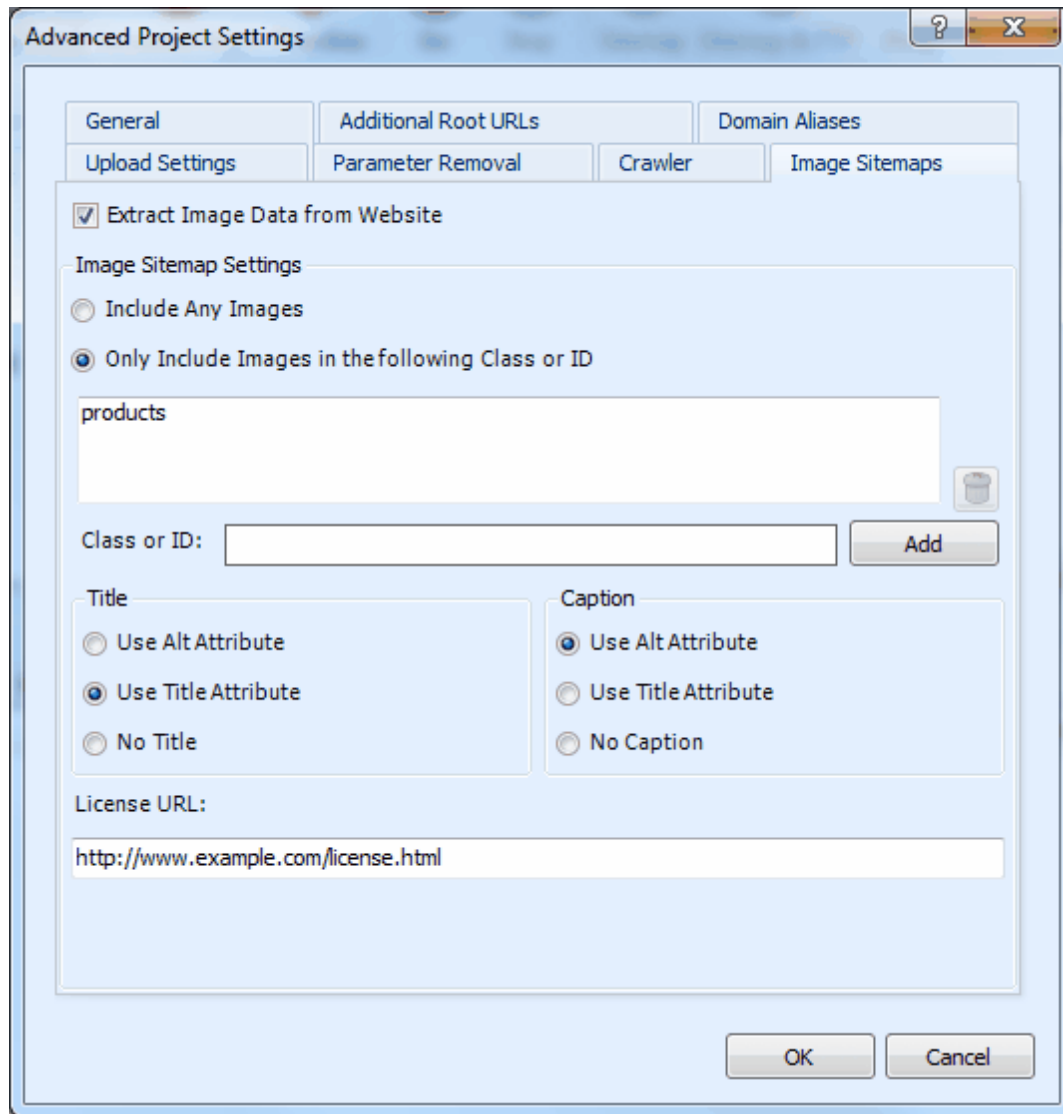
typically 'public\_html' on Linux/Unix servers. This field can be left empty if you do not require changing directories before uploading files to your site. Specifies whether Sitemap Creator should use an Active or Passive FTP connection when communicating with your server. If you encounter problems, try toggling this option. (This option applies to FTP transfer modes only.)

**Sitemap URL**

**Override URL** - *(This field is strictly optional. Leave it blank if you are unsure.)* Override URL is used to override your sitemap file location. By default Sitemap Creator automatically determines the sitemap URL for your project based on the Root URL and the filename you used when saving your sitemap. However, if you're uploading your sitemap to a 3rd party server (so the URL hosting your sitemap does not match the Root URL you've entered into Sitemap Creator) you will need to use the Override URL field. Enter the website address that will host the sitemap file (such as "<http://sitemaps.example.com/>"). Do not include the sitemap filename, Sitemap Creator will still determine that automatically. Changes made here will affect the URL used for "pinging" as well as the sitemap URLs used in the sitemap index file (if you create one).

**4.1.1 Image Sitemap**

Image Sitemaps allow you to provide information about important images on your website to Google and other search engines that support the Image Sitemap extensions. To capture this information Sitemap Creator provides some powerful features to automatically extract this data without any manual intervention or modifications to your website.



The first step to extracting image data from your website is to ensure that "Extract Image Data from Website" is checked. This option is found on the "Images" tab in [Advanced Project Settings](#). Note: Users with very large sites that do not require image data may wish to disable this feature to improve performance.

To determine which images should be included in your sitemap, there are a number of filtering features that will be explained below. By default, Sitemap Creator does not include every single image it finds because only important images (that are unique and relevant to your site's content) should be included. Including images that are used as backgrounds, navigation elements, etc. will only reduce the relevance of your website and could impact how well your sites images are ranked.

The images that are extracted by Sitemap Creator are filtered by their HTML "class" or "ID" attributes and whether or not the "Alt" and "Title" attributes are present. These options are explained below.

**Include Any Images** - This option tells Sitemap Creator to extract imaged that are found in any part of the website, regardless of the class or ID name in which they reside.



**Only Include Images in the following Class or ID** - If you only wish to include images that are located under a specific class or ID name select this option and add the names using the "Class or ID" field and the "Add" button. The images can be located anywhere under the class or ID block. For example, many e-commerce sites wish to include only images from their product catalog. If those images are located in a "div" block with the class name "productinfo", we would add "productinfo" to the list of Class/ID names.

**Title and Caption** - The Image Sitemap protocol specification optionally allows you to provide title and caption text for each image in your sitemap. Sitemap Creator can automatically extract this text from your images' Alt or Title attributes. Selecting the appropriate radio buttons tells Sitemap Creator which attribute to use in each field. If your site doesn't have image title or alt tags, select "No Title" or "No Caption" as necessary.

**Please Note:** If you select to use the "Alt" or "Title" attributes only images with those attributes present will be included in your sitemap.

**License URL (Optional)** - The License URL field can be used to populate the "License" attribute in your sitemap. This URL should point to a page that contains the license information for your images. If you do not have any license information to include with your images, leave this field blank.

## 4.1.2 Soft 404 Detection

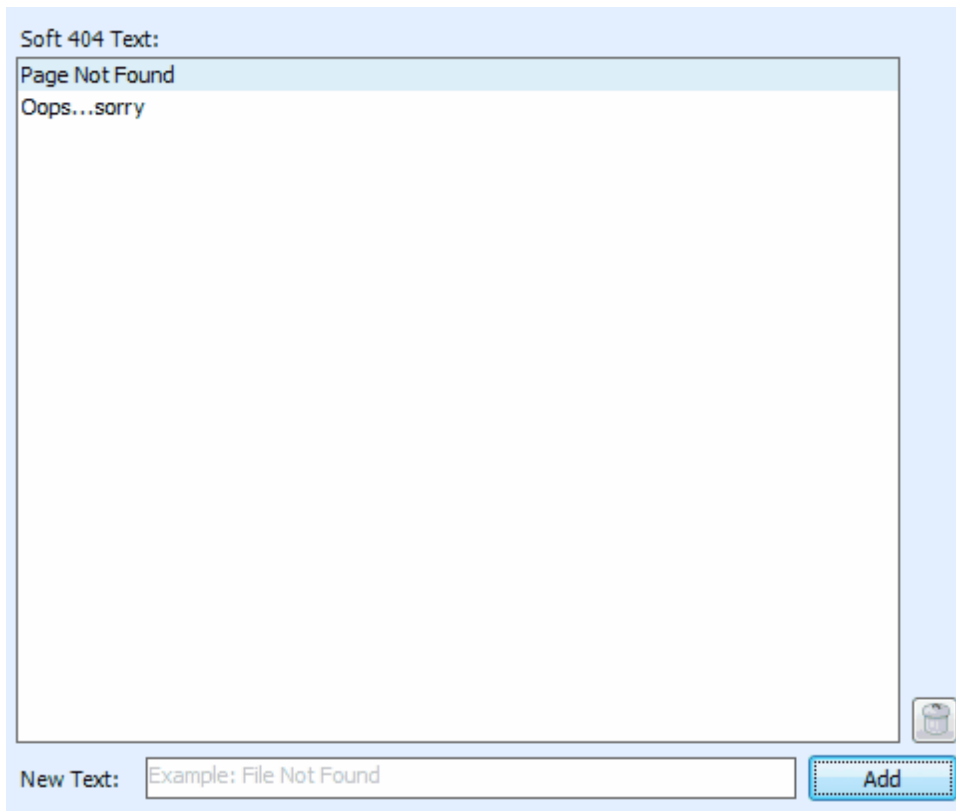
Soft 404 Detection allows Sitemap Creator to detect which links are broken even if your web server does not correctly return a "404 - Not Found" HTTP header. This is useful if you use a content management system that could report "Missing Pages" but don't necessarily return an HTTP status error. A common example of this is when an invalid product ID number is specified in some online shopping cart systems.

If you find that when crawling your website Sitemap Creator crawls lots of incorrect or invalid URLs it's a good indication that you need to configure Soft 404 Detection. Another way to test this is to change your Root URL to point to a page on your website you know does not exist, for example:

`http://www.yoursite.com/thispagedoesnotexist`

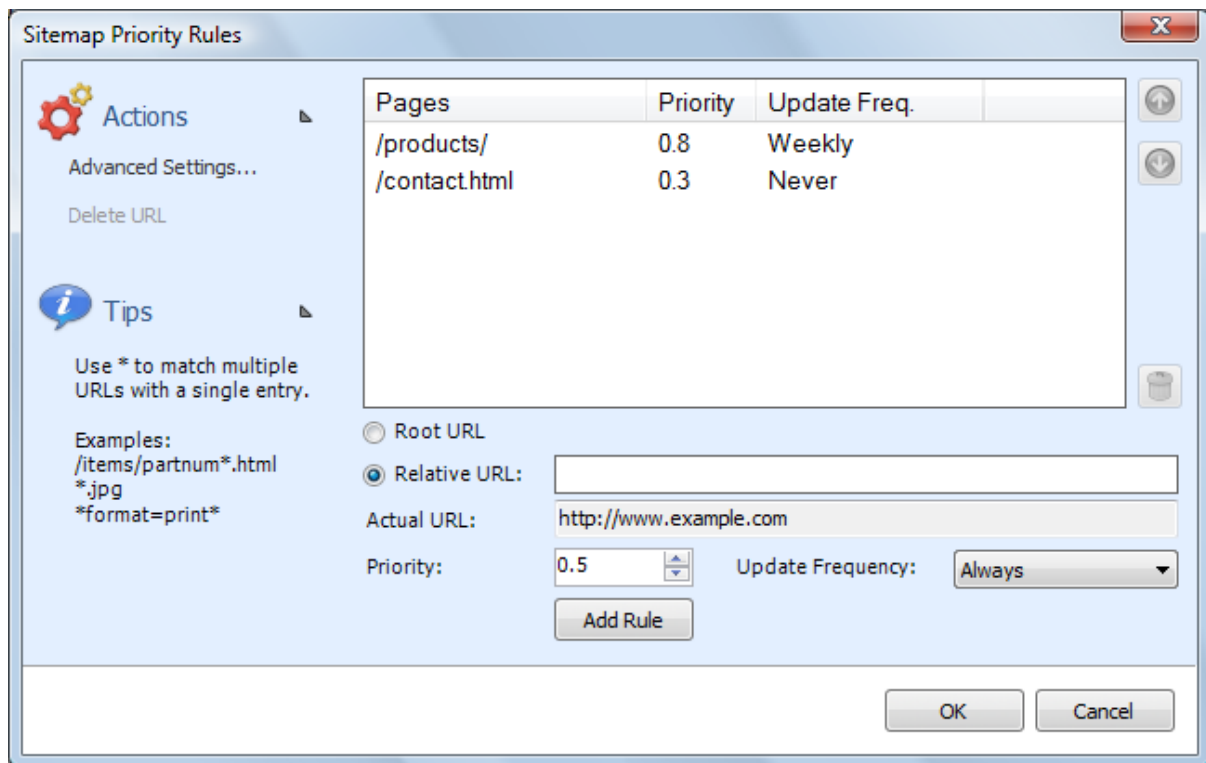
(Replace "yoursite.com" with your actual hostname). If Sitemap Creator doesn't detect an error on this bogus Root URL, then you'll need to configure Soft 404 Detection.

To configure Soft 404 Detection just enter some unique text that appears on your website's error page. In the example screenshot below we have configured "Page Not Found" and "Oops...sorry" as our Soft 404 Text. Any time Sitemap Creator sees a page with this text on it, it will treat it as an error page. To find out what text your website uses, use a regular web browser to access a URL on your website that you know does not exist.



## 4.2 Priority Rules

Priority Rules allow you to specify "Priority" and "Update Frequency" values that should be applied to pages automatically each time your website is crawled. Priority Rules greatly simplify the process of customizing your sitemap, because once the rules are put in place, they allow the Priority/Update Frequency value to be reassigned when your sitemap is recreated. This is especially useful if you are taking advantage of Sitemap Creator's scheduler for automatically updating your sitemap.



To create a new priority rule, enter the Relative URL in the appropriate field (or, if you wish to assign a specific rule to the Root URL, select the "Root URL" radio button). For example, if your website had content located at:

<http://www.example.com/support/>

and you wanted all pages under "support" to have priority 0.3, you would enter "/support" in the relative URL field.

Next, assign the Priority and Update Frequency values that should be applied to pages matching this rule. Click "Add Rule" to save this rule.

Priority Rules are matched from the top down. The first rule to match a page is used. You can reorder them by selecting a rule, then changing its position using the up and down arrows located to the right of the rule list (or press the delete key).

Double clicking on a rule allows you to edit the rule. Click "Add Rule" after editing to save it back to the list. To remove an entry from the rule list, select a rule then click the little "Trash" button to the lower right of the list.

The "Advanced Settings..." Action opens the [Advanced Project Settings](#) where Automatic Priority can be enabled/disabled.

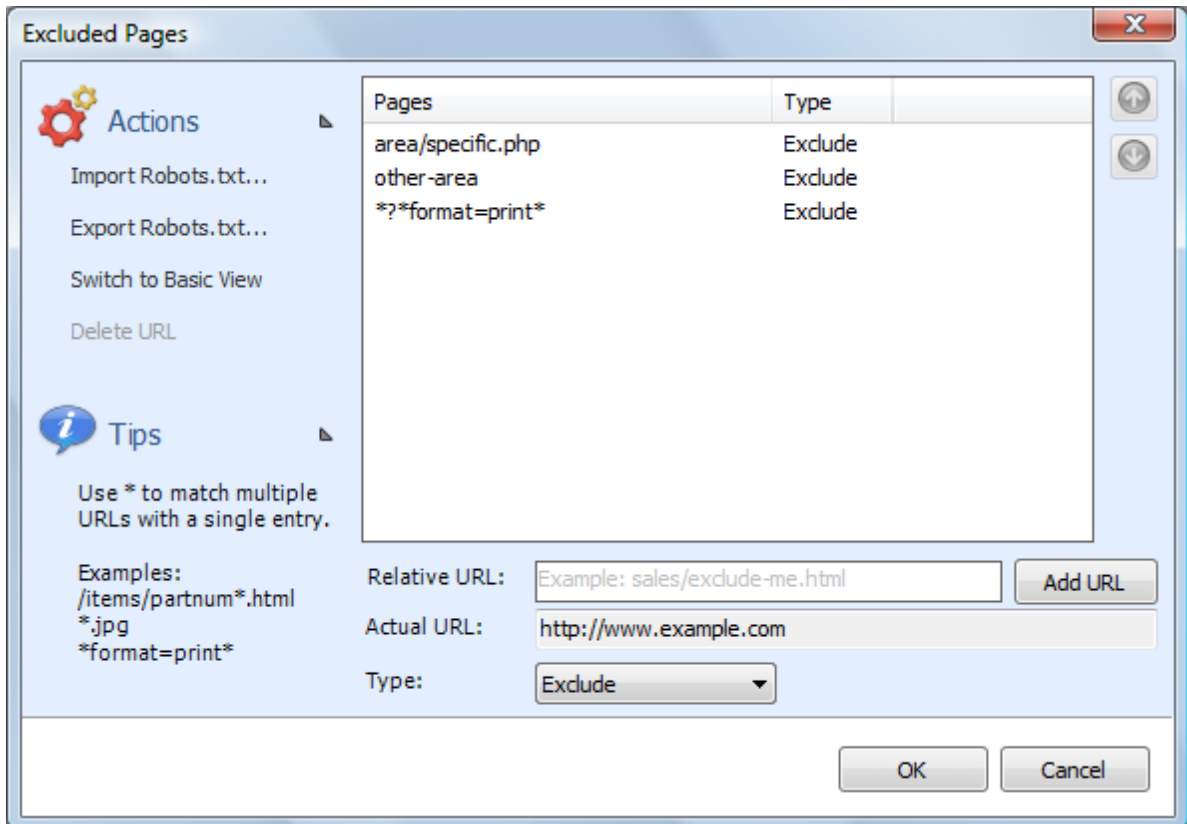
## 4.3 Excluded Pages

Often it is necessary to exclude some sections of a website from crawling. To do this, Sitemap Creator provides the ability to exclude specific URLs or groups of URL from crawling. To access the Excluded Pages window, click "**Project | Edit Excluded Pages**" from the main menu bar.

The Excluded Pages window contains full or relative URLs that should be skipped when crawling. This

is important if you have pages or scripts that generate a large amount of redundant content (such as a message board or article pages that are formatted for printing). You might not want these included in the crawl. To add a new URL to the list of Excluded Pages enter the full or relative path in the **Relative Path** field. (As you type the "Actual URL" field will show how your URL is combined with the project's Root URL.) Click the **"Add Rule"** button to save the new URL to the list.

To remove a URL from the list, select it then click **"Delete URL"** (under Actions), or press the "Del" key on your keyboard.



By clicking the **"Switch to Advanced View"** link on the left hand **"Action"** menu 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.

When a directory (like "/blog") is specified as an Excluded Page, then all of its sub-directories will be excluded (such as: www.example.com/blog/index.html and www.example.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 the Relative Path field.

*inventory/warehouse\*index* will exclude the following:

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

*userreports/\*.txt* will exclude all files with an extension of ".txt"

It is also possible to use the Exclusion List to exclude URLs that contain a particular parameter. For example, to exclude all urls that contain "format=print" in the URL parameter list, you could enter:

*\*?\*format=print\**

This will exclude all URLs that contain 'format=print' somewhere in the URL parameter list (after the '?').

**Advanced Tip:** The \* matches every character including "/". If you want to match any text but **NOT** the path separator ("/"), use an exclamation point in front of the asterisk: "!\*". This special case will match any characters **EXCEPT** the path separator.

The "\$" can be used to match the end of the URL, like this:

*inventory/index.php\$* will exclude the following:  
*inventory/index.php*

*but will NOT exclude:*  
*inventory/index.php?product=xyz*

Internal URLs should be relative to the root URL. URLs beginning with '/' are taken from the domain root (this is recommended). URLs **not** beginning with '/' are relative to your project's Root URL. For example, if the Root URL is "www.example.com/somepage/":

'apage.html' becomes 'www.example.com/somepage/apage.html'  
 '/apage.html' becomes 'www.example.com/apage.html'

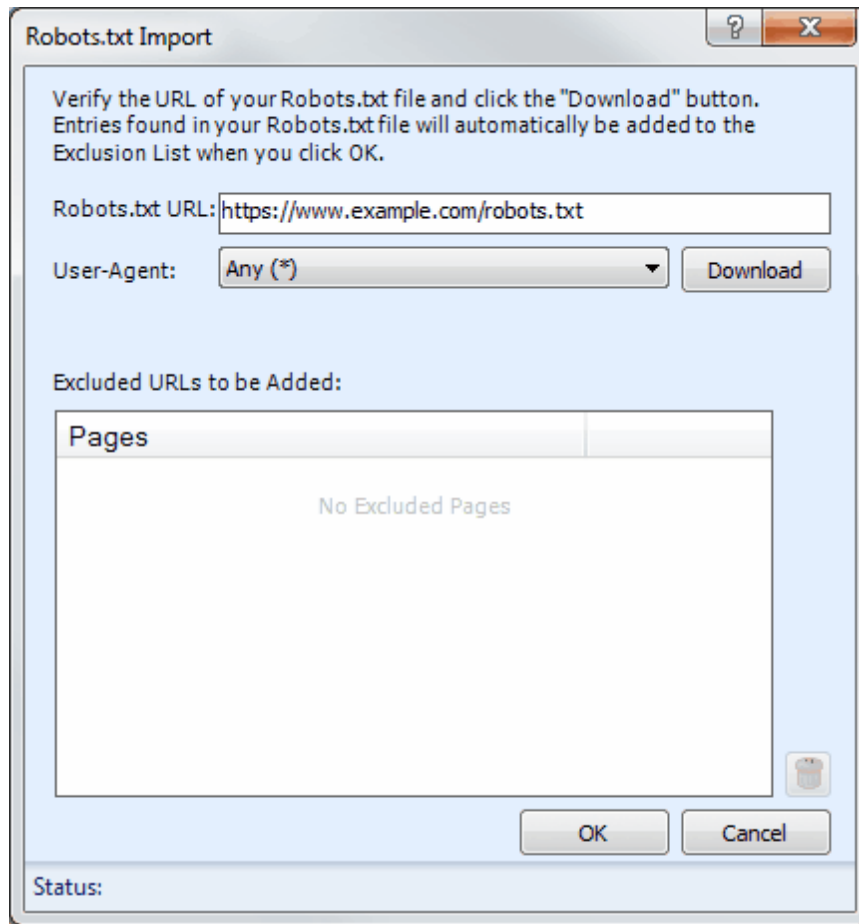
URLs beginning with 'http://' are treated as full URLs, and can be used to exclude links to external sites. For example:

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

As a shortcut, a page can also be added to the exclusion list by right clicking the URL in the Crawl Results.

### 4.3.1 Importing Robots.txt

Some websites already contain a special file called "robots.txt" which tells crawling software which URLs (or paths) to ignore. Sitemap Creator does not automatically obey robots.txt because it is often desirable to crawl the entire website. However, Sitemap Creator has the ability to import a robots.txt file into the Excluded Pages list. To access this feature, click "**Import Robots.txt...**" under "**Actions**" on the [Excluded Pages](#) window.



When the Robots.txt Import window opens, it will automatically populate the "Robots.txt URL" field based on your Root URL. You can modify this field if necessary. Click "Download" to fetch and load the file from the website.

The User-Agent field is used to select which rules you wish to import from the robots.txt file. By default the "\*" user-agent will be loaded (any bot).

Once Sitemap Creator loads the file successfully, the results will be placed in the "Excluded URLs to be Added" list. You can remove any unwanted entries now by highlighting the entry and clicking the "Trash" button (or pressing "Del" on your keyboard). If you click "Cancel" at any time no changes will be made to your Excluded Pages.

When you click "OK", any entries found in the "Excluded URLs to be Added" list will be automatically added to your "Excluded Pages" list.

### 4.3.2 Exporting Robots.txt

In addition to being able to import an existing robots.txt file, Sitemap Creator is also capable of creating (or exporting) it based on the current entries in the Excluded Pages list.

The contents of the Excluded Pages list will be converted into "Allow" and "Disallow" rules and normalized automatically.

The robots.txt file generated by Sitemap Creator must be placed at your domain root. For example, if you Root URL was "[www.example.com/products/](http://www.example.com/products/)", the robots.txt file created by Sitemap Creator must be placed so that it's accessible from "[www.example.com/robots.txt](http://www.example.com/robots.txt)" (not in the "products" sub-folder).

The robots.txt file created by Sitemap Creator will only have entries under the "\*" user-agent field. If you require different rules for different user-agents, you will need to manually edit your robots.txt file in a text editor (such as Notepad).

## 4.4 Passwords and Forms

If your website is password protected or contains content hidden behind a form submission, you can configure Sitemap Creator to access that content using the "Passwords and Forms" feature. To access the Passwords & Forms window, click "**Project | Passwords & Forms**" from the main menu bar.

Passwords and Forms allows you to create rules which Sitemap Creator will use each time it crawls your website. If a page is encountered with a form (or POST) rule, it will post the specified form data (usually a username and password) to the server when that page is encountered.

Page	Method	Credentials
/admin/	HTTP	example:*****
dologin.php	POST	username=user&passw...

Refer to the Excluded Pages section of the manual for more information on how to format the Relative URL field.

The "Method" field tells Sitemap Creator what type of credentials to supply to the remote server. There are two methods for supplying login information to a website, POST and HTTP.

### Post Method

This is the most common method of logging into a website. If you enter your login information by

filling out some type of web based form, then use this method. To configure the login credentials click the '**Form Wizard...**' link under **Actions** to launch the Inspyder mini-browser. Use this browser to log into your website. When you complete the login form, Sitemap Creator will ask if you want to save the form data, click 'Yes' and Sitemap Creator will automatically create the necessary login rule for you.

#### **Prompt for Manual Login**

If you don't want Sitemap Creator to store your login credentials, you can enable the '**Prompt for Manual Login**' option. This option only works with websites using cookies for authentication (most websites using form/post based authentication also use cookies). When you click the 'Go' button to start crawling your site Sitemap Creator will launch the mini-browser to capture your login session on-the-fly. When you close the mini-browser Sitemap Creator will resume crawling with your logged in session.

#### **HTTP Method**

The second method is the HTTP method. This is used when the server (rather than the website) is providing the authentication method. If you see a popup login window (with username and password fields) when you access your website with Internet Explorer or FireFox, then your website is using this authentication method. To configure this rule, enter the Relative Path of your login page (if it is your Root URL, just enter "/"), along with the username and password in the corresponding fields (the domain field is optional). Click 'Add' to save the rule.

To remove a rule, select the rule to remove from the list, then click the "**Delete Page**" link under "**Actions**".



## 5 Crawl Results

The Crawl Results section contains 5 tabs: Sitemap Info, Updated Pages and Log

### Sitemap Pages

The contents of this tab provide information on what your sitemap file will contain. Each crawled page will appear here along with the frequency, last updated and priority values. Any pages matching a "Priority Rule" rule will have their new priority and update frequency values applied here. Any changes you make to the Priority Rules 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 Priority Rule from URL** - This option will automatically create a Priority Rule 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 Pages list

### Sitemap Images

This tab provides the image data that was extracted from the website that can be included in the XML file if "Image Sitemap" is selected on the Sitemap Setting page. For more information on extracting sitemap image data, see [Image Sitemap settings](#).

### Sitemap Videos

This tab provides the video data that was extracted from the website that can be included in the XML file if "Video Sitemap" is selected on the Sitemap Setting page. Embedded videos are automatically detected on your site.

### Updated Pages

The Updated Pages tab contains a list of each page that changed since the last time the site was crawled. When crawling a website for the first time, all pages will be listed as "Added".

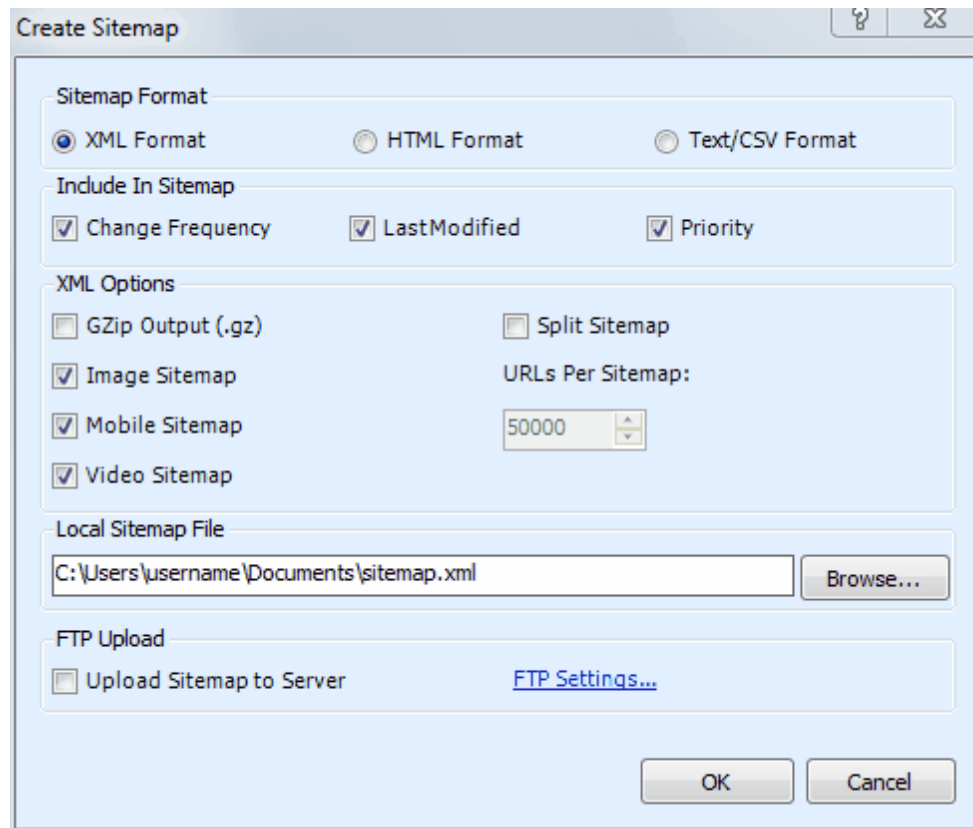
### Log

The Log tab contains information that can be useful when troubleshooting crawling problems. The log indicates each time a crawl starts and stops, as well as information regarding errors that have occurred. To save the log to a file, right click and select "Save Log...".

To clear all results from the Crawl Results section, click the "**Clear Results**" button located in the lower right hand corner of the Sitemap Creator window.

## 6 Creating a Sitemap

To create your sitemap after crawling your website with Sitemap Creator, click either the "Sitemap" or "Sitemap & FTP" buttons. This will open the "Create Sitemap" window which will allow you to fine tune how your sitemap is saved.



### Export Format

**XML Format** - The XML Format sitemap conforms to the standards listed on [www.sitemaps.org](http://www.sitemaps.org). The output will always contain the "loc" tags, and depending which Components are selected, the "changefreq", "lastmod" and "priority" tags will 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 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.

**Text/CSV Format** - The Text/CSV format saves the sitemap contents to a simple comma separated text file that can be opened with Microsoft Excel, Notepad or any other text editor. The file will contain priority, update frequency and/or last modified data depending on which "Components" are selected.

### Include In Sitemap

These settings are specific to the XML and Text/CSV formats. These options select which fields are included in the sitemap file. (The URL is always included for each entry in the sitemap because it is a required field.)

To create a simple list of URLs, select the "Text/CSV" format, and uncheck all the optional components. This will produce a text file with a list of URLs from your website.

### XML Options

**GZip Output** - If this option is checked the sitemap file will be compressed into a gzip file. This option can significantly reduce the size of your sitemap, and is compatible with the XML sitemap standard. Typically this option is only used if you have a large sitemap or a slow Internet connection.

**Split Sitemap** - If this option is checked, Inspyder Sitemap Creator will automatically generate multiple sitemap files and a sitemap index file based on the "URLs Per Sitemap" field. Inspyder Sitemap Creator will automatically number the individual sitemap files, and save the sitemap index in the file specified in the Sitemap Filename field.

For example, if the sitemap filename was "sitemap.xml", then the individual sitemap files would be named "sitemap1.xml, sitemap2.xml, etc.", and the sitemap index would be saved to "sitemap.xml".

**Image Sitemap** - If Inspyder Sitemap Creator was able to extract relevant image data from your website, you can include it in your sitemap file by checking this option. See [Image Sitemap](#) for more information on how to configure Inspyder Sitemap Creator to automatically detect and capture image information.

**Mobile Sitemap** - Check this option if the URLs contained within this sitemap contain content that is specially formatted for mobile devices. **Note:** You should NOT check this option if your website does not contain special content. If your server returns specialized content and URLs based on the browser's User-Agent, consider changing Inspyder Sitemap Creator's User-Agent value in [Advanced Project Settings](#) and creating a special sitemap with Mobile only content.

**Video Sitemap** - If Inspyder Sitemap Creator was able to extract relevant video data from your website, you can include it in your sitemap file by checking this option.

### XML Options

If the "Upload Sitemap to Server" box is checked, Inspyder Sitemap Creator will automatically transfer your generated sitemap to your server. For this option to work, you must configure your FTP settings in

Advanced Project Settings. (You can click the "FTP Settings..." link to jump to this screen.)

## 6.1 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)  
#DESCRIPTION#   The page description specified in the description meta tag  
#N#  
#LASTMODIFIED# The date the page was last changed  
#ED#  
#FILENAME#     The filename of the page (without the path)  
#PROJECT#      The name of the project

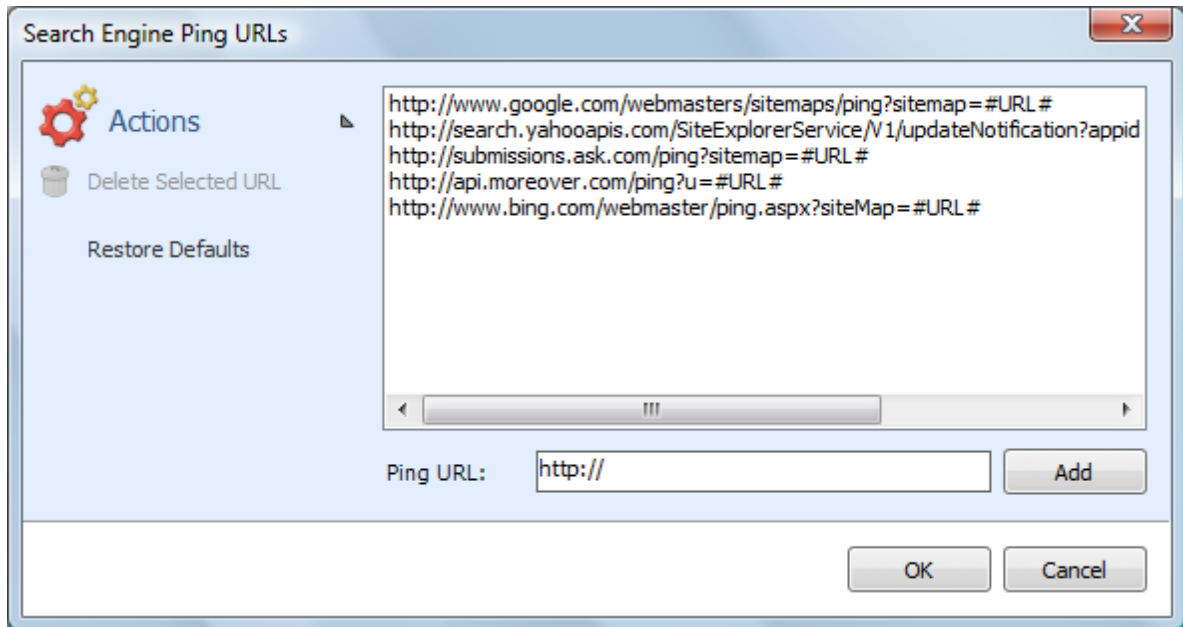
The HTML report templates must be placed in the application's "Templates" directory. By default, it's located at:

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.

## 7 Ping URLs

After creating your sitemap, Sitemap Creator can let search engines know that your sitemap file has been updated. This is called "Pinging". It is possible to configured which search engines Sitemap Creator pings by clicking 'Settings | Edit Ping URLs' from the main menu. The ping settings apply to every project configured in Sitemap Creator.



The Search Engine Ping URLs window allows you to add or remove search engine ping URLs. To add a new URL, enter the address in the Ping URL field and click **Add**. Use **#URL#** in the URL itself to indicate where you want your sitemap's URL to be substituted. Sitemap Creator will automatically apply the necessary encoding to the **#URL#** value.

Sitemap Creator combines your sitemap filename to your project's Root URL, and uses the result as the **#URL#** value. Sitemap Creator is smart enough to strip any trailing filename from your root URL (if you have entered one).

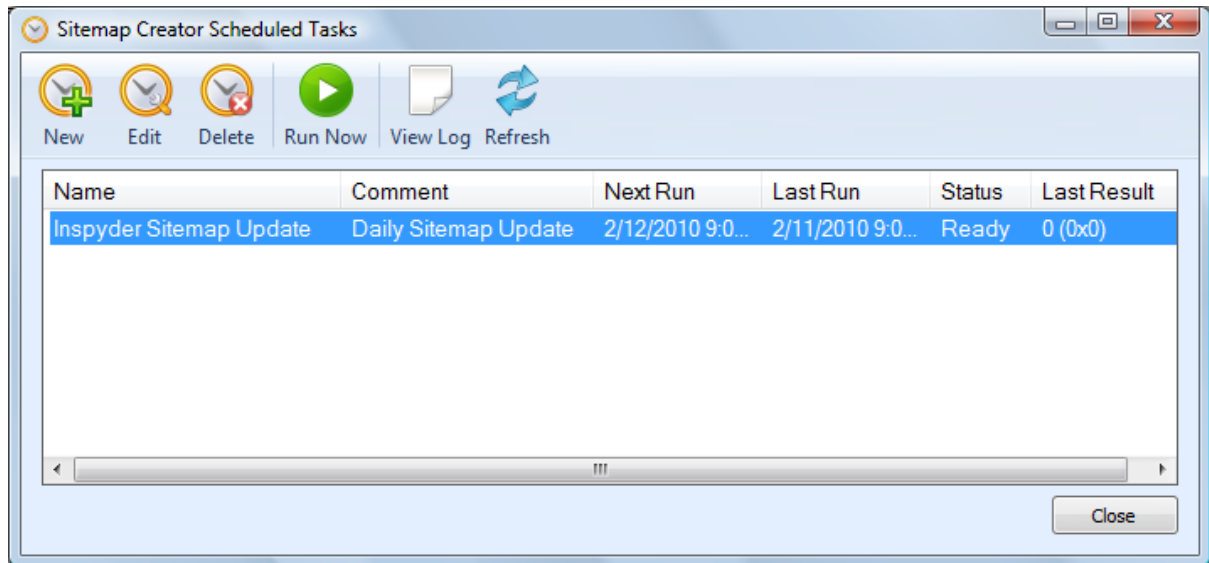
For example: If your root URL is 'http://www.example.com/index.shtml' and you export 'mysitemap.xml' Sitemap Creator will create the **#URL#** parameter as: 'http://www.example.com/mysitemap.xml'

To edit an existing Ping URL, double click it in the list of URLs. To remove a URL, select it, then click **Delete Selected URL** under the "Actions" list.

The **Restore Defaults** action will reset the ping URL list to the default settings.

## 8 Scheduler

Sitemap Creator has the ability to run in an unattended mode and automatically export the results. To configure Sitemap Creator for scheduled operation simply select "**Settings | Scheduler**" from the main menu bar.



### 8.1 Scheduling Basics

Scheduling Sitemap Creator to run automatically is done by creating one or more tasks in the Sitemap Creator Task Scheduler. Each scheduled task contains the settings that will be used when running Sitemap Creator automatically. Basic settings include:

- The Sitemap Creator Project to load before crawling
- The date and time the task will run next
- How often the task should be re-run

Tasks are run by the Windows Task Scheduler, so Sitemap Creator does not need to be open for the scheduled tasks to run, but your computer must be turned on (it will not wakeup automatically).

Tasks configured within Sitemap Creator are visible in the Windows Task scheduler. For scheduling options more advanced than those provided by Sitemap Creator, please see the Windows Task Scheduler documentation included with Windows.

Name	Comment	Next Run	Last Run	Status	Last Result
Sample Task	Nightly Crawl	1/4/2010 12:00:00 PM	N/A	Ready	1

Any scheduled tasks will appear in the scheduler window. The list of scheduled tasks shows the following information:

**Name** - The user-friendly name of the scheduled task

**Comment** - Any comments that were entered when this task was created

**Next Run** - The next time the task will automatically run

**Last Run** - The last time the task ran automatically

**Status** - Indicates if the Task is "Ready" (ready to run), "Running" (is currently running) or "Disabled" (will never run)

**Last Result** - The result code of the last run. If this number is not zero then an error may have occurred (the Scheduler Log may have more details on any errors that occurred).

## 8.2 Adding/Editing a Task

To add a new task to the Sitemap Creator Task Scheduler, click the **"Add"** button. The scheduling wizard will launch, which will walk you through the steps to creating your new task. When you are finished, click "Close" to save your task to the Windows Task Scheduler. The new task will appear in the list of scheduled tasks.

You can edit a task at any time by highlighting the task item in the list and clicking the **"Edit"** button. This will launch the same scheduling wizard as before with your corresponding settings already populated. Please note, if you have made changes to your scheduled tasks outside of Sitemap Creator then those changes will be lost when you edit your task with the scheduling wizard.

It is possible to force a task to run immediately by selecting the Task item, then clicking the **"Run"** button.

Tasks can also be set to **"Disabled"** by right clicking and then un-checking the **"Task Active"** item. When a task is disabled its settings are still saved by the scheduler, but it will not run automatically.

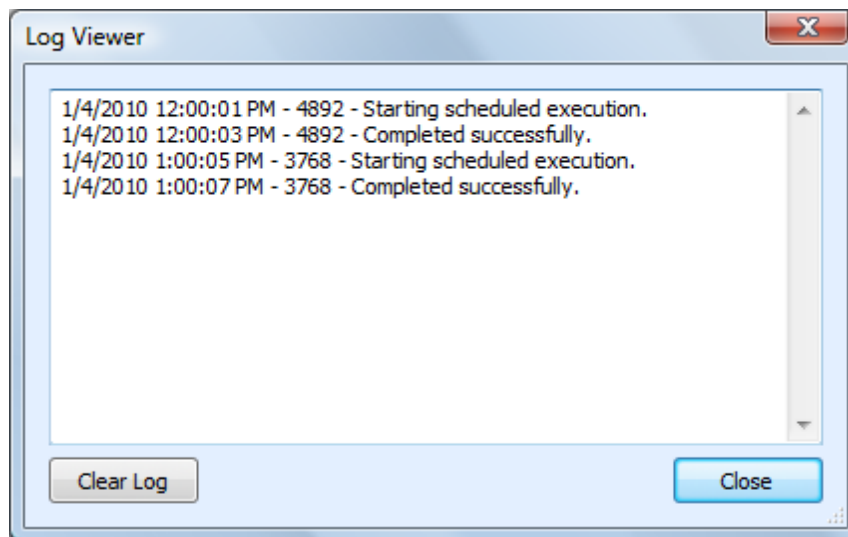
## 8.3 Deleting a Task

To remove a task from the scheduler, select the task item from the list of tasks and click the **'Delete'** button. This will permanently remove the task from the Windows Task Scheduler, but will not remove any projects, reports or other data associated with that task.

## 8.4 Scheduler Log

The scheduler log provides a mechanism to review what happened while a scheduled Task was running. To view the log click the **'View Log'** button.





Each line in the log file provides the following information:

- Date/Time the event occurred
- Process ID of the running Task (this is useful for tracing events if you have multiple scheduled tasks running at the same time)
- The log message

To remove all entries from the log, click the "**Clear Log**" button.

## 8.5 Command Line Interface

If you wish to run Sitemap Creator from the Windows command line interface, the following parameters can be used. These parameters are also used by the Task Scheduler when scheduling tasks. If you want to manually adjust your scheduled tasks in the Windows Task Schedule refer to these parameters.

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>-mobile</b>	(Optional) If this parameter is set Sitemap Creator will generate a mobile sitemap (note this option only applies to the XML sitemap format)
<b>-image</b>	(Optional) If this parameter is set Sitemap Creator will generate an image sitemap (Note: this option only applies to the XML sitemap format)
<b>-video</b>	(Optional) If this parameter is set Sitemap Creator will generate a video sitemap (Note: this option only applies to the XML sitemap format)
<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; CSV, 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)
<b>-nocrawl</b>	(Optional) If this option is specified, Sitemap Creator will not crawl, just load the project and generate a report. This command line option should only be used when you want to generate a second sitemap format without recrawling (you would run Sitemap Creator twice from the command line, once as usual, and a second time with this parameter specified).

Note that if the command line argument's parameter contains one or more spaces, then the enter argument must be enclosed in quotes. (i.e. "-output=My Report.csv")

## 8.6 Result Codes

These are the codes returned to the 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 Export Error  
**Solution:** Try running the project and exporting in interactive mode. Check for file in-use, disk full and other possible 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.).
- 4 Email Error  
**Solution:** An error occurred while trying to send the email notification. Check your Email Settings or check the scheduler log for more details.
- 5 Other Error  
**Solution:** Check the scheduler log for more details.

## 9 Email Settings

You can customize the emails that Sitemap Creator sends when scheduled crawling has completed. The customization screen is accessed by clicking on the "**Settings | Email Settings**" menu item.

### 9.1 Server Settings

Configuring your email server settings (SMTP) allows Sitemap Creator to send automated emails. To access your mail server settings, click "**Settings | Email Settings...**" from the main menu bar.

The screenshot shows the 'Email Settings' dialog box. It features a sidebar on the left with a gear icon and the title 'Actions', containing two options: 'Test Email Settings' and 'Restore Defaults'. The main content area is divided into several sections: 'SMTP Server' with a text input field containing 'smtp.example.com'; 'Security Mode' with a dropdown menu set to 'None'; 'SMTP Username' and 'SMTP Password' with text input fields, both containing the placeholder text '(Required by some ISPs)'; 'Sender Address' with a text input field containing 'someone@example.com'; 'Subject' with an empty text input field; and 'Message Body' with a large, empty text area. At the bottom right of the dialog are 'OK' and 'Cancel' buttons.

#### Settings

**SMTP Server** - This is the hostname of your outgoing (SMTP) mail server. If you are unsure what to use, check with your Internet service provider (ISP) or your email provider (such as Google if you use Gmail).

If your provider uses a non-standard port (25 by default, or 465 if SSL is enabled) you can append it to the hostname by separating it with a colon (for example: "mail.example.com:1234").

**Security Mode** - If your provider uses SSL (Implicit SSL) or TLS (Explicit SSL) select that option here. SSL mode uses port 465 by default. "None" and TLS modes use port 25.

**SMTP Username and Password** - This is your SMTP username and password. Use these fields if your service provider requires SMTP authentication. If no username and password are supplied Sitemap Creator will not attempt to use SMTP authentication.

**Sender Address** - This is the email address that Sitemap Creator will use as the "From:" address for any emails.

**Subject and Body** - See the Message Body section in the manual for details.

### Actions

**Test Email Settings** - This action tests your configuration by sending an email to the email address specified in the **Sender Address** field. If you receive this message then your email settings are correct. If an error occurs or you do not receive the email, verify your settings and try again.

**Restore Defaults** - This action resets the **Subject** and **Message Body** fields to their default values.


## 9.2 Message Body

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 Sitemap Creator 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 unique files (URLs) crawled on your website
#TOTAL_PAGES#	The total number of pages included in the sitemap (HTML and PDF pages unless you have "Include All Content" enabled)
#TOTAL_IMAGES#	The number of images that are included in the sitemap ("N/A" will be displayed if an image sitemap is not selected)
#TOTAL_VIDEOS#	The number of videos that are included in the sitemap ("N/A" will be displayed if a video sitemap is not selected)
#TOTAL_UPDATED#	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
#UPLOAD_RESULTS#	Indicates the status (success or failure) of FTP upload (if applicable)

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

Email Settings

 **Actions**

- Test Email Settings
- Restore Defaults

SMTP Server:

Security Mode:

SMTP Username:

SMTP Password:

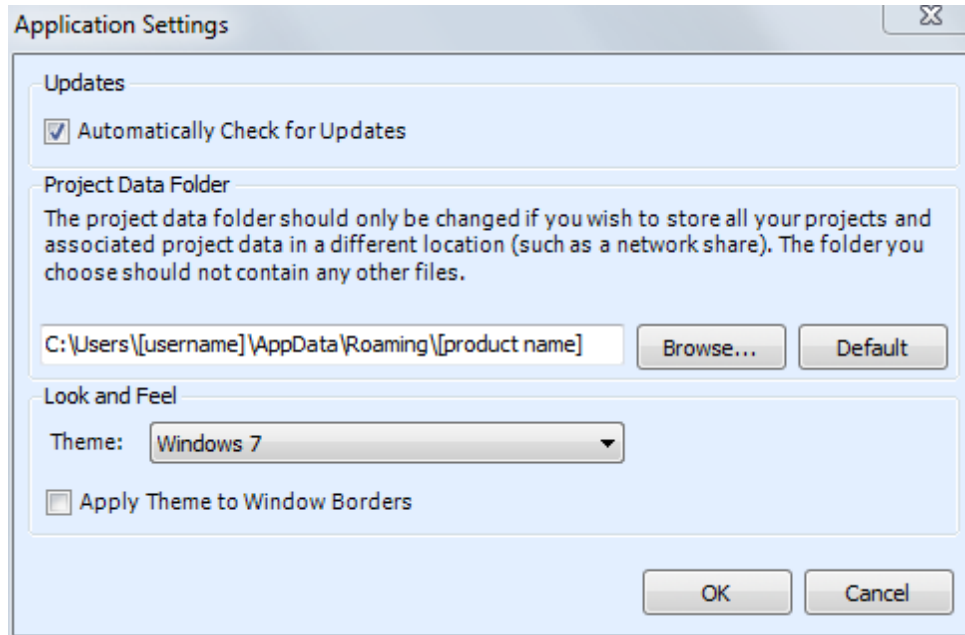
Sender Address:

Subject:

Message Body:

## 10 Application Settings

The Application Settings window allows you to change application wide settings that do not affect individual projects. You access this window from the main menu, under "Settings | Application Settings...".



**Updates** - If this option is enabled, Sitemap Creator will automatically check for a new version each time it runs. Unselect this option if you do not wish to see the "A new version is available" message when you start Sitemap Creator.

**Project Data Folder** - This setting allows you to change where Sitemap Creator stores its project and data files. This is useful if you wish to change the default location to a shared network location, or to a folder that is easier for you to backup. Note that you must restart Sitemap Creator for any changes to take effect. If you change the Project Data Folder your existing projects will not be moved (you must move them manually). Finally, the folder you select as the Project Data Folder should not be used for anything else. It should only contain projects and files created by Sitemap Creator.

**Look and Feel** - These options allow you change the theme of Sitemap Creator. "Apply Theme to Window Borders" allows you to over-ride the Windows border with a style that matches the current theme.

## 11 Project Import and Export

Sitemap Creator provides the ability to import and export projects. These features are useful if you need to move projects to a different PC, wish to backup your existing projects, or copy projects between different Inspyder applications.

### 11.1 Exporting Project Files

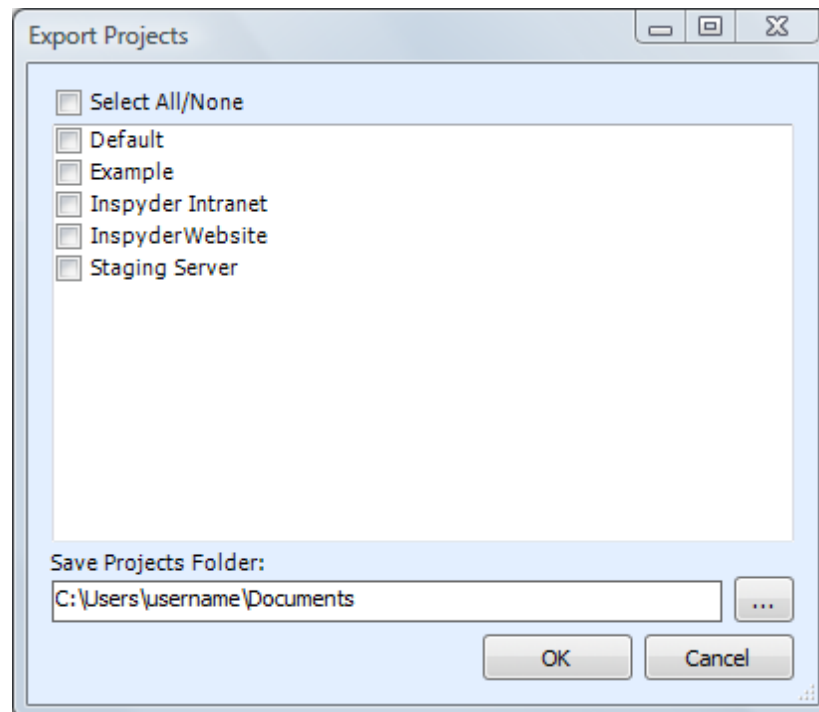
To export one or more projects, click "**File | Export Project Files...**" from the main menu bar. The Export Projects window will open. Here you can select one or more project files to export and the location where they will be saved.

Each exported project is saved to a unique file. If you chose to export multiple projects, multiple files will be created in the selected output folder. The filenames will have the following format:  
"Project Name"."Application".projx

"Project Name" is the name of the project in Sitemap Creator. "Application" will be the Inspyder application abbreviation and is used to determine which application generated the project export. All Inspyder project files end with the ".projx" file extension.

Exporting projects from Sitemap Creator will not remove them from your list of saved projects. Exporting a project makes a copy of the project file. Any changes made to the project file will not be reflected unless you Import the project back into Sitemap Creator.

If you wish to remove a project after exporting it, right click the Project name in the list of Saved Projects and select 'Delete Current Project'.



## 11.2 Importing Project Files

To Import one or more project files into Sitemap Creator, select "**File | Import Project Files...**" from the main menu bar. An Open File window will appear that will enable you to select one or more Inspyder project files.

It is possible to import a project file from an Inspyder application other than Sitemap Creator. If you do, only the common elements between applications will be imported (such as Root URL, Excluded Pages, etc.). Application specific settings will be ignored. If you are importing a Sitemap Creator project back into Sitemap Creator, then all settings will be used.



## 12 Default Project Settings

Default Project settings specify the initial configuration used for any newly created projects. Options that are unique to every Project (such as Root URL) cannot be defined as a default setting, however common options (such as Crawler Timeout) can be configured here.

Changes made to the Default Project Settings will also effect **current projects** that use the **same default values**. For example, if you have 10 projects already created in Sitemap Creator and you have not changed the "Crawler Delay" setting in any of them (so they are still using the original "0 seconds" value), then those projects are still using the "Crawler Delay" value from the Default Project. Now, if you change the crawler delay in the Default Project from 0 to 1 second, that change will affect all 10 of your current projects.

If you had another project where you had explicitly changed the Crawler Delay, for example, from 0 to 5 seconds, then any change made to the Crawler Delay in the Default Project Settings would not affect this project. This project had the "Crawler Delay" changed within it, so the Default Project Setting value is ignored.

To edit the Default Project Settings click '**Settings | Default Project Settings...**' from the main menu bar. The settings in the Default Project Settings window are nearly identical to those found in **Advanced Project Settings**. For details on what each specific options means, please review the settings in the 'Advanced Project Settings' section of the manual.

## 13 About Inspyder Software

Founded in 2004, Inspyder Software is a leading provider of web crawling technologies for content analysis, SEO and website management. These technologies have formed the foundation of our software products and made 'Inspyder' a recognizable brand in the software industry.

Our products are used all over the world by large and small businesses, from independent web designers to Fortune 500 companies. Our user base includes customers in over 50 countries around the world.

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

## 14 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), DVD(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 license.

**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 another computer provided the Product is used only on a single computer. Installing the Product in a shared environment such as a Terminal Server or otherwise making it accessible to more than the single computer, even if only one user is using the Program at any given time, violates this license agreement. 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 © 2014 Inspyder Software Inc.



Inspyder Software Inc.  
200-3342 Mainway,  
Burlington, Ontario,  
Canada, L7M 1A7