Inspyder Web2Disk
User's Reference Manual

Inspyder Web2Disk (Web2Disk)
User's Reference Manual

Copyright © 2015 by Inspyder Software Inc.

Table of Contents

1   Introduction

Web2Disk from Inspyder Software Inc. is a Windows based utility that enables you to download entire websites to your computer for offline browsing. Web2Disk provides an easy way to back up a website, put a website on CD, or simply take a website where an Internet connection is not available.

Web2Disk is compatible with Apache, IIS and other web server software. It can download websites created with PHP, ASP, JSP or any other technology. Just enter a URL and let Web2Disk do the rest.

Features and Benefits

- Easy to Use — Just enter a website URL and click "Go!"
- Offline Browsing — Web2Disk fixes downloaded content for easy offline browsing.
- Scheduled Website Downloads — Take a website with you, where no Internet connection is available!
- Monitor a Website for Updates — Configure Web2Disk to email you when a site is changed.
- Download Dynamic Pages — Download database driven websites with ease. Web2Disk converts dynamic content to static content for offline browsing.
- Powerful Filtering — Save bandwidth by excluding the files, pages and folders you don't need.

Web2Disk requires Windows XP/Vista/7 or Windows Server 2003/2008.

2   Quick Start Guide

When you first run Web2Disk the default project is loaded but it does not contain any settings. The following steps will guide you to setting up Web2Disk to download your first website.
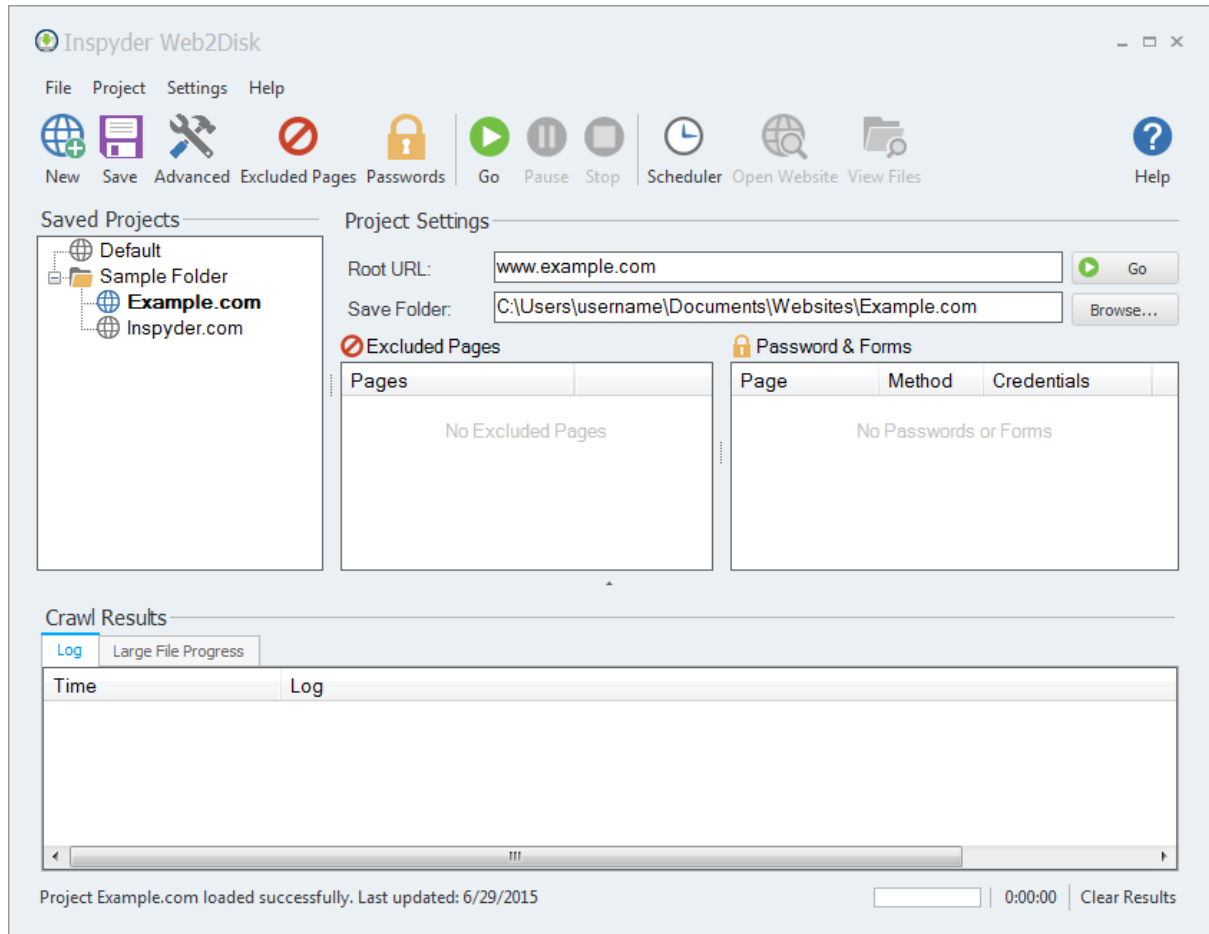
Step 1   Enter the URL of the website you wish to save in the '**Root URL**' field.
Step 2   Enter the folder where you wish to save the website to in the **'Save Folder'** field (or click the "**Browse...**" button to open the folder browser).
Step 3   Click the **"Go"** button.
Step 4   When crawling is finished, click the '**Open Website**' button on the toolbar to see the saved website!

As Web2Disk crawls the website the "Crawl Results" field will show each file as it is downloaded. When Web2Disk is finished it will reprocess the links in the website so that you will be able to browse it with any web browser, directly from your hard drive.

To download a different website create a new project ("File | New") or simply change the Root URL.

3   Toolbar

This section describes the functions of the toolbar shortcuts. These are the most often used features of Web2Disk 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.).
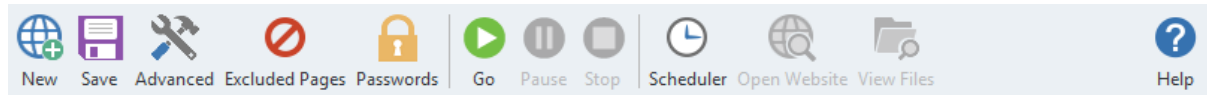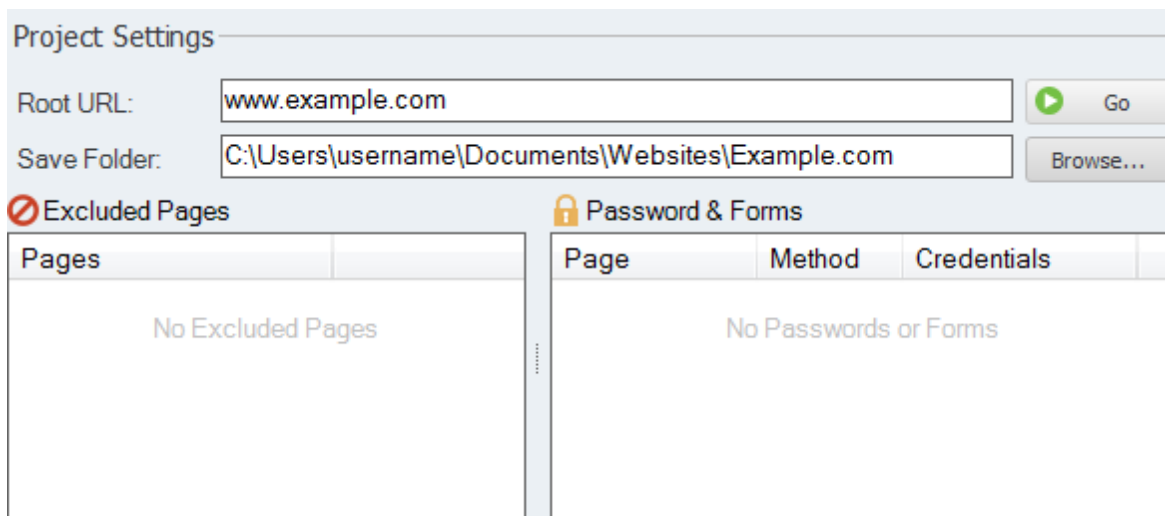
- **New** - The **New Project** button is used to establish a new website that you wish to download. You will be prompted for a shortcut name, which will appear in the Projects list, and the root URL of that web site.
- **Save** - The **Save Project** button is used to save the current project settings. 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 of this document.
- **Excluded Pages -** Opens the Excluded Pages window which can be used to control which pages the crawler visits on a website.
- **Passwords** - Opens the Passwords and Forms window which is used for configuring access to password protected websites.
- **Go** - This button starts the crawling process. The crawl can be interrupted at any time by clicking on the **Stop** button.
- **Pause** - This button pauses the crawler process (allowing you to sleep or hibernate your computer). You cannot close Web2Disk while a project is paused.
- **Stop -** Use this button to stop the crawling process.
- **Scheduler -** Opens the scheduler interface for automatically capturing websites on a recurring basis.
- **Open Website** - Launches the local copy of the website in your default browser. This button will be disabled if the website associated with the current Project has not been downloaded.
- **View Files** - Opens the directory where your website files have been stored in Windows Explorer. To burn this website to CD (or copy to USB), use the contents of this folder.
- **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.

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.

Project Settings

Root URL: www.example.com ● Go

Save Folder: C:\Users\username\Documents\Websites\Example.com Browse...

🚫 Excluded Pages   🔒 Password & Forms

Pages		Page	Method	Credentials	
No Excluded Pages		No Passwords or Forms			

Root URL

The Root URL field contains Web2Disk's initial starting point on a website. Web2Disk will only download from the Root URL and deeper into a website.

Save Folder

This is the local folder where the local copy of the website will be saved.

Maximum Link Depth

This value specifies how many "clicks" deep into the site Web2Disk should download. If this value is set to 0 Web2Disk will download the entire website. This option is typically used on very deep sites when only some content is required.

Maximum File Count

This value specifies how many files to download from the website. If this value is set to 0 Web2Disk will download the entire website. This provides an alternative way to restrict how much of a website is downloaded.

Excluded Pages

This list shows any rules that are created to exclude pages or sections of a website. More information is available in the Excluded Pages section.

Passwords & Forms

This list shows any passwords or form submission data that is configured for this website. More information is available in the Passwords and Forms section.

4.1   Advanced Project Settings

The Advanced Project Settings provide you with a way to further refine how Web2Disk downloads a website. These settings are project specific. To access these settings click "Project | Advanced Project Settings" on the main menu bar, or click the "Advanced" button on the toolbar.
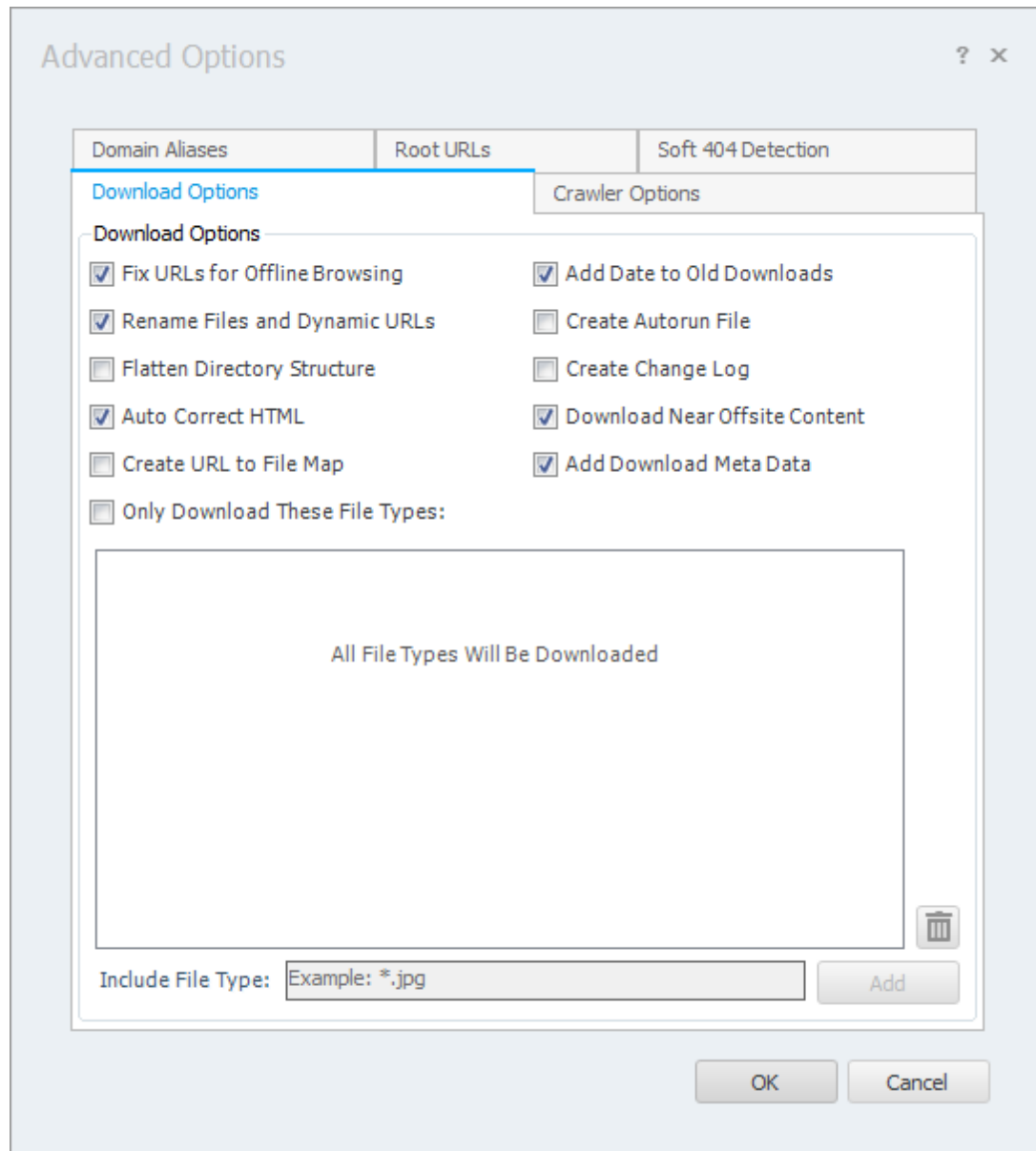
Download Options

Fix URLs for Offline Browsing
If this option is checked Web2Disk will go through the downloaded content and re-write any internal links so that they point to the downloaded files on your hard drive. If this options is unchecked, the downloaded content will be left "as-is", offline browsing may or may-not work. The default value is checked.

Rename Files and Dynamic URLs
If this option is checked Web2Disk will automatically rename file extensions so that Windows makes the correct file-type association when opening the offline files. Dynamic URLs (with parameters, such at http://www.example.com/products.aspx?productID=xyz&view=normal) are also renamed to over come any file name limitations of Windows.

If this option is unchecked, the files will be saved with the same names used by the server, but dynamic URLs will no longer be downloaded. Additionally, Windows may fail to open some file types correctly. The default option is checked.

Flatten Directory Structure

If this option is checked, all the downloaded files will be stored in the same folder. Files with duplicate names, such as:
- http://www.example.com/**index.php**
- http://www.example.com/support/**index.php**

will be over-written by the last file with that name that is downloaded. This option is useful if you wish to strip the content files from a site (such as images, CSS, etc.) but are not interested in the HTML files.

If this option is selected, it is not possible to use the 'Fix URLs for Offline Browsing' feature.

Add Date to Old Downloads

If this option is checked each time a website is downloaded, any previously downloaded copies of that site will be automatically renamed (with the date appended to the download directory in ISO format, YYYY-MM-DD). If the same website is downloaded multiple times on the same day, the time will also be appended to the filename.

If this option is unchecked, then any new downloads will automatically overwrite the previous copy.

Create Autorun File

If this option is checked Web2Disk will create an "Autorun.inf" file in the download directory. If this file is burned to CD-ROM with the website it will cause the website to automatically start when the CD is inserted into a Windows based PC.

Create Change Log

If this option is checked Web2Disk will create a file in the save path named "Changes.txt". This tab delimited file will include any files that have been added, removed or updated since the last time the website was crawled. The file includes the type of change, the original URL, and the file path where the offline copy of the file was saved.

Auto Correct HTML

If this option is checked Web2Disk will automatically attempt to correct errors in the HTML (such as missing close tags and other common mistakes). This option is enabled by default. If the offline copy of your website does not render correctly in your browser, try turning this option off and downloading the site again. When this is turned off Web2Disk will save each page's HTML exactly as it is received.

Download Near Offsite Content

If this option is checked Web2Disk automatically download supporting content (such as images, CSS and JavaScript) that is hosted on servers outside the Root URL's domain. This feature is enabled by default and provides maximum compatibility with sites that use content delivery networks (CDN) for distribution. If you disable this setting, only files located within the Root URL will be downloaded.

Create URL to File Map

If this option is checked Web2Disk will generate a CSV file in the Save Folder named "URL

to File Mapping.csv". If opened in Excel this file will contain two columns; the original URL and the file on disk that the website was saved to. This is useful if you wish to find which file maps to a particular URL on the live website.

Add Download Meta Data

If this option is checked Web2Disk will insert an HTML comment with the download date and the original URL into the HTML header of each downloaded page.

Only Download These File Types

This option is used to restrict what Web2Disk actually downloads and saves. (The Exclusion List is used to filter how Web2Disk crawls a website.) Use this feature if you only want to download a certain type of file from a website. For example, to download only PDF documents, you would enter "*.pdf" on a line in the File Filters.

Multiple filters can be placed one per line, or on a single line separated by commas. The * character is used to match one or more letters. For example, "*.jpg" would match any files that end with ".jpg".

Crawler Options Tab

The Crawler Options allows you to fine tune how Web2Disk downloads a website.

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 Web2Disk places too high a load on your server, increase this value to slow the crawl down.

User Agent - The User Agent value indicates how Web2Disk should identify itself to your webserver. By default, Web2Disk identifies as the Inspyder Crawler. If your site uses browser detection scripts, you may wish to have Web2Disk 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 your want Web2Disk 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 Web2Disk, try setting a limit of 1024 KB. For most websites setting this value is not required (use 0 for no limit).

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 Web2Disk 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, Web2Disk is set to use 5 threads. Increasing this value can make Web2Disk 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 Web2Disk will attempt to locate URLs within JavaScript code.

Use JavaScript Heuristic - If this option is checked Web2Disk 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 Web2Disk to automatically discover all your pages.

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 Web2Disk 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 Keep-Alive - This option tells Web2Disk 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.

Use HTTP 1.0 - This option tells Web2Disk to use version 1.0 of the HTTP protocol (instead of 1.1). Some websites may be required to enable this option if they experience difficulty with crawling.

Obey 'nofollow' Meta Tag - If this option is checked, links on pages that include the Robots Meta tag with "nofollow" in the content attribute will be ignored.

Ignore Protocol Prefix - If checked Web2Disk will ignore the "http://" or "https://" protocol prefix when determine if a URL has already been crawled. With some websites that mix secure and non-secure sections together is may be necessary to uncheck this option to get a complete crawl.

Parse Form Action Attribute - If this option is checked, Web2Disk will parse the HTML "action" attribute from "form" tags. Turning this on can cause Web2Disk to discover more content on your website, but may also create crawler errors. Leaving this option disabled is recommended.

Ignore HTTP Content Type Encoding - If this option is checked, Web2Disk will ignore the HTTP "Content-Type" header returned by your web server. This option is only necessary if Web2Disk returns unreadable characters in it's results. (This option is rarely needed and should almost always be left unchecked.)

Parse Form "Action" Attribute - By default Web2Disk does not include URLs found in the "Action" attribute of HTML forms. If this option is checked Web2Disk will include these URLs when crawling.

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

Case Sensitive URLs - This option tells the crawler to treat URLs that differ in case only as sperate URLs. For example, when this option is turned on, the crawler will access both these URLs separately:

http://www.example.com/default.aspx
http://www.example.com/Default.aspx

For most websites this option should be disabled.

Root URLs Tab

Use this page to specify Additional Root URLs. Additional Root URLs are used the 'seed' the crawler before checking your site. You only need to specify Additional Root URLs if your site is not fully cross-linked. For example, if you had some pages that were only linked to by an external website, then you may want to add one of those pages as an Additional Root URL so that the crawler is able to find them and include them in the download.

Domain Aliases Tab

Domain aliases tell Web2Disk that your site may use one of these as alternate domain names. If you consistently use the same domain name, or only use relative paths on your website, then you do not need to enter anything here. However, if you use multiple domain names ("www. example.com" and "example.com") to refer to the same site on the same server, than it is necessary to enter those alternate names here. You do not need to include the domain name given in the Root URL.

4.2   Clear Page Change History

The Clear Page Change History menu item (from the "**Project**" menu) allows you to delete the history file that Web2Disk uses to determine which files have been added, removed or changed since the last crawl. After cleaning the change log and crawling your website again, all files will appear as "New" in the "Changed Files" result tab.

Clearing the Page Change History only affects the current project.

4.3   Soft 404 Detection

Soft 404 Detection allows Web2Disk 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 Web2Disk 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 Web2Disk 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 Web2Disk 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.

Soft 404 Text:

Page Not Found
Oops...sorry

New Text: Example: File Not Found   Add

4.4   Excluded Pages

Often it is necessary to exclude some sections of a website from crawling. To do this, Web2Disk 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.
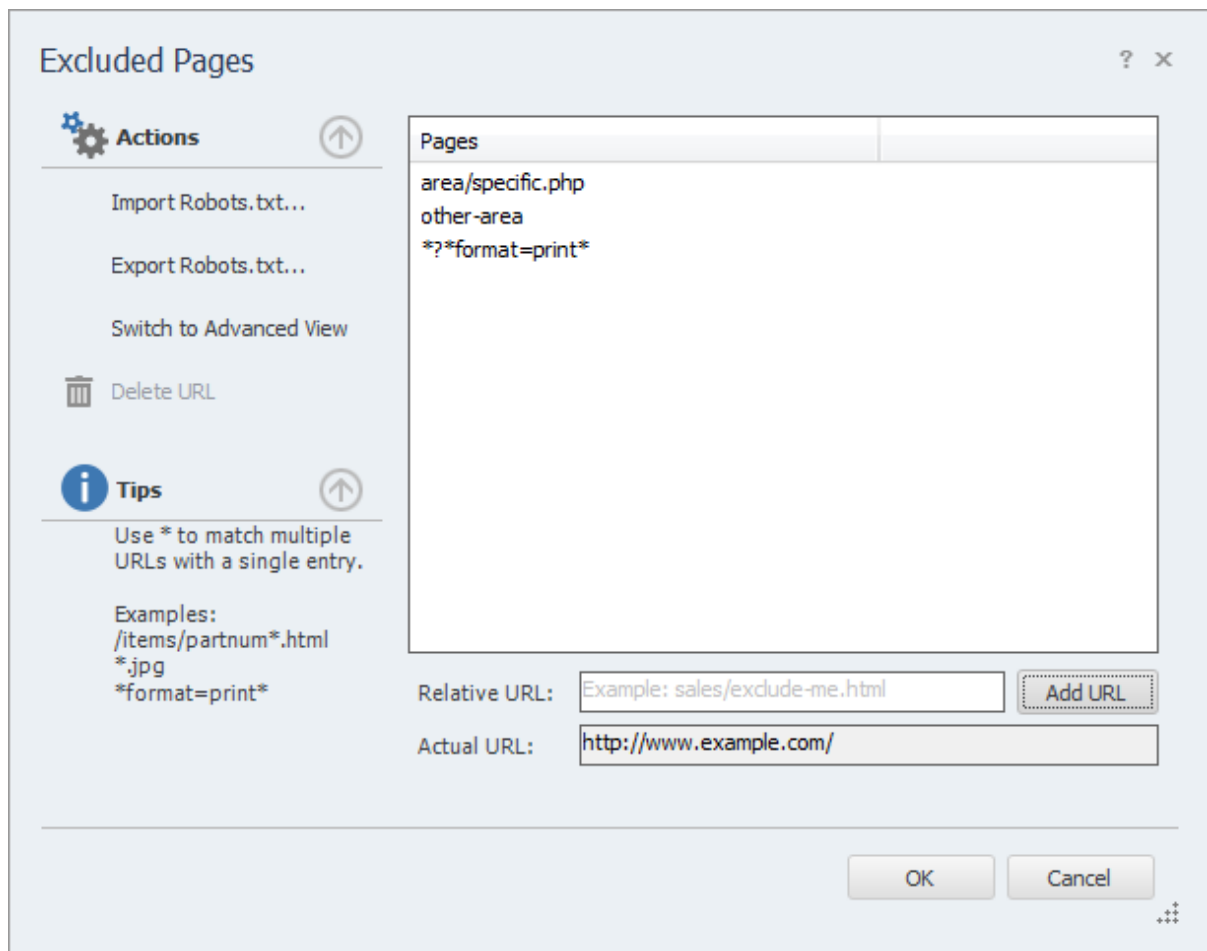
By clicking the "**Switch to Advanced View**" link on the left hand "**Action**" menu you are able to mark a page or section as specifically **included**. The list is matched from top to bottom. If you want to exclude all pages under "www.example.com/content/" except "www.example.com/content/important" than 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.

4.4.1 Importing Robots.txt

Some websites already contain a special file called "robots.txt" which tells crawling software which URLs (or paths) to ignore. Web2Disk does not automatically obey robots.txt because it is often desirable to crawl the entire website. However, Web2Disk 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.
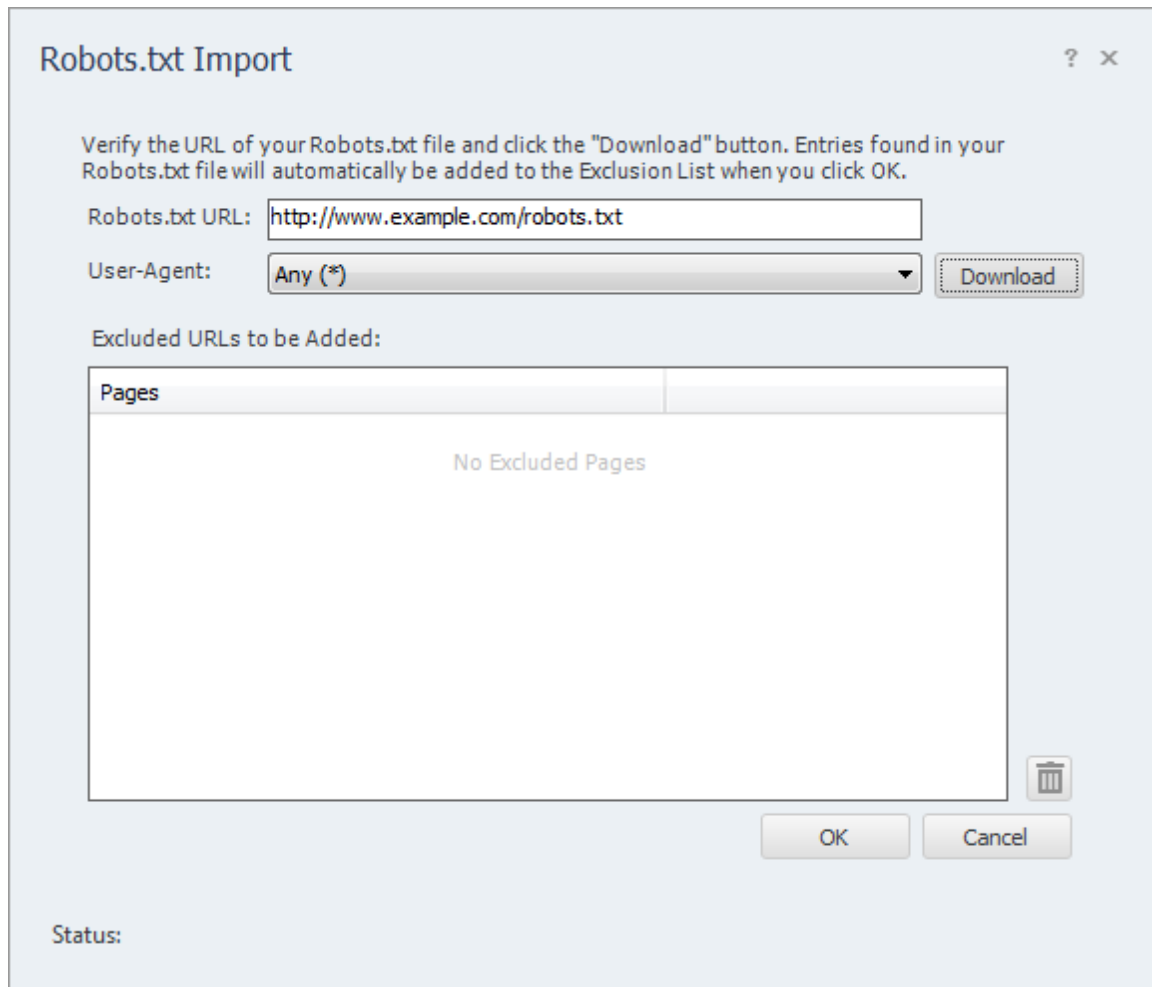
When the Robots.txt Import window opens, it will automatically populate the "Robots.txt URL" field based on your Root URL. You can modify this field if necessary. Click "Download" to fetch and load the file from the website.

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 Web2Disk 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.

4.4.2 Exporting Robots.txt

In addition to being able to import an existing robots.txt file, Web2Disk 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 Web2Disk must be placed at your domain root. For example, if you Root URL was "www.example.com/products/", the robots.txt file created by Web2Disk must be placed so that it's accessible from "www.example.com/robots.txt" (not in the "products" sub-folder).

The robots.txt file created by Web2Disk 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).

4.5   Passwords and Forms

If your website is password protected or contains content hidden behind a form submission, you can configure Web2Disk 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 Web2Disk 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.

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

The "Method" field tells Web2Disk 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, Web2Disk will ask if you want to save the form data, click 'Yes' and Web2Disk will automatically create the necessary login rule for you.

Prompt for Manual Login

If you don't want Web2Disk 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 Web2Disk will launch the mini-browser to capture your login session on-the-fly. When you close the mini-browser Web2Disk 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 Copying a Website to CD

Copying a website to CD-ROM that has been downloaded with Web2Disk is easy. Web2Disk automatically fixes the website's link structure, so you can easily redistribute your website on CD-ROM and all the links will work. Only a web browser is required to view websites downloaded with Web2Disk, so there is no need to redistribute or install Web2Disk on the target PC.

When you download a website, that website's files are placed in the "Save Folder" specified on the main Web2Disk window. By default that folder is:

Windows Vista/7: C:\Users\[username]\Documents\Websites\[Project Name]\
Windows XP/2003: C:\Documents and Settings\[username]\My Documents\[Project Name]\

Where [username] is your Windows username and [Project Name] is the current Web2Disk project name. In that folder you should find three items:
1. A file named "autorun.inf"
2. A directory with the domain name of the Root URL website
3. A shortcut named "Launch www.yourwebsite.com"

Note: You may see other folders in this directory. These folders will contain files that were required by the website but hosted on external servers (see the "Download Near Offsite Content" option in Advanced Project Settings). It is recommended that you included these folders too when burning your website to CD.
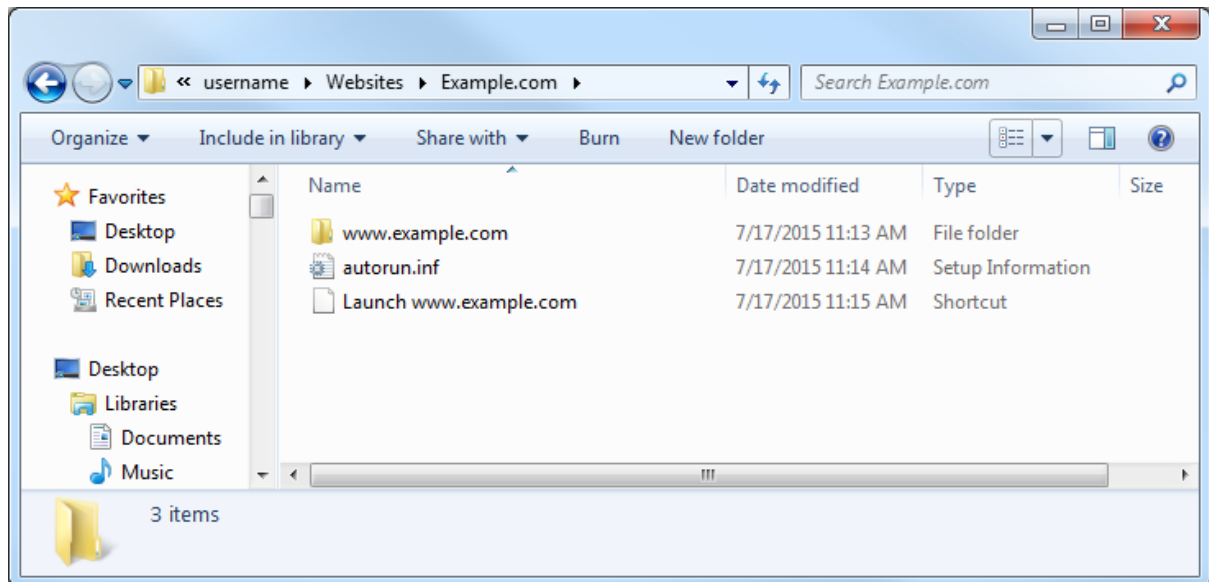
Burn the first two items exactly as they are using your favorite burning software. You will want to burn these to the root of a new CD-ROM. For example, if we were downloading "www.example.com" and your CD-ROM drive is named "D:", then when you browse your burned CD you should find:
D:\autorun.inf
D:\www.example.com (folder)
D:\Launch www.example.com

The "autorun.inf" tells Windows to automatically run your website when the CD is inserted. If you do not want your website to automatically run, just omit this file from the burning process.

After downloading your website, if you do not see the "autorun.inf" file, ensure that "Create Autorun File" is enabled in your Advanced Project Settings. To launch the website from the CD manually, double click on the "Launch" shortcut. Your default browser should open and you should be taken to the homepage.

6   Defaut 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 Web2Disk 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.

7 Project Import and Export

Web2Disk 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.
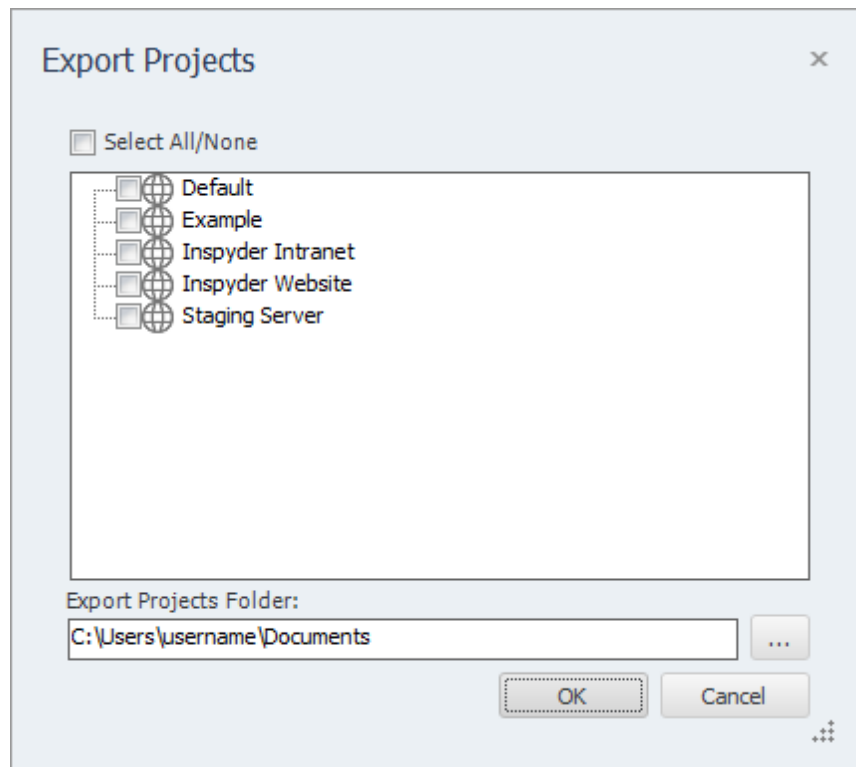
7.1 Exporting Project Files

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 Web2Disk. "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 Web2Disk 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 Web2Disk.

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'.

7.2   Importing Project Files

To Import one or more project files into Web2Disk, 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 Web2Disk. 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 Web2Disk project back into Web2Disk, then all settings will be used.

8   Email Settings

8.1   Server Settings

Configuring your email server settings (SMTP) allows Web2Disk to send automated emails. To access your mail server settings, click "**Settings | Email Settings...**" from the main menu bar.

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 us 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 Web2Disk will not attempt to use SMTP authentication.

Sender Address - This is the email address that Web2Disk 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.

8.2 Message Body

The Web2Disk email subject line and message body can be customized in the Subject and Body fields respectively. By using special tags inside the message body, it can be dynamically customized based on the current project settings and the results of the crawl.

The following tags can be used:

Tag	Location	Description
#project#	Subject and Body	The project name
#rooturl#	Subject and Body	The Root URL of the project
#offlineurl#	Body	The location where the site was saved on your PC
#changes#	Body	A listing of the files that changed since the last crawl

Keep in mind that the report with the full crawl results listing can be included with the email as an attachment.

9   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...".
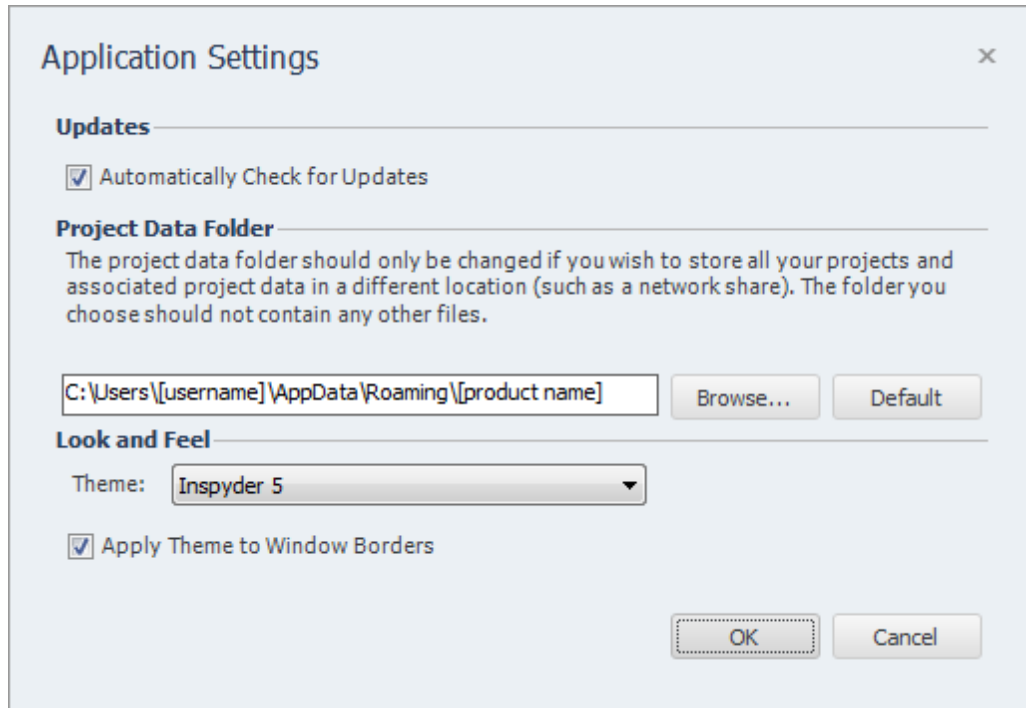

Updates - If this option is enabled, Web2Disk will automatically check for a new version each time it runs. Unselect this option if you do not wish to see the "A new version is available" message when you start Web2Disk.

Project Data Folder - This setting allows you to change where Web2Disk stores it's 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 Web2Disk 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 Web2Disk.

Look and Feel - These options allow you change the theme of Web2Disk. "Apply Theme to Window Borders" allows you to over-ride the Windows border with a style that matches the current theme.

10   Scheduler

Web2Disk has the ability to run automatically in an unattended mode. To configure Web2Disk for scheduled operation simply select "**Settings | Scheduler**" from the main menu bar.
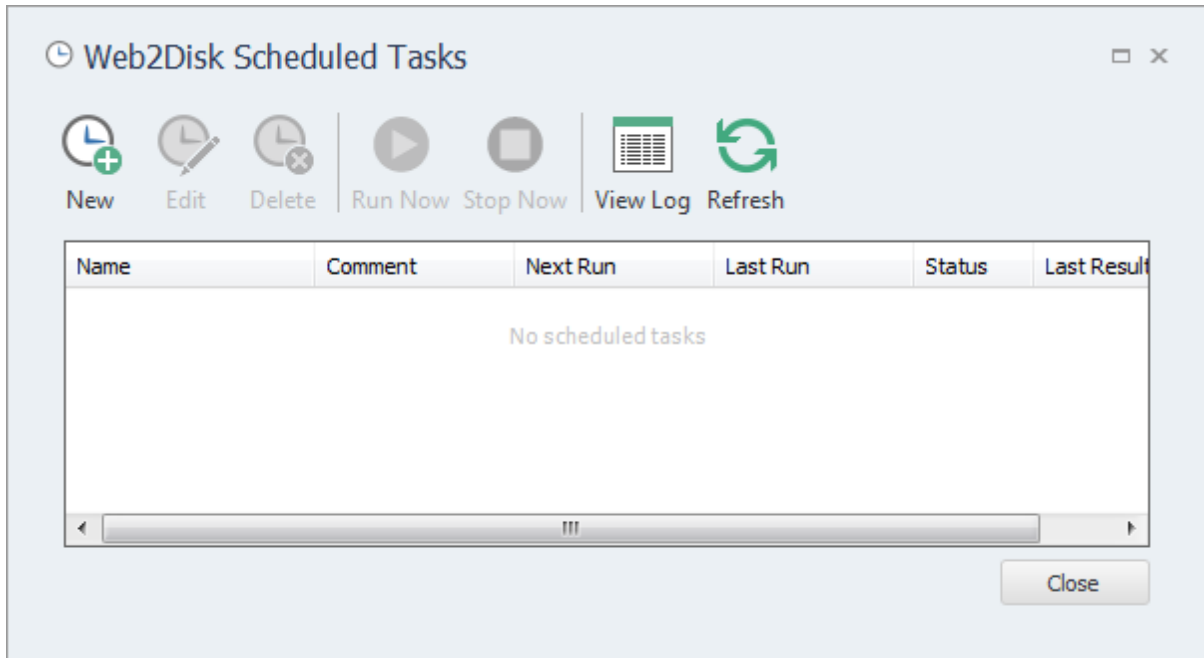
10.1   Scheduling Basics

Scheduling Web2Disk to run automatically is done by creating one or more tasks in the Web2Disk Task Scheduler. Each scheduled task contains the settings that will be used when running Web2Disk automatically. Basic settings include:

- The Web2Disk 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 Web2Disk 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 Web2Disk are visible in the Windows Task scheduler. For scheduling options more advanced than those provided by Web2Disk, 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 if not zero than an error may have occurred (the Scheduler Log may have more details on any errors that occurred).

10.2 Adding/Editing a Task

To add a new task to the Web2Disk 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 Web2Disk 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.

10.3 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.

10.4 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.
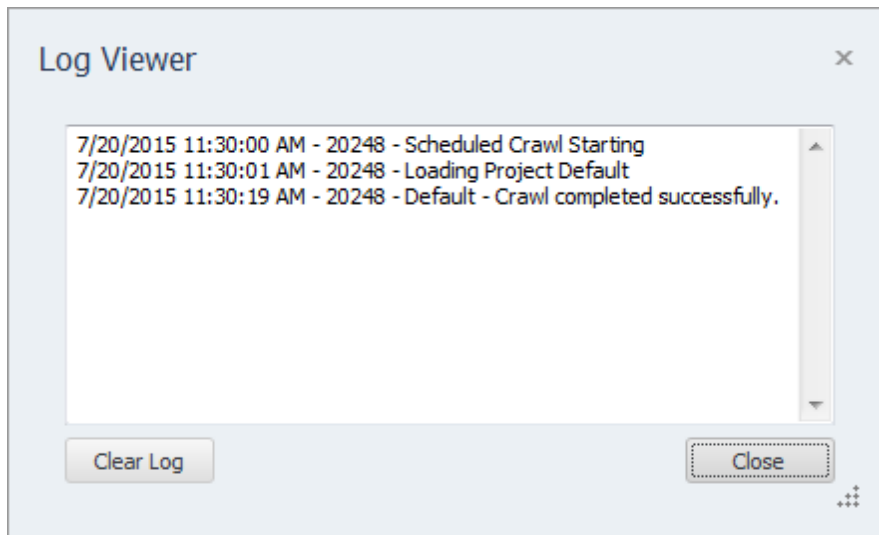
Each line in the log file provides the following information:
- Date/Time the event occurred
- Process ID of the running Task (this is useful for tracing events if you have multiple scheduled tasks running at the same time)
- The log message

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

10.5   Command Line Interface

If you wish to run Web2Disk 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:

-project=	The project name to be downloaded (i.e. -project=default)
-rooturl=	(Optional) Can be used to override the project's Root URL
-savepath=	(Optional) Can be used to override the project's Save Path
-filelimit=	(Optional) Overrides the project's File Limit setting
-depthlimit=	(Optional) Overrides the project's Depth Limit setting
-email=	(Optional) The email address to use for notification
-checkchanges	(Optional, requires -email) Indicates that an email notification should only be sent if changes are detected
-rename	(Optional) Overrides the project's "Rename Old Downloads" setting and forces Web2Disk to rename any previous downloads

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

10.6   Result Codes

These are the codes returned to the Task Scheduler after the Web2Disk 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: Try running the report in interactive mode, and check for errors. The report file may have
been in use or an incorrect output path may have been selected.
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   Emailing Error
     Solution: Check to make sure your SMTP settings are correct in Email Settings and try again.

11   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.

12   License Agreement

This legal document is an agreement between you, as licensee and Inspyder Software Inc. ("Inspyder"), as licenser. 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 Licenser 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 Licenser'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 Web2Disk to provide an exclusive service such as using Web2Disk 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 Licenser's entire liability and Your exclusive remedy shall be: (a) the replacement of any media not meeting Licenser's 'Limited Warranty' and which is returned to Licenser 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 Licenser 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 Licenser or any authorized representative of Licenser 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 © 2015 by Inspyder Software Inc.

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