



**Inspyder InSite  
User's Reference Manual**

Copyright © 2008 Inspyder Software Inc.

# **Inspyder InSite (InSite) User's Reference Manual**

**Copyright © 2008 Inspyder Software Inc.**

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

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

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

Printed October 2008 in Canada.

# Table of Contents

## Part 1 Introduction

## Part 2 Quick Start Guide

## Part 3 Usage

|       |                                 |    |
|-------|---------------------------------|----|
| 3.1   | Button Bar & Menu .....         | 3  |
| 3.2   | Project .....                   | 4  |
| 3.2.1 | Advanced Project Settings ..... | 7  |
| 3.2.2 | Import Robots.txt .....         | 10 |
| 3.2.3 | Exporting Results .....         | 11 |
| 3.2.4 | Word Count .....                | 12 |
| 3.3   | Custom Dictionary .....         | 13 |
| 3.3.1 | Custom Word List .....          | 13 |
| 3.3.2 | InSite_Words Meta Tag .....     | 14 |
| 3.4   | Scan Results .....              | 15 |
| 3.5   | Settings .....                  | 16 |
| 3.5.1 | Email Settings .....            | 16 |
| 3.5.2 | Scheduler .....                 | 17 |
| 3.5.3 | Master Project Settings .....   | 20 |

## Part 4 About Inspyder Software

## Part 5 Product License Agreement

## Index

## 1 Introduction

InSite from [Inspyder Software Inc.](#) is a Windows based utility to analyse web sites and provide detailed results on spelling and link errors. InSite supports multiple languages and a custom dictionary for words and phrases specific to your organization/industry. Its link analysis will follow references in your website's HTML and look for missing pages and files including images, stylesheets, etc.

InSite is a web crawler so it checks the pages your customers see, regardless of your server configuration. Therefore, InSite is compatible with Apache, IIS and other web server software. It can crawl pages created by PHP, ASP, JSP, or other technologies. There's no need to enter in hundreds of links. Just give it your homepage URL and let InSite do the rest.

With InSite you can crawl password protected pages and websites with ease. InSite supports both HTTP authentication and posting form data, making it easy for you to check all types of protected pages that your customers might see. InSite can also be configured to exclude certain pages from its crawl, such as message boards and other pages that might not require checking.

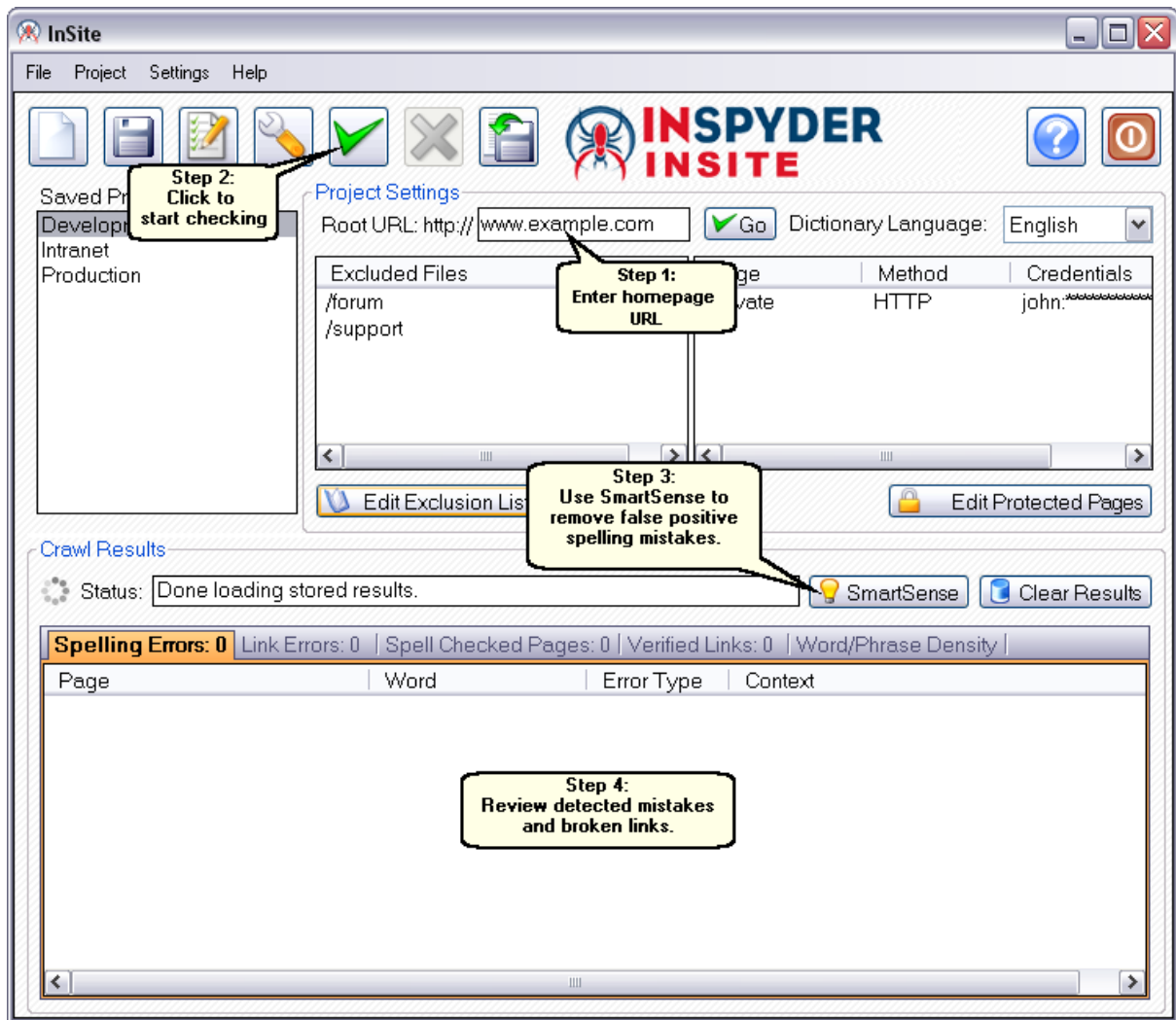
### Features and Benefits

- Easy to use - Just enter your homepage URL and go.
- Search your entire site for spelling mistakes - InSite crawls all your website's pages for spelling mistakes, no more hand-checking each page.
- Quickly find broken links - InSite automatically finds broken links and missing images, saving you time and effort.
- Custom dictionary - The custom dictionary allows you to easily add words that are specific to your business or site.
- Scan password protected pages - InSite provides robust facilities for crawling password protected pages.
- Crawl only within your site - InSite crawls only pages below the root page, so you don't have to worry about crawling pages that aren't yours.

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

## 2 Quick Start Guide

- Step 1** The default project is now loaded but does not contain any settings. Enter the URL of your homepage in the 'Root URL' field.
- Step 2** Click the "**Go**" button.
- Step 3** Use *SmartSense* to remove false spelling mistakes and to automatically build the custom word list.
- Step 4** Verify your results in the Scan Results section and make corrections to your website.



**Hint:** Try modifying the [Advanced Project Settings](#) to refine how InSite analyses your site. The Advanced Project Settings are accessible from the main menu bar. Click "Project | Advanced Project Settings..." at the top of the screen.

### 3 Usage

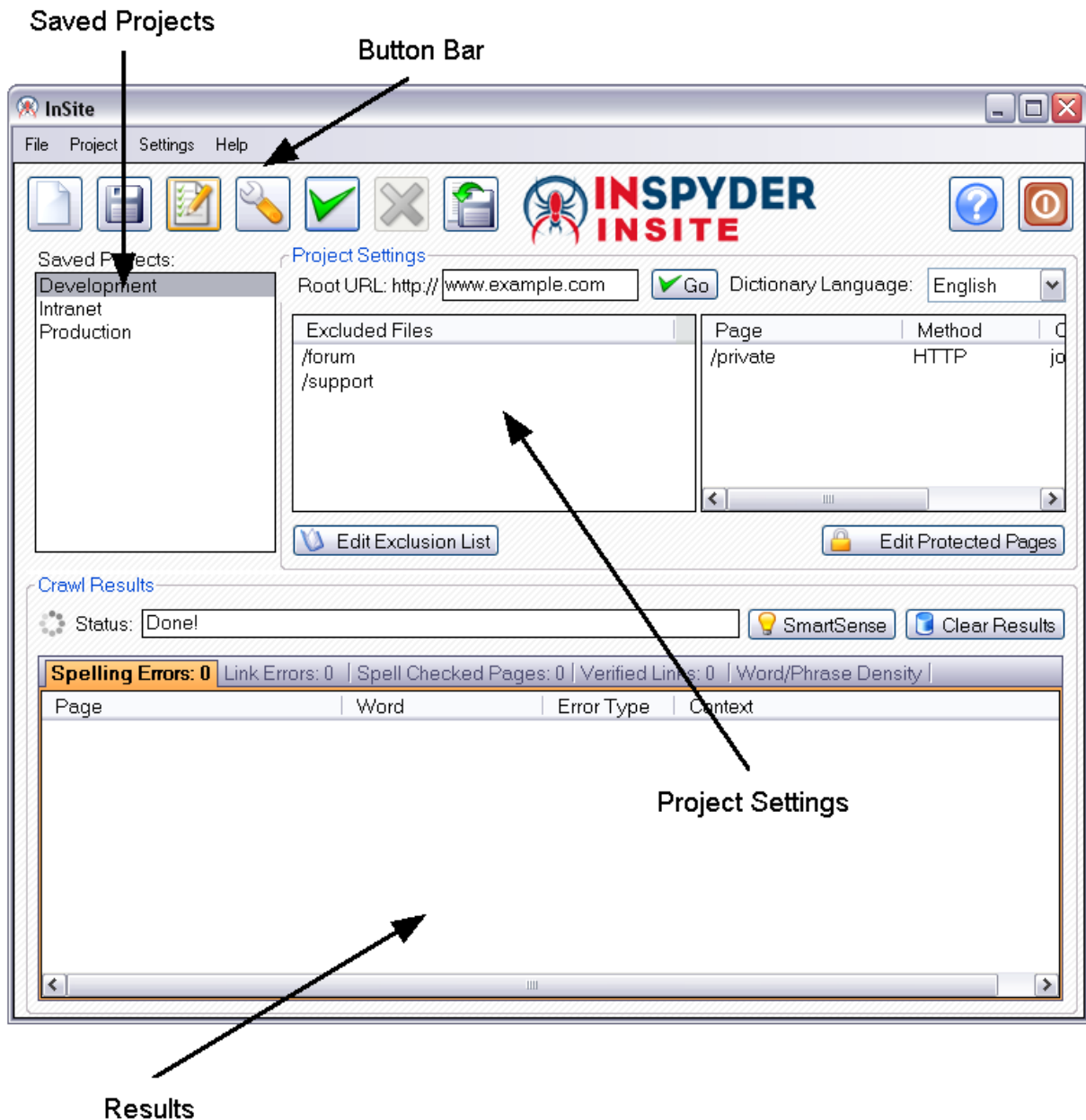
The InSite user interface area is broken down into four main areas:

**Button Bar** These buttons provide quick access to frequently used functions within InSite.

**Saved Projects** This is a list of the currently saved projects. When you create a new project, it will appear in this list. Projects can be deleted or renamed by right clicking the project name in the project list. To switch projects, click on a different project name in the list.

**Project Settings** This is where the settings for the currently select project will appear.

**Scan Results** This is where the output from InSite will be presented.



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

### 3.1 Button Bar & Menu

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



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



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



Use this button to edit the Custom Word List. The Custom Word List is a private dictionary used by the spell checker. For more information please refer to the [Custom Word List](#) section of this document.



This button allows you to access the advanced configuration options for the current project. For more information please refer to the [Advanced Project Settings](#) section of this document.



Select a project from your list of previously defined projects and then click on this button to start the analysis process. The analysis can be interrupted at any time by clicking on the **Stop** button.



Use this button to stop the web site checking process.



This button provides a way to export the crawl information. For more information please refer to the [Exporting Results](#) section of this document.



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




Closes the program.

The main menu contains most of the functionality of the button bar, but extends that capability with a few additional commands. The most notable menu additions are the "**File | Save As**", "[Settings | Scheduler](#)" and "[Project | Word Count](#)" commands.

## 3.2 Project

This section describes the options, fields and features of the Project Settings group.

**Root URL** - Specify the address of the homepage for the website to be analysed. You can enter an IP address or the URL of the web page. The URL does not need to begin with the "http://" prefix. If your website requires the use of SSL, than you must specify the "https://" prefix.

**Go Button** - Use this control to begin the analysis of the Root URL. This button has the same functionality as the  control on the button bar. It is included in this location for convenience.

**Dictionary Language** - This is the language of the common dictionary. At this time only the English is provided standard with InSite. We are currently working to make additional dictionaries available. When they are completed they will be available, free for registered users, from our website.

**Excluded Pages** - The Excluded Paths list contains URLs that should not be part the of the analysis. 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 your results.

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

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

*inventory/warehouse\*index* will exclude the following:  
*inventory/warehouse1index*  
*inventory/warehouse125index*  
*inventory/warehouseindex*

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

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", you could enter: *\*format=print\**. This will exclude all URLs that contain 'format=print' somewhere in the URL.

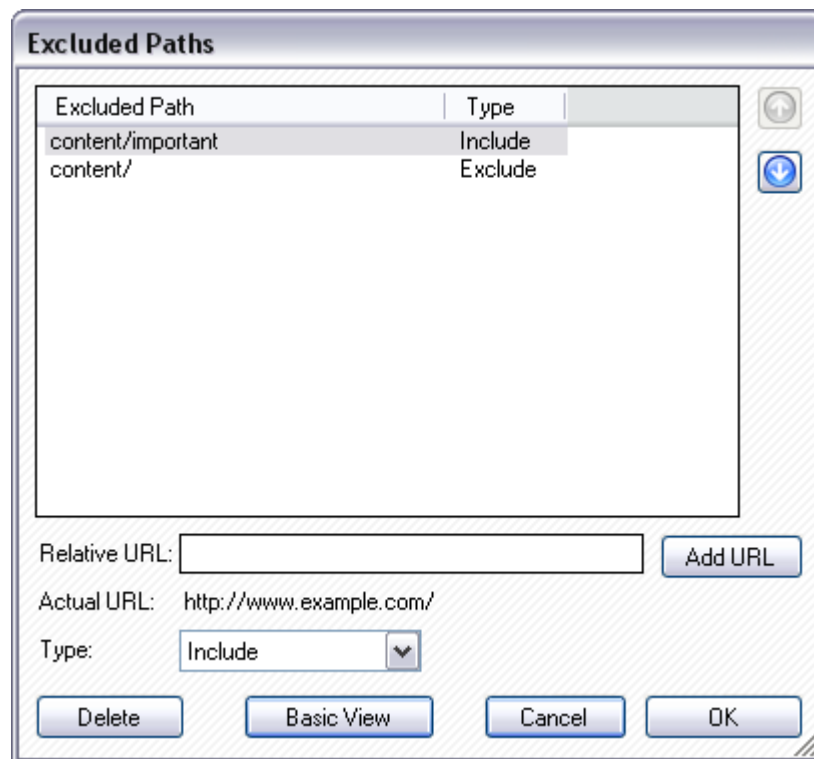
Click the **Edit Exclusion List** button to add URLs to the exclusion list. Internal URLs should be relative to the root URL. URLs beginning with '/' are taken from the domain root (this is not recommended). URLs **not** beginning with '/' are taken from the start of the crawl. For example, if the root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

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

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

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

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



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

A page can also be added to the exclusion list by right clicking the URL in the [Spell Check](#) tab.

**Protected Pages** - Password protected pages can be configured by clicking the **Edit Protected Pages** button. A dialog box will appear that will let you enter the relative URL of the password protected page, as well as the authentication method and the credentials. If the HTTP method is selected then a username, password and domain can be entered into the appropriate fields. Note that the domain field is optional.

If the POST method is selected then the credentials must be entered as form data (For example, the following form data can be posted to a page: *username=user&password=pass*. Where 'username' and 'password' are the form items, and 'user' and 'pass' are the values).

To simplify this process InSite provides a helper tool. Clicking the 'Wand' icon will prompt you for a variable name and value. These values will be entered automatically in the 'Form Data' field in the correct format. To add multiple variables, simply click the 'Wand' again. The POST method can also be used to fill out entire forms on your site if required.

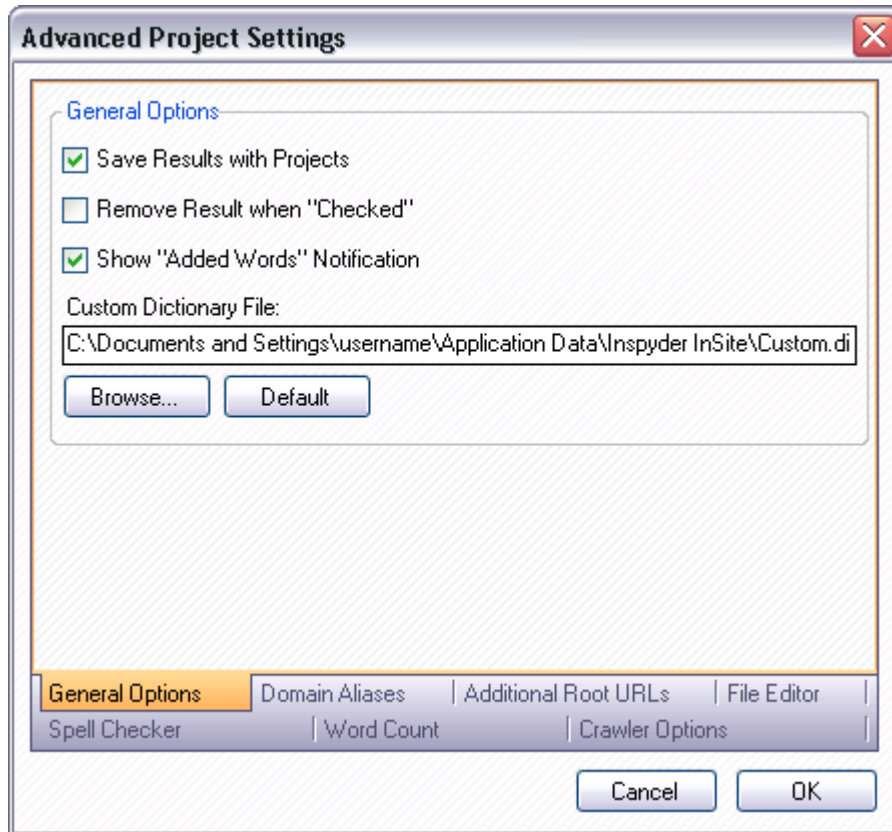
Refer to the [Excluded Pages](#) description above for examples of proper URL formation.

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

C:\Documents and Settings\username\Application Data\Inspyder InSite\

### 3.2.1 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 Options

**Save Results with Projects** - If this box is checked, InSite will save the results of each crawl with that project. If this options is disabled, InSite will not save any project settings. This latter behaviour is the same as versions of InSite prior to version 1.4.0.

**Remove Results when "Checked"** - If this box is checked, InSite will remove any spelling mistakes and link errors that have their checkbox checked from the list of results. If this options is unchecked (default), then result simply has a line drawn through it.

**Show "Added Words" Notification** - If this box is checked, InSite will popup a window confirming when words are added to the custom dictionary from the result list.

**Custom Dictionary File** - New in version 1.5.0 is the ability to have multiple custom dictionaries. By default, all projects use the same custom dictionary file. This can be changed by using the "Browse..." button. If the dictionary file does not already exist, it will be automatically created when words are added to it.

#### Crawler Options

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

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

**Max 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 'mouse-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.

**Enable Link Checking** - This option enables/disables the link checking component of InSite. If this box is not checked, then broken links will not be found or reported on your site.

**Ignore Invalid Relative Path Errors** - This option allows you to tell InSite to ignore and internally correct invalid relative path errors. By default this option is disabled, since these types of errors should be flagged for correction. However, in some cases it is desirable to remove these errors from the result list. With this option enabled InSite will correct invalid relative paths internally the same way a web browser does. For example, the relative link: `"../..../index.php"` on the page:

`http://www.example.com/support/kb/`  
will become:  
`http://www.example.com/index.php`

## Domain Aliases

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

## 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.

## File Editor

**Editor** - This is the HTML editor that InSite should launch when you select 'Open Page in Editor' from the results shortcut menu ('Right-Click' on a spelling mistake to see the shortcut). Your editor must support a filename as a command-line argument in order for it to open the file from InSite.

**Path** - This is the location on the local hard drive (or network share) where the website's source files reside. It is used to send the correct filename the HTML editor when opening a page from the results.

**Default Filename** - This filename will be appended to URLs where there is no filename when attempting to open the file in the Editor. For example, if a spelling error was reported on '[www.example.com/support/](http://www.example.com/support/)' and the default filename was 'index.html'. Then InSite would attempt to open a file named 'index.html'.

**Parameter Separator** - This indicates the character that your server uses to distinguish URL parameters from filenames. (The default is '?'.) InSite will strip any characters following the parameter separator from URLs before opening the file.

## Spell Checker

**Enable Spell Checking** - This option disables InSite's spell checking features. Uncheck this box if you do not wish to find any spelling mistakes on your site.

**Flag Double Words** - This option causes repeated words (i.e. the the) to be flagged as an error.

**Flag Mixed Case** - This option cause words with mixed case to be flagged as a potential error. For example, the word 'compuTer' would be flagged as an error. This setting can have a large impact on the number of errors reported, although they will be flagged as "Mixed Case" errors due to the proliferation of custom spellings used today.

**Ignore File and Domain Names** - This feature causes file names and domain names to be excluded from any spell checking.

**Ignore Words with Numbers** - If this option is checked, words that contain numbers will not be flagged as spelling mistakes (for example: 23lbs).

**Ignore Email Addresses** - If this option is checked email addresses to be excluded from any spell checking.

**Ignore Proper Names** - This feature causes proper names, words beginning with a capital letter but not at the start of a sentence, to be excluded from any spell checking. This is useful if your website contains a large number of names (not already in the dictionary) or names of locations (for example, if you have a travel related website).

**Ignore Acronyms** - This feature causes acronyms, words entirely in capital letters, to be excluded from any spell checking.

**Ignore Meta Text** - If this option is checked, any Meta Keywords or Meta Description tags found on your site will not be spell checked. This is useful if your meta tags contain deliberate spelling mistakes and you do not wish these to be found by the spell checker.

## Advanced Spell Check Options

**Excluded CSS Classes** - This feature allows you to exclude text enclosed within a specific CSS class from spell checking. This is useful on sites where some text should never be checked for spelling errors and is encapsulated within a unique CSS class name. An example would be a website with technical documentation. Some portions of the technical documentation (such as examples of HTML code, which would generate a large number of false spelling mistakes) are wrapped within a CSS class.

## SmartSense Options

**Sensitivity** - This slider allows you to set the sensitivity of the SmartSense analysis technology. In general, increased sensitivity will decrease the number of words that InSite will detect as unique to your web site. Whereas setting the sensitivity level too low may yield more

words, but there is an increased chance that these words may include misspelled words that have been detected as unique to your site. The default has been set at a level that has empirically returned the best results.

### Word Count Options

**Enable Word Counting** - Uncheck this box to prevent InSite from counting the words on your site. This can improve performance slightly if you have a very large site and do not require any word count information.

**Include Numbers** - If this option is selected the word count will include numeric values such as "1234". By default this setting is disabled so that numbers are excluded from the word counts.

**Include Words with Numbers** - If this option is selected then words that contain numbers, such as "3M" are included in the word counts.

**Include Words in Meta Text** - If this option is checked, words that are included on your pages, but not visible by default (such as alternate image text, description tags, etc.), will be included in the word count.

**Include Numbers** - If this option is checked, numbers that appear in the text of the website will be counted as words. By default this is unchecked.

**Word Length Threshold** - This setting is the minimum length of a word, in characters, that is included in the word counts. Words this length or less are excluded.

### Keyword Density Options

Keyword Density Options panel is used to configure options pertaining to InSite's keyword density analysis feature. Keyword density is the ratio of occurrences of a particular word or phrase to the total number of words or phrases in a particular page. The densities for phrases are calculated using the number of phrases that occur on the site (not the number of words).

**Enable Keyword Density Analysis** - Enables/disables keyword density analysis. If you have a very large website, and do not require keyword density information, disabling this feature can provide a small performance improvement.

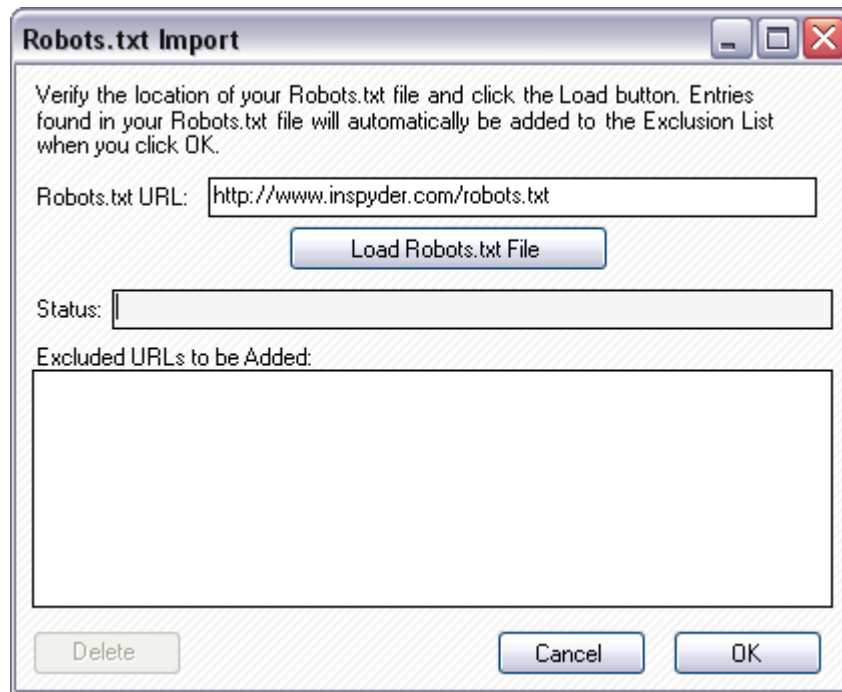
**Maximum Phrase Length** - This setting is the maximum length of a phrase, in words, that is included in word/phrase density analysis. If this value were set to "3", then a phrase like "red brown fox" would be reported as a phrase, where "red brown fox jumps" would not be included.

**Minimum Density** - This setting is the lowest phrase density that will be included in the results.

**Word Length Threshold** - This setting is the minimum length of a word, in characters, that is included in the word/phrase density analysis. Words "equal to" or "less than" this value are excluded. For example, if the Word Length Threshold was 3, than "Quick Brown Fox" would not be considered a phrase since the word "Fox" is 3 characters (it excludes the phrase).

## 3.2.2 Import Robots.txt

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



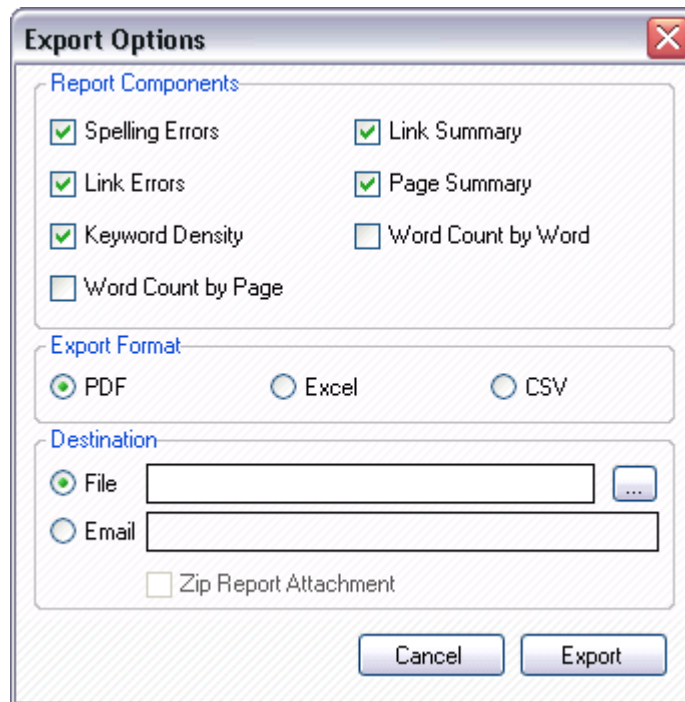
The "Robots.txt URL" field should contain the location of your website's robots.txt file. InSite will try and guess this URL based on the Root URL for the current Project. If the URL that InSite guesses is not correct, you can correct it in place. To load the file, click "Load Robots.txt File". InSite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

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

### 3.2.3 Exporting Results

Web scan results can be exported as a PDF, Excel or a comma delimited format (CSV) file. Any of these formats can be written to the selected file or they can be emailed to a specified email address. Emailing the results is especially useful if the web analysis is being run on a [scheduled](#) basis.

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



### Report Components

This allows you to select which components you would like included in the report. By default, the two Word Count components are not included because they greatly increase the size of the report.

### CSV Format

If the CSV Output format is selected, you can load the exported file into a word processor, database, or spreadsheet for additional manipulation. Since each tabbed divider contains data in a slightly different format titles separating the data types are inserted into the export file.

Once the data has been imported into your favorite utility, it is a simple matter to cut and paste data from the different tabbed dividers for further sorting, analysis, and reporting.

### PDF/Excel Format

The PDF/Excel formats produce a website report that is suitable for printing or distribution. Both version provide formatting not found in the CSV format to improve readability and appearance. To customize the logo image in these reports, replace the file located at:  
 C:\Program Files\Inspyder Software Inc\Inspyder InSite\Reports\ReportLogo.jpg  
 with your own logo image. For printed reports, a 300 DPI resolution (or higher) image is recommended.

### Destination

This allows you to select where the report should be saved/sent. If File is selected, the report is saved to the location specified (click the '...' button to open the file browser). The Email is selected, the report is automatically emailed as an attachment (see [Email Settings](#) for more information). If Zip Report Attachment is checked, the report file is automatically compressed into a zip file before being sent.

## 3.2.4 Word Count

After your site has been crawled, it is possible to view the number of words found on each page. To view the Word Count dialog, click "**Project | Word Count**". Remember, you must let InSite crawl your site before accessing the word count dialog. Failure to do so will result in no words being shown or an

incomplete/inaccurate count.

Any pages excluded from checking (through the Excluded Pages feature) will not be included in the word count.

| Page  | Count |
|---|-------|
| http://www.inspyder.com/check.out.php?ProductID=30003       | 164   |
| http://www.inspyder.com/Company.php                         | 88    |
| http://www.inspyder.com/Contact.php                         | 41    |
| http://www.inspyder.com/downloads/                          | 97    |
| http://www.inspyder.com/index2.php                          | 126   |
| http://www.inspyder.com/mailorder.php                       | 112   |
| http://www.inspyder.com/PrivacyPolicy.php                   | 427   |
| http://www.inspyder.com/products/                           | 140   |
| http://www.inspyder.com/products/Bundles.php                | 129   |
| http://www.inspyder.com/products/Finder/                    | 385   |
| http://www.inspyder.com/products/Finder/revisionhistory.php | 97    |
| http://www.inspyder.com/products/Finder/CategoryList.php    | 90    |

Word Count by Page | Word Count by Word

Total Word Count: 10809      Total Unique Word Count: 1345      Export...      OK

**Word Count by Page** - This view shows the total number of words found on each individual page.

**Word Count by Word** - This view shows each word that was found on the website and the number of times that word appeared.

**Total Word Count** - The total word count is the total number of words found on all pages.

**Total Unique Word Count** - This is the total number of distinct words that appeared on all pages.

Word count information can be exported in the same way as spelling and link checking results. Simply click the 'Export Results' button and chose 'Word Count' as a Report Component. Both the 'Word Count by Page' and 'Word Count by Word' summaries will be exported.

## 3.3 Custom Dictionary

Often it is desirable to create custom lists containing terms that are specific to your business or website that may not be found in common grammar. InSite supports this capability with both a [Custom Word List](#) (Custom Dictionary) which is applied to all of the projects, and through the use of [meta tags](#) which are embedded in and applied only to specific websites.

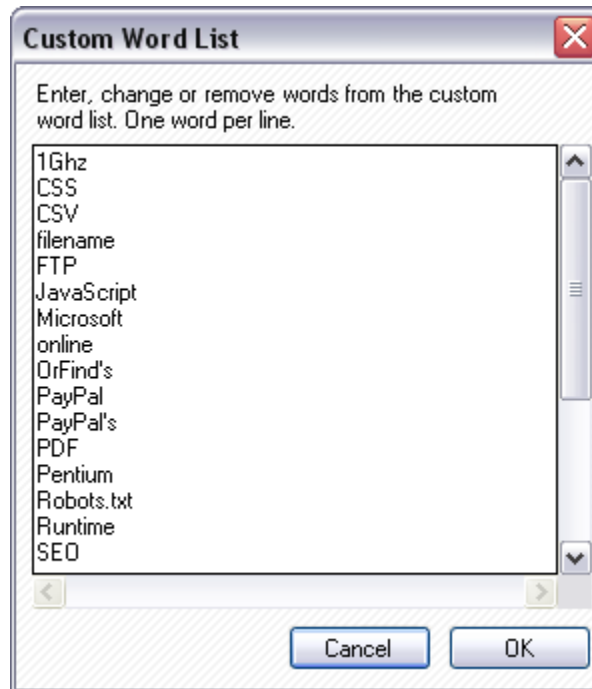
As with the main language dictionary the comparisons using both the Custom Word Lists and the meta-tag technique are not case sensitive.

### 3.3.1 Custom Word List

The Custom Word List is a private dictionary for words that are specific to your site. The Custom Word List is common to all Projects on the system, regardless of the dictionary language chosen. For example: If you were to add the name of your company "MyCo" to the custom word list, this word could then appear on any of your project pages in any language and it would not be flagged as an incorrect spelling. Keep in mind that if you are accessing pages with dynamic lists, such as parts inventory or

employee directories then you should use the Excluded Pages feature to avoid scanning these pages to prevent the large number of 'false-positive' spelling errors that may result.

The Custom Word List can be viewed by clicking the 'Check-list' button on clicking 'Project | Edit Custom Dictionary' from the menu-bar. Words can be added by entering them into the text box, one word per line. It is possible to copy and paste from the word list using Ctrl-C (copy) and Ctrl-V (paste).



Custom words can quickly be added to the list by selecting them in the [Spell Check](#) result tab. Right-clicking the word will present the option to add the word to the custom word list. Multiple words can be added at once by using Ctrl-Click to select several words at a time.

Alternatively, you can add them by clicking the **SmartSense** button. The *SmartSense* results window will appear with the option of entering all of the words into the Custom Word List. *SmartSense* is a very effective way of adding unique words to the Custom Word List.

### 3.3.2 InSite\_Words Meta Tag

The InSite\_Words meta tag provides a mechanism to add words to the custom dictionary on-the-fly (as InSite crawls a site). This provides a mechanism to share the custom dictionary words between copies of InSite (useful if your organization has several individuals working on one website), or to make special words specific to one particular website.

By using a meta tag, the custom words are stored with the page, but not shown to users and ignored by browsers. InSite is the only software which reads and processes this particular meta tag.

The tag is simple to use and has a very straightforward structure. The following is an example of the *InSite\_words* meta tag:

```
<META name="InSite_words" content="mycompany, yourcompany, Mozilla, Firefox, specialword">
```

The *name* of the meta tag is *InSite\_words*, and the custom word spellings must be in the *content*

attribute. Note that the spellings are enclosed in double quotes and separated by a single comma ",".

The meta tag must be in the <HEAD> section of the webpage. The content attribute can span multiple lines and all spellings are case insensitive. Keep in mind that once a word is added to the custom dictionary by the meta tag, it is kept for the duration of the crawl. If you have a word that appears across all pages, it only needs to be included once on the Root URL page.

## 3.4 Scan Results

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

### SmartSense

Once a site has been analyzed, click the **SmartSense** button. A list of the words detected by *SmartSense* will be displayed. Delete any words that you *do not want* in the **Custom Word List** and then click **OK** to add the remainder to the list. You can try adjusting the sensitivity of *SmartSense* in the [Advanced Project Settings](#) dialog. In general, *SmartSense* will provide you with a list of words that are not in either dictionary, but that its analysis has determined are likely special words perhaps unique to your website, and not spelling errors.

### Clear Results

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

### Spelling Errors

The contents of this tab provide information on potential spelling errors. The Page column contains the URL where the error was detected. The Word column contains the word that caused the error and the Context column contains that word in the phrase. Using the Context column is the easiest way to locate the error in the associated Page.

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

- **Add that word** to the Custom Dictionary. This will prevent it from being detected on successive scans and is a very convenient method of building up the Custom Dictionary.
- **Exclude the page** from future crawls by adding it to the Excluded Pages list. This is a convenient way to avoid pages that generate large numbers of inadvertent errors
- **Set page password or form information.** Refer to the [Protected Pages](#) section for additional information
- **Recheck the spelling** on this page only
- **Open that page** in your default browser
- **Open that page** in your HTML editor

### Link Errors

The Page column provides the URL of the page that cannot be accessed.

The Error column displays the type of error encountered. Note that not all errors are from missing or invalid references. You may get a 'Timeout Error' if the site is extremely busy or very slow. The number preceding the error message is the HTTP response code received from your site's server.

The Referrers column provides the URL of the page containing the reference to the inaccessible page.

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

- **Open the referring page.** This is a convenient method of examining the potential source of

the problem by locating a reference error.

- **Exclude the page** from future crawls by adding it to the Excluded Pages list. This is a convenient way to avoid pages that generate large numbers of inadvertent errors
- **Set page password or form information.** Refer to the [Protected Pages](#) section for additional information
- **Open that page** in your default browser
- **Open that page** in your HTML editor

### Spell Checked Pages

This tab contains a complete list of all of the pages accessed in the analysis that were checked for spelling errors. You can right click on any entry to open a page using the default browser or set page password and form information.

### Verified Links

This tab contains a complete list of all of the pages/files that were accessed to check for link validity. These files are any content that make up your website, such as pages, images, zip files and other similar files. You can right click on any entry to try and open a page using the default browser.

### Word/Phrase Density

The contents of this tab provide detailed information on the usage and placement of each word. Click on an **entry in the Page panel** to review the **Word/Phrase statistics** for that page. The contents of this tab are highly dependent on the settings in the [Advanced Project Settings](#) dialog. Please refer to this section for details on how the words and phrases are accumulated. The following is a description of the columns associated with the Word/Phrase Density tab:

|                    |   |
|--------------------|---|
| <b>Word/Phrase</b> | The word or phrase that was detected                                    |
| <b>Count</b>       | The number of occurrences of this particular word or phrase on the page |
| <b>Density</b>     | The relative number of occurrences of this word or phrase               |
| <b>Body</b>        | Indicates that the word/phrase was found in the page body               |
| <b>Alt</b>         | Indicates that the word/phrase was found in an ALT attribute            |
| <b>Meta</b>        | Indicates that the word/phrase was found in a META tag                  |

## 3.5 Settings

Enter topic text here.

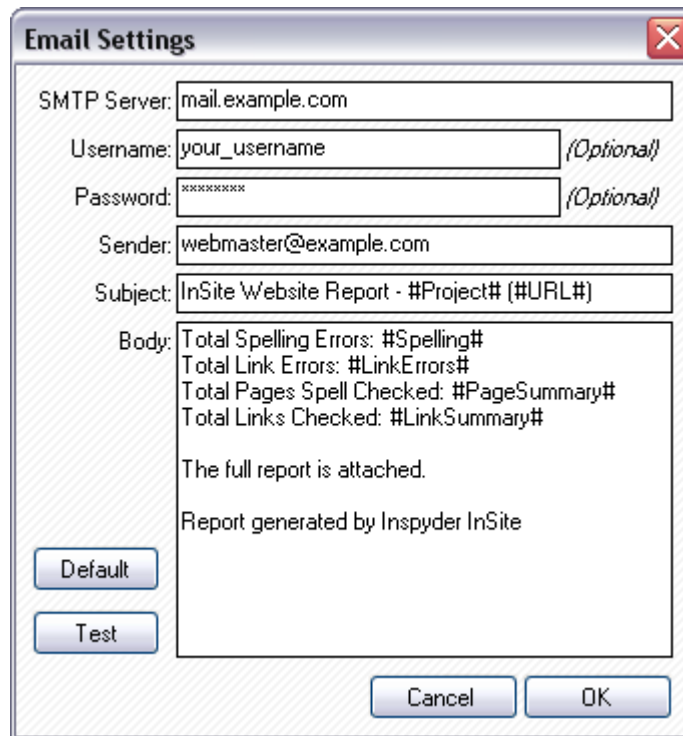
### 3.5.1 Email Settings

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

Enter the **SMTP Server** (mail server) that is used for your outgoing email, and fill out the **Sender** field with the email address to be used as the sender. These settings will be used when automatically exporting the results of a [scheduled analysis](#). The recipient's address is setup when setting InSite to run on a scheduled basis.

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

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



The image shows a dialog box titled "Email Settings" with a close button (X) in the top right corner. It contains several input fields and buttons:

- SMTP Server: mail.example.com
- Username: your\_username (Optional)
- Password: \*\*\*\*\* (Optional)
- Sender: webmaster@example.com
- Subject: InSite Website Report - #Project# (#URL#)
- Body: Total Spelling Errors: #Spelling#  
Total Link Errors: #LinkErrors#  
Total Pages Spell Checked: #PageSummary#  
Total Links Checked: #LinkSummary#  
The full report is attached.  
Report generated by Inspyder InSite
- Buttons: Default, Test, Cancel, OK

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

The following 'tags' can be used within emails:

|                      |  |
|----------------------|--|
| <b>#Date#</b>        | Creation date of the report                  |
| <b>#Time#</b>        | Creation time of the report                  |
| <b>#Project#</b>     | The project name used within InSite          |
| <b>#URL#</b>         | The Root URL of the project                  |
| <b>#Spelling#</b>    | The number of spelling mistakes found.       |
| <b>#LinkErrors#</b>  | The number of broken links found             |
| <b>#PageSummary#</b> | The number of pages that were spell checked  |
| <b>#LinkSummary#</b> | The number of unique links that were checked |

### 3.5.2 Scheduler

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

**Schedule InSite**

Available Projects:

- Acme Parts
- Acme Parts Developer
- Acme Parts Intranet

Report Components

- Spelling Errors
- Link Errors
- Link Summary
- Page Summary
- Word Count
- Word Density

Export Format

- CSV Output
- HTML Output

Default  
Inspyder

Destination

- File
- Email: someone@acmeparts.com

Schedule Options

- Daily
- Weekly
- Monthly

Task Name: Weekly Website Checkup

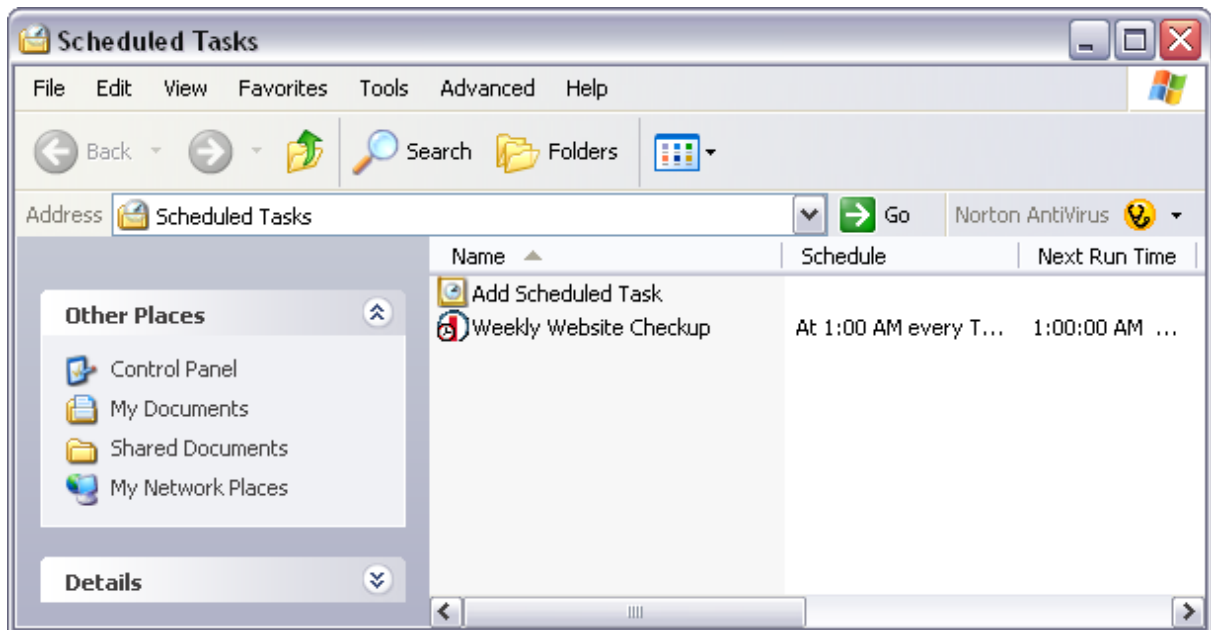
Comment: A task to check our site each week

Start On: 8/17/2006 1:00:00 AM

View Schedule Add Task Close

The **Destination** section is used to specify output to a **File** for later review, or to specify an **Email** address. Refer to the [Email Settings](#) section for additional fields that are used by InSite to construct the email.

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



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

|                         |  |
|-------------------------|--|
| <b>-spelling</b>        | Selects Spell Check content from Scan Results to be included in output                 |
| <b>-linkerrors</b>      | Selects Link Errors content from Scan Results to be included in output                 |
| <b>-wordcountbyword</b> | Selects Word Count by Word from Scan Results to be included in output                  |
| <b>-wordcountbypage</b> | Selects Word Count by Page from Scan Results to be included in output                  |
| <b>-worddensity</b>     | Selects Word Density from Scan Results to be included in output                        |
| <b>-pagesummary</b>     | Selects Crawled HTML Content from Scan Results to be included in output                |
| <b>-linksummary</b>     | Selects Verified Links content from Scan Results to be included in output              |
| <b>-output=</b>         | Specified the output file for the results (i.e. C:\MyFile.html or someone@company.com) |
| <b>-outtype=</b>        | Specifies output format; CSV, PDF or Excel (i.e. -outtype=PDF)                         |
| <b>-project=</b>        | The project name to be analyzed (i.e. -project=Inspyder Local)                         |

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.pdf")

Note to Windows XP and 2003 users: The scheduled tasks will not run on an account with a blank password. For a workaround to this please see our online Knowledge Base.

#### Scheduled Execution Return Codes:

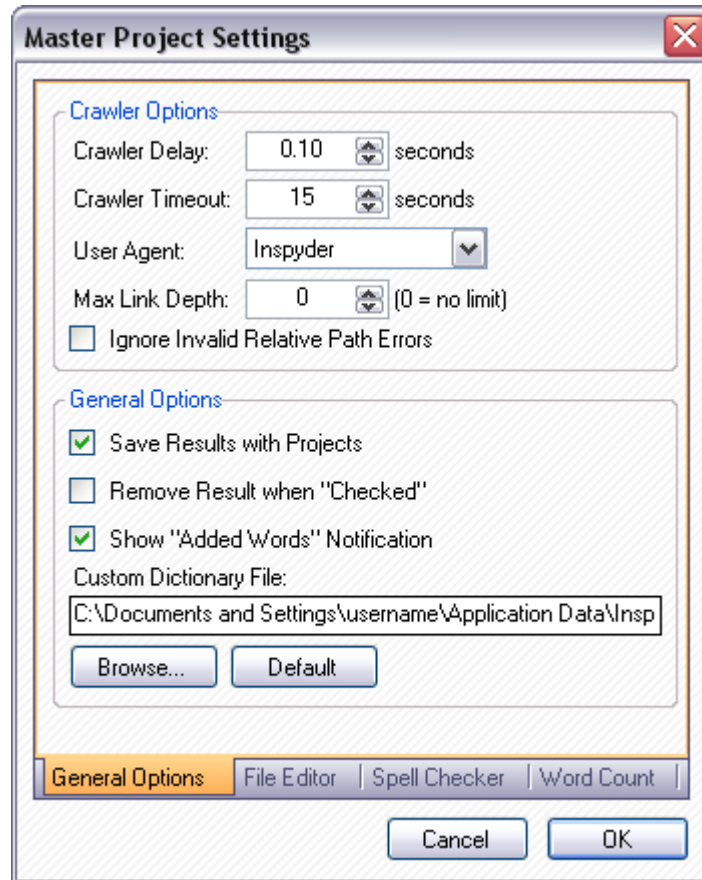
These are the codes returned to the Windows Task Scheduler after the InSite task has been executed. You can view them in the Scheduler window to help troubleshoot any problems.

- 0** Success (no errors)
- 1** Not registered  
**Solution:** Run the program interactively and enter your registration information when prompted.
- 2** Report Generation Error  
**Solution:** If exporting to email, verify that the SMTP server is configured and that a valid return email address is used. If not exporting to email, verify that a valid template is chosen. Try running the report in interactive mode, and check for errors.
- 3** Crawler Error

**Solution:** The crawler detected some type of unhandled error during crawling. Run this project in interactive mode and check for errors (corrupt HTML, etc.).

### 3.5.3 Master Project Settings

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



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

## 4 About Inspyder Software

Founded in 2004, [Inspyder Software](#) is a leading provider of web crawling technologies for content analysis and website management. These technologies have formed the foundation of our software products and made 'Inspyder' a recognizable brand in the software industry. Our software is designed from the ground up using the latest Microsoft .NET technologies to deliver effective, easy to use products.

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

## 5 Product License Agreement

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

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

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

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

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

**USAGE RESTRICTIONS:** The product is licensed only for Your personal use. You cannot use the product to provide a service to others. To be clear, you can use the product as part of your web development activities, including web development for others. However, you cannot use InSite to provide an exclusive service such as using InSite 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 © 2008 Inspyder Software Inc.

# Index

## - A -

Additional Root URLs 7

## - B -

Button Bar 2, 3

Buttons

Custom Word List 3

Exit Program 3

Export 3

Help 3

New Project 3

Save Project 3

Start Analysis 3

Stop Analysis 3

## - C -

Clear Results 15

comma delimited file 11

Command Line Arguments

-linkerrors 17

-linksummary 17

-output= 17

-outtype= 17

-pagesummary 17

-project= 17

-spelling 17

-template= 17

-wordcount 17

-worddensity 17

Crawler Options

Link Depth Limit 7

CSV 11

CSV Output 11

Custom Word List Button 3

Custom words 13

Customized HTML Reports 11

## - D -

Dictionary Language 4

## - E -

Email Settings

Body 16

Sender 16

SMTP Server 16

Subject 16

Excluded Pages 4

Exit Program Button 3

Export Button 3

Export Options 11

Export Options, CSV Output 11

Export Options, Customized HTML Reports 11

Export Options, Email 11

Export Options, File 11

Export Options, HTML Output 11

Export Options, HTML Template 11

## - F -

Features and Benefits 1

File Editor Options

Default Filename 7

Editor 7

Parameter Separator 7

Path 7

## - H -

Help Button 3

HTML Output 11

HTML Tags

#Date# 11

#LinkErrors# 11

#LinkSummary# 11

#PageSummary# 11

#Project# 11

#Spelling# 11

#Time# 11

#URL# 11

#WordCount# 11

#WordDensity# 11

**- I -**

InSite\_Words 14  
Inspyder Software Inc. 20  
Internet Usage Management 20  
Introduction 1

**- L -**

Link Errors  
    Exclude page 15  
    Open page 15  
    Open referring page 15  
    Set form information 15  
    Set page password 15

**- M -**

meta tag 14

**- N -**

New Project Button 3

**- P -**

private dictionaries 14  
private dictionary 13  
Product License  
    Copy Restrictions 21  
    Definitions 21  
    Grant of License 21  
    Jurisdiction 21  
    Limitations of Remedies 21  
    Limited Warranty 21  
    Transfer Restrictions 21  
Project Settings 2  
    Dictionary Language 4  
    Exclusion List 4  
    Go Button 4  
    Protected Pages 4  
    Root URL 4  
Protected Pages 4

**- R -**

Return Codes 17  
Root URL 4

**- S -**

Save As 3  
Save Project Button 3  
Saved Projects 2  
Scan Results 2  
Scheduled Execution Return Codes 17  
SmartSense 13, 15  
SMTP Server 16  
Spell Checked Pages 15  
Spelling Errors  
    Add word 15  
    Exclude page 15  
    Link Errors 15  
    Open page 15  
    Recheck spelling 15  
    Set form information 15  
    Set page password 15  
Start Analysis Button 3  
Stop Analysis Button 3

**- V -**

Verified Links 15

**- W -**

Word Count  
    by Page 12  
    by Word 12  
    Count 12  
    Unique Count 12  
Word/Phrase Density  
    Alt 15  
    Body 15  
    Count 15  
    Density 15  
    Meta 15  
    Word/Phrase 15



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

Tel: 888.732.6134

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