**PWGSC WPSS Validation Tool**

**User Guide**

Version 6.13.0

April 21, 2021

Contents

[1 Introduction 4](#_Toc39130661)

[1.1 Tool Limitations 4](#_Toc39130662)

[1.2 Tool Risks 4](#_Toc39130663)

[2 Installation 6](#_Toc39130664)

[2.1 System Requirements (Windows) 6](#_Toc39130665)

[2.2 Installing the WPSS Tool 6](#_Toc39130666)

[2.3 Optional Requirements 6](#_Toc39130667)

[2.3.1 Chrome Headless User Agent 6](#_Toc39130668)

[2.3.2 Pa11y accessibility tool 7](#_Toc39130669)

[2.3.3 Deque AXE accessibility tool 7](#_Toc39130670)

[2.4 Uninstalling the WPSS Tool 7](#_Toc39130671)

[2.4.1 Uninstalling Perl 8](#_Toc39130672)

[2.4.2 Uninstalling Python 8](#_Toc39130673)

[3 Using the WPSS Validation Tool 8](#_Toc39130674)

[3.1 Site Details Tab 10](#_Toc39130675)

[3.1.1 Entering URLs 10](#_Toc39130676)

[3.1.2 Saving Site Details 12](#_Toc39130677)

[3.1.3 Loading Site Details 12](#_Toc39130678)

[3.2 Login/Logout Tab 12](#_Toc39130679)

[3.2.1 Login/Logout URLs 13](#_Toc39130680)

[3.3 Crawling a Site and Analysing Documents 14](#_Toc39130681)

[3.4 Direct HTML Input Tab 14](#_Toc39130682)

[3.5 URL List Tab 15](#_Toc39130683)

[3.6 Configuration Tab 16](#_Toc39130684)

[3.7 Results Window 18](#_Toc39130685)

[3.7.1 Saving Analysis Results 20](#_Toc39130686)

[3.7.2 Stopping Crawler and Analysis 20](#_Toc39130687)

[3.8 Application Login 20](#_Toc39130688)

[3.9 Web Server Protected Sites 21](#_Toc39130689)

[3.10 Reporting Passes and Fails 21](#_Toc39130690)

[3.11 Capture HTML Content on Errors 22](#_Toc39130691)

[4 Command Line Interface 22](#_Toc39130692)

[4.1 Crawl a Site 22](#_Toc39130693)

[4.1.1 Crawl a site with a login 23](#_Toc39130694)

[4.2 Analyse a block of HTML code 24](#_Toc39130695)

[4.3 Analyse a list of URLs 25](#_Toc39130696)

[4.4 Program Status and Progress 25](#_Toc39130697)

[4.4.1 Viewing Results 25](#_Toc39130698)

[5 Language and Configuration 26](#_Toc39130699)

[5.1 Language Switching 26](#_Toc39130700)

[5.2 Configuration Files 26](#_Toc39130701)

[5.2.1 Domain Network Scope 26](#_Toc39130702)

[5.2.2 Domain Alias Map 26](#_Toc39130703)

[6 Troubleshooting 27](#_Toc39130704)

[6.1 Entry Page Rewritten 27](#_Toc39130705)

[6.2 Forbidden by robots.txt 27](#_Toc39130706)

[6.3 User Agent String 27](#_Toc39130707)

[6.4 Perl Command Line Interpreter Error 27](#_Toc39130708)

[6.5 500 Internal Server Error 28](#_Toc39130709)

# Introduction

The PWGSC WPSS Validation Tool provides web developers and quality assurance testers the ability to perform a number of web site and web page validation tasks at one time. The WPSS Validation Tool crawls a site to find all of the documents, then analyses each one with a number of validation tools. The analysis includes:

1. HTML validation,
2. CSS validation,
3. robots.txt validation,
4. link violation checking,
5. metadata checking,
6. Accessibility checking, WCAG 2.0, Deque AXE, Pa11y,
7. HTML document feature reporting, for example, forms,

Open Data checking using a separate application tool.

Details of these analysis tools are available in the document, *Web-WPSS Standalone Validation Tool Testcases*.

## Tool Limitations

There are limitations within the tool that may affect the validation results.

* The tools do not support JavaScript. The Validation Tool may not accurately crawl or analyse sites that rely on JavaScript.
* The tools do not use a standard Web browser User Agent name. The Validation Tool may not accurately crawl or analyse sites that rely on the User Agent name.
* The tools’ default behaviour is to respect robots directives. If a site has a robots.txt to prohibit crawlers, the WPSS Validation Tool will not validate the site’s documents. You can configure the WPSS Validation Tool to ignore robots.txt directives. For more information, see the section, “Configuration Tab”.
* Some of the output of the individual Validation Tools may be in English only. This tool is using third party software components. The source of these components is available only in the language that it was authored.

## Tool Risks

The validation tool includes a crawler that follows links and retrieves web documents from sites. Care should be taken with this tool to:

* Ensure entry page URLs for the sites are accurate so the crawler does not go beyond the site being analysed.
* Ensure that document retrieval from the site being analysed, and any links to other sites, do not impose excessive loads on web servers and the network.

# Installation

The WPSS Validation Tool requires Perl, Python and Java distributions installed. A number of optional components have additional requirements.

## System Requirements (Windows)

To install the WPSS Validation Tool on Windows, the follow requirements must be met:

* Windows 7 or later,
* Oracle Java runtime environment 1.8.0 (other versions may not work), or OpenJDK version 11 or newer,
  + Oracle Java is available at <https://www.java.com/en/download/>
  + OpenJDK is available at <https://openjdk.java.net/>
* Latest version of Python 2.7.x (does not work with Python 3),
  + available from http://python.org/downloads/
* Strawberry Perl 5.26.1 or later, either 32 or 64 bit, (does not work with ActiveState Perl),
  + available from <http://strawberryperl.com>
  + Only one installation of Perl on the system. Multiple installations may cause problems.

## Installing the WPSS Tool

To install the WPSS Validation Tool, double-click the WPSS\_Install.exe file and follow the instructions on the screen.



The default installation folder for the WPSS Validation Tool is C:\Program Files\WPSS\_Tool.

## Optional Requirements

The following are the requirements for optional components of the WPSS Tool.

### Chrome Headless User Agent

The following are requirements for using the Chrome headless user agent. This user agent is newer and more feature rich than the default [PhantomJS](https://phantomjs.org/) user agent.

* Node version 8 or newer
  + available from <https://nodejs.org/en/download/>
* Chrome browser version 69 or newer
  + available from <https://www.google.com/chrome/>
* ChromeDriver node module that matches the Chrome browser version
  + List locally installed chromedriver version ‘npm list chromedriver -g’
  + List all available chromedriver versions ‘npm view chromedriver versions’
  + Remove module ‘npm unistall -g chromedriver’
  + Install a specific version ‘npm install -g [chromedriver@89.0.0](mailto:chromedriver@89.0.0)’
* Puppeteer-core node module. Installed via command prompt.
  + npm install –g puppeteer-core

### Pa11y accessibility tool

The following are requirements for using the [Pa11y accessibility tool](https://pa11y.org/). This tool is required for the Pa11y testcase profile of the WPSS Tool.

* Pa11y node module (required for optional Pa11y accessibility tool) Installed via command prompt.
  + npm install –g pa11y
  + <https://github.com/pa11y/pa11y>

### Deque AXE accessibility tool

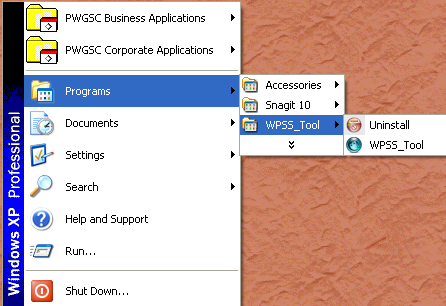
The following are requirements for using the [Deque AXE accessibility](https://github.com/dequelabs/axe-core-npm/tree/develop/packages/cli) command line tool. This tool is required for the Deque AXE testcase profile of the WPSS Tool.

* Deque AXE core module and command line interface (CLI) module. Installed via command prompt.
  + npm install @axe-core/cli -g
  + Install the required chromedriver module (see above).

## Uninstalling the WPSS Tool

To remove the WPSS Tool from a workstation, run the uninstall script.

Go to **Start > Programs > WPSS\_Tool > Uninstall**.



### Uninstalling Perl

**To remove the Perl installation:**

1. Go to **Start > Settings > Control Panel**.
2. Click **Add or Remove Programs**.
3. Locate Strawberry Perl and click **Remove**.

### Uninstalling Python

**To remove the Python installation:**

1. Go to **Start > Settings > Control Panel**.
2. Click **Add or Remove Programs**.
3. Locate Python and click **Remove**.

# Using the WPSS Validation Tool

**To start the PWGSC WPSS Validation Tool:**

1. Go to **Start > Programs > WPSS\_Tool**.
2. Click the **WPSS\_Tool icon**.

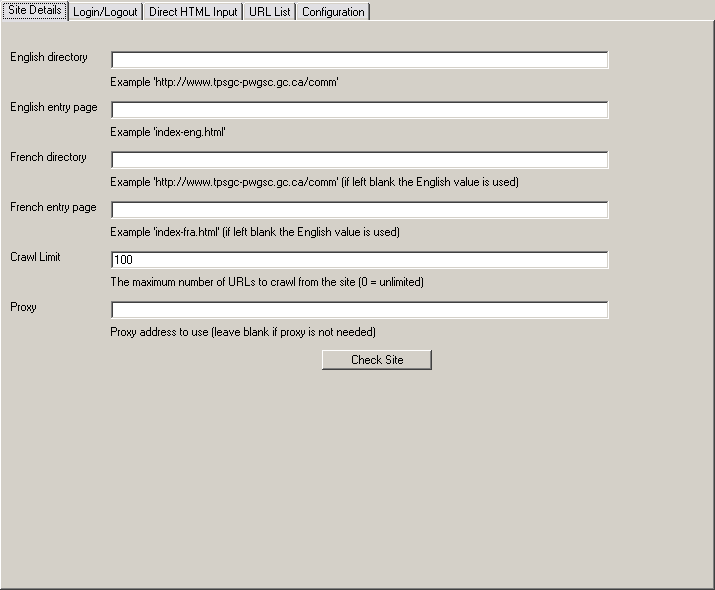
Alternatively, using Windows Explorer, navigate to the **C:\Program Files\WPSS\_Tool** folder and double click the **wpss\_tool.pl** file.

The main window consists of five tabs:

* **Site Details** is for entering URL information for crawling a site.
* **Login/Logout** is for entering site or application login/logout information.
* **Direct HTML Input** is for pasting in HTML code.
* **URL List** is for entering a list of URLs.
* **Configuration** is for configuring the WPSS Validation Tool options.

## Site Details Tab

To crawl and analyse the documents on a site, enter the URLs for the site and entry page addresses into the WPSS Validation Tool. Use the **Site Details** tab to enter this information. You need to enter the following details to properly define a site for crawling.



### Entering URLs

Enter the English and French URLs as well as the entry page names for the site.

**English Directory** – The English URL to the directory containing all of the site’s documents (HTML documents, image files, CSS files, etc.).

For example:

* <http://www.tpsgc-pwgsc.gc.ca/comm/index-eng.html>   
  The English Directory is <http://www.tpsgc-pwgsc.gc.ca/comm>/
* <http://source.tpsgc-pwgsc.gc.ca/policy/text/index-e.html>  
  The English Directory is <http://source.tpsgc-pwgsc.gc.ca/policy>/
* <http://id-reo.tpsgc-pwgsc.gc.ca/public/ffchrdspl.do?page=acclhm&lang=en>  
  The English Directory is <http://id-reo.tpsgc-pwgsc.gc.ca/public>/
* <http://crownassets.pwgsc.gc.ca/mn-eng.cfm?snc=hm>  
  The English Directory is <http://crownassets.pwgsc.gc.ca>/

**English Entry Page** – The directory and file name of the English entry page. This is not the full URL, just the directory, file name and arguments portion. To find the entry page name, go to the site’s splash page to get the page referenced by the English language button.

For example:

* <http://www.tpsgc-pwgsc.gc.ca/comm/index-eng.html>   
  The English Entry Page is **index-eng.html**.
* <http://source.tpsgc-pwgsc.gc.ca/policy/text/index-e.html>  
  The English Entry Page is **text/index-e.html**.
* <http://id-reo.tpsgc-pwgsc.gc.ca/public/ffchrdspl.do?page=acclhm&lang=en>  
  The English Entry Page is **ffchrdspl.do?page=acclhm&lang=en**.
* <http://crownassets.pwgsc.gc.ca/mn-eng.cfm?snc=hm>  
  The English Entry Page is **mn-eng.cfm?snc=hm**.

**French Directory** – This is the French URL to the directory containing all of the site’s documents (HTML documents, image files, CSS files, etc.). If the French Directory is the same as the English Directory, leave the field empty.

For example:

* <http://www.tpsgc-pwgsc.gc.ca/comm/index-eng.html>  
  Leave the French Directory field left blank if it is the same as the English directory.
* <http://source.tpsgc-pwgsc.gc.ca/politique/text/index-f.html>  
  The French Directory is <http://source.tpsgc-pwgsc.gc.ca/politique>. Since the directory name, “politique”, differs from the English name, “policy”, you must specify the French directory.
* <http://id-reo.tpsgc-pwgsc.gc.ca/public/ffchrdspl.do?page=acclhm&lang=fr>  
  Leave the French Directory field left blank if it is the same as the English directory.
* <http://biensdelacouronne.tpsgc.gc.ca/mn-fra.cfm?snc=hm>  
  The French Directory is <http://biensdelacouronne.tpsgc.gc.ca>. The site has separate English and French domain names.

**French Entry Page** – The directory and file name of the French entry page. This is not the full URL; only the directory, file name and arguments portion. To find the entry page name, go to the site’s splash page to get the page referenced by the French language button.

For example:

* <http://www.tpsgc-pwgsc.gc.ca/comm/index-fra.html>  
  The French Entry Page is **index-fra.html**.
* <http://source.tpsgc-pwgsc.gc.ca/politique/text/index-f.html>  
  The French Entry Page is **text/index-f.html**
* <http://id-reo.tpsgc-pwgsc.gc.ca/public/ffchrdspl.do?page=acclhm&lang=fr>  
  The French Entry Page is **ffchrdspl.do?page=acclhm&lang=fr.**
* <http://biensdelacouronne.tpsgc.gc.ca/mn-fracfm?snc=hm>  
  The French Entry Page is **mn-fra.cfm?snc=hm**.

**Crawl Limit** – The maximum number of URLs the WPSS Validation Tool will retrieve and analyse from the site. The URLs include all file types, HTML, PDF, images, CSS files, etc. To crawl the entire site, enter 0 (zero).

### Saving Site Details

You can save the site details in a configuration file for sharing or easy access if you want to use the WPSS Validation Tool on the site again. Note that the Tool only saves the information from the **Site Details** tab.

**To save the site configuration:**

1. Go to **File > Save Site Config**.
2. Select a folder and file name for the configuration file.
3. Select **OK**.

To best manage site configuration files, it is suggested that you save the configuration files in the **C:\Program Files\WPSS\_Tool\profiles** folder.

### Loading Site Details

You can load a previously saved site configuration. Loading a saved configuration file loads the **Site Details** tab fields. Once loaded, you can modify the information if required.

**To open a saved site configuration:**

1. Go to **File > Load Site Config**.
2. Locate the folder and file.
3. Select **OK**.

## Login/Logout Tab

The WPSS Validation Tool can crawl application sites that have a simple login page consisting of a single form for login credentials. If the site has a login, then you must include additional data items. If the site does not have a login, these fields can remain blank.



### Login/Logout URLs

**English Login Page** – The directory and file name of the English login page. This is not the full URL; just the directory, file name and arguments portion. For example:

* <http://crownassets.pwgsc.gc.ca/mn-eng.cfm?snc=hm>   
  The English Login Page is **mn-eng.cfm?snc=ouv-log&ouvlog=1**.

**English Logout Page** – The directory and file name of the English logout page. This is not the full URL; just the directory, file name and arguments portion. For example:

* <http://crownassets.pwgsc.gc.ca/mn-eng.cfm?snc=hm>  
  The English Logout Page is **mn-eng.cfm?snc=ouv-log&ouvlog=1&vndsld=0**.

**French Login Page** – The directory and file name of the French login page. This is not the full URL; just the directory, file name and arguments portion. For example:

* <http://biensdelacouronne.tpsgc.gc.ca/mn-eng.cfm?snc=hm>  
  The French Login Page is **mn-fra.cfm?snc=ouv-log&ouvlog=1**.

**French Logout Page** – The directory and file name of the French logout page. This is not the full URL; just the directory, file name and arguments portion. For example:

* <http://biensdelacouronne.tpsgc.gc.ca/mn-eng.cfm?snc=hm>  
  The French Logout Page is **mn-fra.cfm?snc=ouv-log&ouvlog=1&vndsld=0**.

**Login Form Name** – If the login page contains more than one form, for example a search form along with the login form, then you must specify the name of the form. The name is the value of the **name** or **id** field on the form. If there is only one form on the login page, you can leave the field blank.

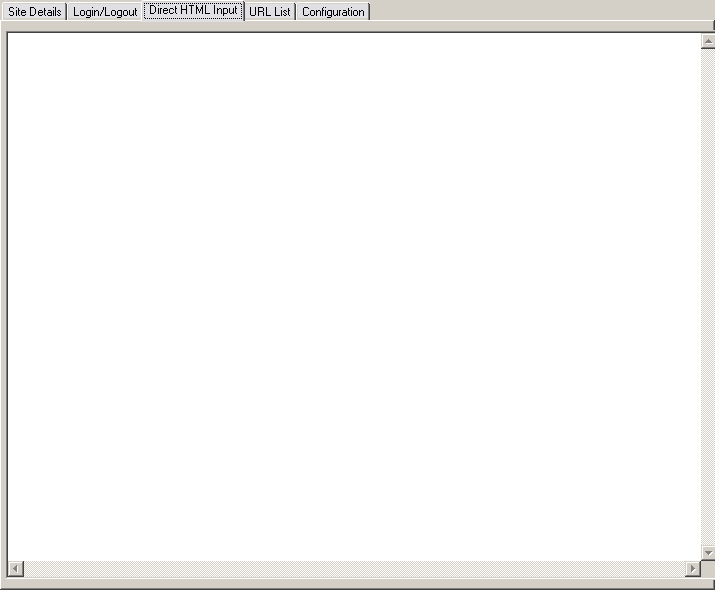
## Crawling a Site and Analysing Documents

Once you have completed the configuration details for the site you want to crawl, click **Check Site** on the **Site Details** tab.

The WPSS Validation Tool begins to crawl the site and analyses each of its documents. The results appear in the Results Window.

## Direct HTML Input Tab

You can analyse a single HTML page or selection of HTML if you need to verify one particular page or chunk HTML code. To analyse specific HTML, select the **Direct HTML Input** tab and paste your HTML code into the text area. Select **Check** to begin the analysis. Results appear in the Results Window.

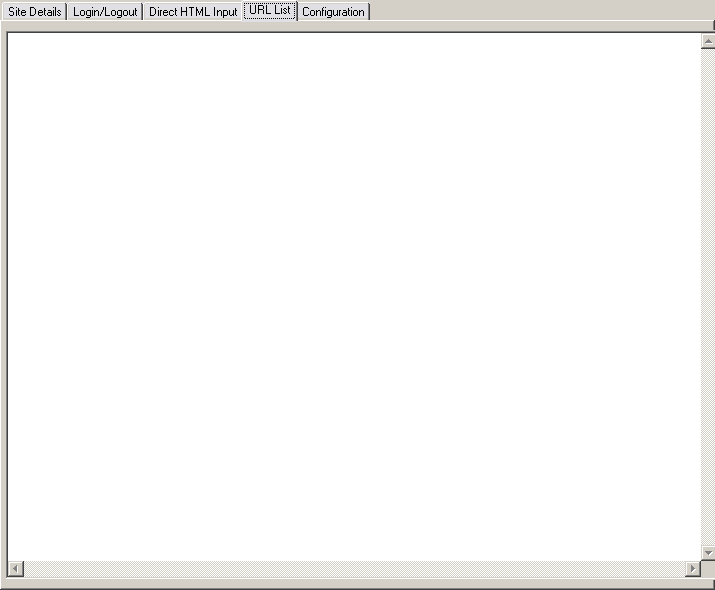


You can also configure how the WPSS Validation Tool analyses the HTML by going to the **Configuration** tab. For more information on the configuration options, see the section, “Configuration Tab”.

## URL List Tab

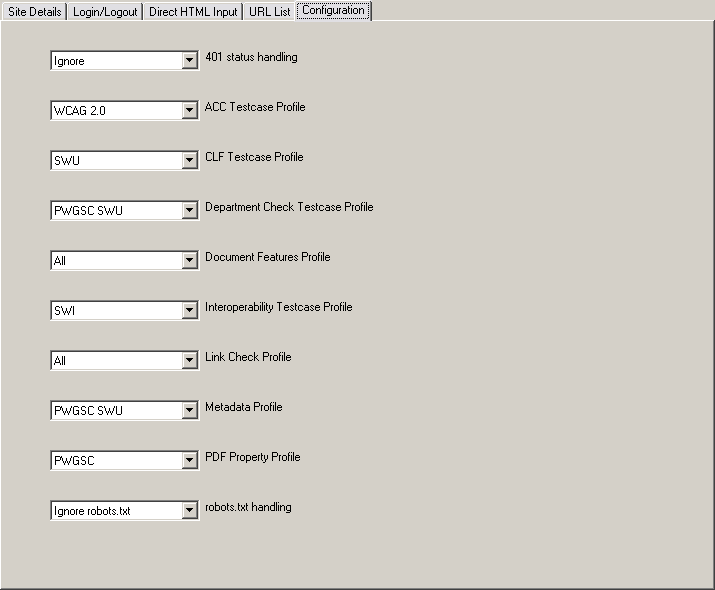
The **URL List** tab enables you to enter a list of URLs for analysis. You can either type the URLs directly into the text area or load them from a file.

To begin the analysis, click **Check URL List**. The results appear in the Results Window.



## Configuration Tab

The **Configuration** tab enables you to select options for the analysis tools.



**401 status handling** – Controls the behaviour of the WPSS Validation Tool if it encounters a *401 – Not Authorized* code message.

Options include:

* Ignore
* Prompt for credentials

**ACC Testcase Profile** – Select the desired accessibility test case profile.

Options include:

* WCAG 2.0
* TBS WCAG 2.0 Quick Check (WAAT Tool profile)
* None

**CLF Testcase Profile** – Select the desired look and feel test case profile.

Options include:

* SWU (Standard on Web Usability)
* PWGSC Intranet
* TBS SWU Intranet
* CLF 2.0
* None

**Department Check Testcase Profile** – Select the appropriate content check test case to validate.

Options include:

* PWGSC SWU
* PWGSC Common
* Common
* PWGSC Intranet
* None

**Department Features Profile** – Select the appropriate feature profile for the report.

Options include:

* All – Report all HTML features
* None

**Interoperability Testcase Profile** – Select the appropriate interoperability test case profile.

Options include:

* SWI – Standard on Web Interoperability
* None

**Link Check Profile** – Select the appropriate link check test case profile.

Options include:

* Common – Broken links, cross languge links
* All – all link violations
* Errors – report broken links only
* IPV4 – report IP addresses used as domain names
* None

**Metadata Profile** – Select the appropriate metadata profile to use to validate.

Options include:

* PWGSC SWU – PWGSC Standard on Web Usability
* TBS SWU – TBS Standard on Web Usability
* PWGSC – The PWGSC CLF 2.0 Metadata profile.
* TBS CLF 2.0 – The TBS CLF 2.0 profile.
* None – No Metadata required.

**PDF Property Profile** – Select the appropriate PDF property profile to use to validate.

Options include:

* PWGSC – The PWGSC profile.
* None – No properties required.

**Web Analytics Profile** – Select the appropriate Web Analytics profile to use to validate.

Options include:

* TBS Web Analytics
* None

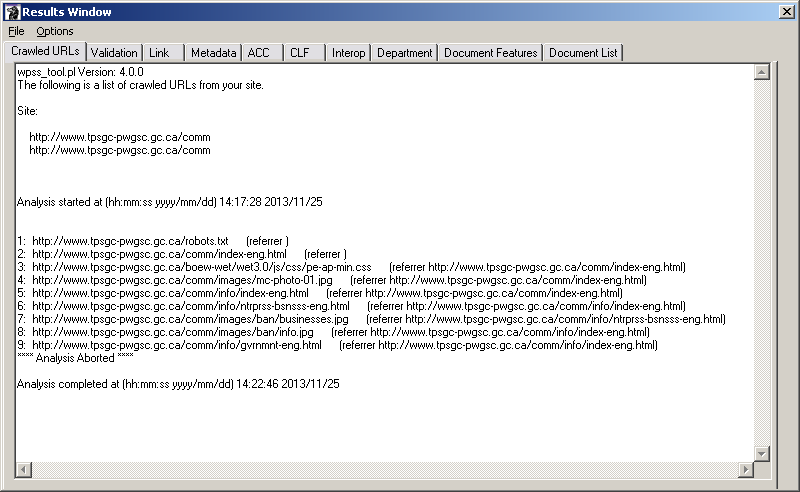
**Robots.txt handling** – Controls the behaviour of the WPSS Validation Tool if it encounters a *403 – Forbidden by robots* code.

Options include:

* Ignore robots.txt
* Respect robots.txt

## Results Window

The **Results Window** contains a number of tabs containing the output of an individual analysis from the WPSS Validation Tool. The output in each tab includes a header that lists the English and French site directory URLs along with the time and date when the analysis started.



**Crawled URLs** tab – Provides the list of URLs the WPSS Validation Tool analysed. It lists the referrer page to indicate how the crawler reached a particular page. Use this tab to monitor the WPSS Validation Tool to ensure it is actively crawling and analysing a site’s pages.

**Validation** tab – Contains the output of the HTML/XHTML, CSS and robots.txt validation tools. The WPSS Tool does NOT perform HTML5 mark-up validation.

**Link** tab – Contains the output of the link check.

**Metadata** tab – Contains the output of the metadata check.

**ACC** tab – Contains the output of the accessibility check.

**CLF** tab – Contains the output of the look and feel check.

**Interop** tab – Contains the output of the interoperability check.

**Department** tab – Contains the output of the department check.

**Document Features** tab – Contains a list of documents that have HTML features such as forms, tables, etc.

**Document List** tab – The WPSS Validation Tool writes information to this tab after completing the site analysis. It contains the sorted list of documents found in the site, and all documents including HTML/XHTML, PDF, images, CSS, etc.

The WPSS Validation Tool includes the time and date at the end of the report in each tab.

### Saving Analysis Results

To save the analysis results, in the Results Window go to **File > Save As**. Select the file name and folder path in the file chooser dialog.

The results are stored in a number of files, one for each result tab. Each file name contains a suffix identifying the report type. For example, if you save the results in the file www\_corp.txt, the actual results files are:

* www\_corp\_**acc**.txt
* www\_corp\_**clf**.txt
* www\_corp\_**crawl**.text
* www\_corp\_**dept**.txt
* www\_corp\_**feat**.txt
* www\_corp\_**int**.txt
* www\_corp\_**link**.txt
* www\_corp\_**meta**.txt
* www\_corp\_**urls**.txt
* www\_corp\_**val**.txt

In addition to the above results, some HTML reports are also generated and saved with the result files. For example:

* www\_corp\_img.html – Image Details report
* www\_corp\_h.html – Headings Outline report

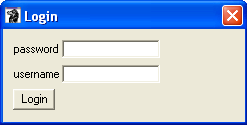
It is suggested that you use the same base name of the results as the site profile name, and save the results in the in the **C:\Program Files\WPSS\_Tool\results** folder.

### Stopping Crawler and Analysis

If you need to stop the analysis while it is running, in the Results Window, go to **Options > Stop Crawl**. This stops the WPSS Validation Tool after processing the current document. The results includes note to the bottom of each output tab indicating that the analysis was aborted.

## Application Login

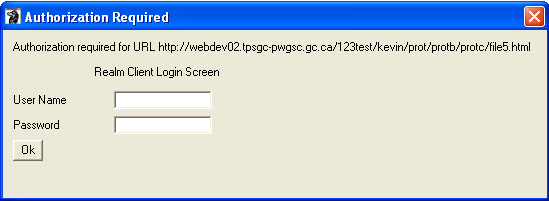
If the site being analysed is an application with a login page specified in the **Site Details** tab, the WPSS Validation Tool must be able login to the application to view pages behind the login. When the WPSS Validation Tool reaches the English login page, a Login window appears to provide application credentials. All of the input fields listed in the login form are presented to the user in a login window. The order of the fields may be different than what appears on the web page.



After entering the login credentials, click **Login**. The WPSS Validation Tool attempts to login to the application site. You are not prompted for login credentials for the French login page; the English credentials are reused.

## Web Server Protected Sites

If the site being analysed contains directories that are password protected by the web server using Basic Authentication, the WPSS Validation Tool must be able to provide the credentials to view the documents. When the WPSS Validation Tool reaches a document that requires authentication, a Login window appears to provide the credentials.



After entering the login credentials, click **OK**. The WPSS Validation Tool attempts to access the document(s).

## Reporting Passes and Fails

The default behaviour of the analysis tools is to report only URLs that fail checks. You can view results for both passes and fails. To see both passes and fails, in the WPSS Validation Tool window, go to **Options > Report Fails and Passes**. The URL for documents that pass checks are recorded in the results output.

To see only failed pages, go to **Options > Report Fails Only**.

## Capture HTML Content on Errors

Sometimes having the URL of a document with errors is not sufficient, especially if the content is generated by an application on the fly. To capture the HTML content of a document with errors, it is necessary to stop the analysis process when the error occurs.

To stop an analysis when an error in an application occurs, in the WPSS Validation Tool window, go to **Options > Stop on Error**.

When selected, the crawler and analysis tools will stop when it detects an error. Once stopped, you have the option to save the HTML content, or to continue on with the next document. The saved content includes HTML comments at the top of the file providing the full URL of the document along with the WPSS Validation Tool error message.

To turn off this option, in the WPSS Validation Tool window go to **Options > Continue on Error**.

# Command Line Interface

The WPSS Validation Tool is available from both the GUI interface and from the command prompt.

To access the command line version:

1. Go to **Start > Programs > Accessories > Command Prompt**.
2. Change to the **C:\Program Files\WPSS\_Tool** directory.
3. Run the program **wpss\_tool.pl**.

## Crawl a Site

To crawl a site and validate documents, use the command…

wpss\_tool.pl –cli –c <crawl file>

…where <crawl file> is the path to a file containing the site details and WPSS Validation Tool configuration. The configuration file is a plain text file containing the following variable/value:

| **Variable** | **Value** |
| --- | --- |
| site\_url\_eng | The URL to the English entry page. |
| site\_url\_fra | The URL of the French entry page. If the French Entry Page is the same as the English Entry Page, you do not need to include it in the command. |
| crawllimit | The maximum number of URLs the WPSS Validation Tool retrieves and analyses from the site. The URLs include all file types, HTML, PDF, images, CSS files, etc. To crawl the entire site, use a value of 0 (zero). |
| output\_file | The folder and file name where the analysis results are written. |
| httpproxy | Sets the proxy server for HTTP traffic. Only use this variable when the user’s workstation requires a proxy to reach the Internet. |

Example file contents:

crawllimit 100

site\_url\_eng http://webdev02.tpsgc-pwgsc.gc.ca/comm/index-eng.html

site\_url\_fra http://webdev02.tpsgc-pwgsc.gc.ca/comm/index-fra.html

output\_file results/webdev02\_comm

httpproxy

### Crawl a site with a login

To crawl a site or application that has a login, you need to provide the WPSS Validation Tool with additional configuration items. All of the configuration items for crawling a site must be provided along with the following:

|  |  |
| --- | --- |
| **Variable** | **Value** |
| loginpagee | The URL of the English login page. |
| logoutpagee | The URL of the English logout page. |
| loginpagef | The URL of the French login page. |
| logoutpagef | The URL of the French logout page. |
| loginformname | If the login page contains more than one form, like a search form with the login form, you need to specify the name of the form. The name is the value of the name/id attribute of the form. If there is only one form on the login page, it does not need to be included in the command. |

Login credentials must also be provided in a credentials file. The credentials file is a text file containing name/value pairs for the login form’s text and password fields. The variable names match the name attribute of the various <input> fields.

Sample credentials file:

username admin

password abc123

To crawl a site or application that has a login, use the command:

wpss\_tool.pl –cli –c <crawl file> -login <credentials file>

## Analyse a block of HTML code

To analyse a block of HTML mark-up, the command…

wpss\_tool.pl –cli –h <html content file>

…where <html content file> is the path to a file containing the WPSS Validation Tool configuration. The configuration file is a plain text file containing the following variable/value pairs:

|  |  |
| --- | --- |
| **Variable** | **Value** |
| html\_file | The path of the file containing HTML mark-up to analyse. |
| output\_file | The folder and file name where the analysis results are written. |
| httpproxy | The HTTP proxy field allows for setting proxy server for HTTP traffic. Only use this field when the user’s workstation requires a proxy in order to reach the Internet. |

All other lines of the file that begin with either http:// or https:// are the list of URLs to analyse.

Example file contents:

html\_file sample.html

output\_file results/webdev02\_comm

httpproxy

## Analyse a list of URLs

To analyse a list of URLs, use the command…

wpss\_tool.pl -cli –u <url list file>

…where the <url list file> is the path to a file containing the list of URLs and WPSS Validation Tool configuration. The configuration file is a plain text file containing the following variable/value pairs:

|  |  |
| --- | --- |
| **Variable** | **Value** |
| output\_file | The folder and file name where the analysis results are written. |
| httpproxy | The HTTP proxy field allows for setting proxy server for HTTP traffic. Only use this field when the user’s workstation requires a proxy in order to reach the Internet. |

All other lines of the file that begin with either http:// or https:// are the list of URLs to analyse.

Example file contents:

output\_file results/webdev02\_comm

httpproxy

<http://www.tpsgc-pwgsc.gc.ca/comm/index-eng.html>

<http://www.tpsgc-pwgsc.gc.ca/comm/index-fra.html>

## Program Status and Progress

As the command line WPSS Validation Tool runs and analyses documents, the URLs of the documents are printed to the console. Use this to monitor the WPSS Validation Tool to ensure it is actively crawling and analysing a site’s pages.

### Viewing Results

The results of an analysis are stored in a number of files. Each file name includes a suffix identifying the report type. For example, if you save the results in the file webdev02\_comm.txt, the actual results files are:

* webdev02\_comm \_crawl.txt **–** list of analysed URLs by the WPSS Validation Tool.
* webdev02\_comm \_acc.txt **–** details of WCAG 2.0 faults detected.

It is suggested that you use the same base name of the results as the site profile name, and save the results in the in the **C:\Program Files\WPSS\_Tool\results** folder.

You can view the results files with any program that can display plain text files, such as WordPad.

# Language and Configuration

## Language Switching

You can toggle the language of the analysis results using a command line option.

|  |  |
| --- | --- |
| **Option** | **Result** |
| -eng | Analysis results are provided in English. |
| -fra | Analysis results are provided in French. |

If no language selection is made, the language is determined from the operating system.

## Configuration Files

There are a number configuration files that modify the behaviour of the WPSS Validation Tool. Most of these files should remain untouched with the exception of the **wpss\_tool.config** file. This file contains the network scope designation for domains as well as the domain and domain alias mapping. The file is a plain text file with simple name value pairs for configuration parameters.

### Domain Network Scope

This defines the network scope of a specific domain. The network scope is one of:

* Internet
* GC Intranet
* PWGSC Intranet
* Internet development
* GC Intranet development
* PWGSC Intranet development

### Domain Alias Map

This provides a mapping of domain names and their aliases.

# Troubleshooting

## Entry Page Rewritten

|  |  |
| --- | --- |
| **Message** | Entry page http://webdev02.pwgsc.gc.ca/comm/index-eng.html rewritten to http://webdev02.tpsgc-pwgsc.gc.ca/comm/index-eng.html |
| **Cause** | The entry page provided for the site configuration is redirected by the web server. |
| **Correction** | Provide the rewritten entry page URL in the site configuration. |

## Forbidden by robots.txt

|  |  |
| --- | --- |
| **Message** | Failed to get URL <http://ssi-iss.tpsgc-pwgsc.gc.ca/index-eng.html>, error is *403 Forbidden by robots.txt*. |
| **Cause** | The site has a robots.txt file that prohibits crawlers from accessing the site. The Tool honours robots, and unable to analyse the site. |
| **Correction** | Go to the **Configuration** tab and select **Ignore robots.txt** for the **robots.txt handling** option. |

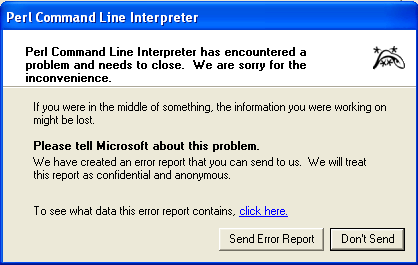
## User Agent String

The validation tools have their own User Agent string that is different from a user’s browser. Some sites look at the User Agent string and present different output depending upon its value, for example, checking for a minimum browser version.

|  |  |
| --- | --- |
| **Message** | Varies |
| **Cause** | The validation tools User Agent string not recognized by the site being analysed. |
| **Correction** | Change site to not depend on a specific User Agent string and/or browser. |

## Perl Command Line Interpreter Error

The WPSS Validation Tool uses a number of Perl modules. These modules may have errors that cause the program to fail. You may encounter the following message:



|  |  |
| --- | --- |
| **Message** | Perl Command Line Interpreter has encountered a problem and needs to close. |
| **Cause** | An error in the supporting Perl modules. |
| **Correction** | Restart the WPSS Validation Tool, if the problem persists you may file an issue in github. |

## 500 Internal Server Error

When analysing some sites, the WPSS Validation Tool may report a *500 Internal server error.* This may be due to a limitation on the Web server and not the WPSS Validation Tool. The problem may be due to the Web server not handling the “Range” setting in the HTTP GET operation. The WPSS Validation Tool sets a size limit on GET operations to avoid getting extremely large documents.

To avoid this error, you can change the WPSS Validation Tool configuration file setting to not include the “Range” setting in a HTTP GET operation.

Using a simple text file editor, such as WordPad:

1. Open the file c:\Program Files\wpss\_tool\conf\wpss\_tool.config.
2. Locate the following lines in the file:

#

# Max User Agent Size limits the size of files accepted in a GET

# request. A value of 0 means we can accept documents of any size.

# A value of 0 also removes the Range field from the HTTP header.

#

#User\_Agent\_Max\_Size 0

1. Remove the leading ‘#’ character from the #User\_Agent\_Max\_Size 0.

This functionality is **not** available through the user interface, only by directly editing the configuration file.