Pitney Bowes Software

Sharepoint File Scanner

User Guide

Spectrum™ Technology Platform

Data Discovery Module

Version 8

April 2013

Global IDs Product Suite® Sharepoint File Scanner Manual

Copyright © 2001 – 2013 Global IDs Inc.

All rights reserved. Printed in USA.

This software and documentation contain proprietary information of Global IDs Inc., they are provided under a license agreement containing restrictions on use and disclosure and is also protected by copyright law. Reverse engineering of the software is prohibited. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopy, recording, or otherwise) without prior consent of Global IDs Inc.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement as provided in DFARS 227.7202-1 (a) and 227.7702-3 (a) (1995), DFARS 252.227-7013(c)(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14(ALT III), as applicable.

This information in this document is subject to change without notice. If you find any problems in the documentation, please report them to us in writing. Global IDs Inc. does not warrant that this documentation is error free.

Global IDs Product Suite is the registered trademark of the Global IDs Inc. in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

Global IDs products contain GUI Components from JIDE™ Software, yFiles from yWorks™, JGo: GoDiagram™ from Northwoods Software Corporation, StelsEngine from J-Stels™ Software, EasyCharts™ from ObjectPlanet, Install4j from ej-technologies GmbH, Nmap™ - Free Security Scanner For Network Exploration & Security Audits(originally written by Gordon Lyon). Java™ is a registered trademark of Sun Microsystems and Apache Derby™ from Apache Software Foundation. All rights reserved.

DISCLAIMER: Global IDs Inc. provides this documentation ‘as is’ without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of non-infringement , merchantability, or use for a particular purpose. The information provided in this documentation may include many technical inaccuracies or typographical errors. Global IDs could make improvements and/or changes in the products described in this documentation at many time without notice.

Contents

List of Figures 4
About Sharepoint File Scanner 6
Scanning SharePoint Files 6
Searching SharePoint files for domains 8
Index 12

List of Figures

Figure 1: Scanning Sharepoint servers *7*
Figure 2: Hierarchy pane *7*
Figure 3: Viewing SharePoint files *8*
Figure 4: Search Domains - step 1 ..... *9*
Figure 5: Searching Domains - step 3 ..... *9*
Figure 6: Searching Domains - step 4 ..... *10*
Figure 7: Searching Domains - step 5 ..... *10*
Figure 8: Viewing searched domains *11*

Sharepoint File Scanner

What it includes:

- About Sharepoint File Scanner
- Scanning SharePoint File Server
- Finding domains in SharePoint files

About Sharepoint File Scanner

Sharepoint File Scanner application enables you to scan files stored on Microsoft SharePoint 2007 servers in your intranet. You can also find semantic domains from the scanned files.

Scanning SharePoint Files

This topic teaches you to scan files stored on SharePoint 2007 servers in your enterprise.

To do so:

- Right click the root node of the left hierarchy pane and select **Scan Sharepoint Server** option. **Scan Sharepoint Server** wizard appears.
- On Step 1 of the wizard:
  - Provide a connection name.
  - Provide the URL where SharePoint is installed.
  - Select an authentication type. You can choose either – **Windows** or **Basic**.
  - Provide username and password.
  - Provide your company's domain.
  - Click **Test Connection** link. This will check connectivity to the SharePoint server. If the connection is successful, a dialog box will appear with the message "**Connection Succeeded**".
  - Specify a directory or folder to scan or select appropriate checkbox to scan the entire system.
  - Specify the types of files to scan. The following file types can be scanned.
    - PDF (.pdf)
    - Flat File (.txt)
    - Rich Text Format (.rtf)
    - Excel (.xls and .xlsx)
    - Power Point (.ppt and .pptx)
    - Word (.doc and .docx)
    - Access 2007 Database File (.accdb)
    - Access (.mdb)
    - Hypertext Markup Language File (.htm and .html)
    - Comma Delimited (.csv and .txt)
    - Pipe Delimited (.txt)
    - Tab Delimited (.tsv, .tab and .txt)
    - Java Source Code File (.java)
    - C++ Source Code File (.cpp and .cc)
    - C/C++ Source Code File (.c)
    - VBScript File (.vb)
    - Java Server Page (.jsp)
    - Hypertext Preprocessor File (.php)

- Python Postscript File (.py)
- C# Source Code File (.cs)
- Text (.txt)

By default both these file types are scanned.

- Select appropriate checkbox to scan compressed files

Figure 1: Scanning Sharepoint servers

Click **Next**.

- On Step 2 of the wizard:
  - Select the appropriate radio button to run the task now or schedule it for a later time
  - Assign a priority level by choosing between five priority options **(Background, Low, Normal, High, Critical)**
  - Select the **Run tasks in parallel** check box to enable running more than one task at the same time and specify **Maximum tasks to run in parallel**
  - You can also add time restrictions by clicking on the **Add Time Restrictions** link

Click **Finish**.

Viewing scanned files

Once you have scanned the SharePoint server, it will be visible from the left hierarchy pane.

Figure 2: Hierarchy pane

You will be able to view the files in the SharePoint server from the **Scanned Files** tab if you select it from the left hierarchy pane. The **Scanned Files** tab is divided into 2 sections. The scanned files of the selected SharePoint server appear in a separate explorer pane. These files are categorized by the directories under which they reside. You can view this explorer pane from the **Scanned Files** tab on the right display pane if you select a system/SharePoint server from the left hierarchy pane. In the explorer pane, you can select a directory from the hierarchy on the left to view all scanned files along with their path, size and last modified date in a list view on the right. You can double click a file name from this list to view its content.

Please note: you will be able to view the content only if the file is residing in your local system.

Figure 3: Viewing SharePoint files

Searching SharePoint files for domains

Actions in grey need to be completed sequentially before the current action is performed. The current action is marked in red.

Pre-requisites

- Scan Sharepoint Server
- Search Domains

This topic teaches you to search domains in files stored on SharePoint servers.

To do so:

- Right click a system from the left hierarchy pane and select **Search Domains** option. **Search Domains** wizard appears.
- On Step 1 of the wizard:
  - Select drive where you want to search the files for domains. You may choose to search the entire drive by selecting the appropriate checkbox.
  - Specify the file types you want to search. By default both these file types are scanned.

  Click **Next**.

Figure 4: Search Domains - step 1

- Step 2 of the wizard appears if you de-select **Search Entire Drive** checkbox from Step 1:
  - Select the files in which you want to search domains. Click **Next**.
- On Step 3 of the wizard:
  - Select **domains** (from a list of pre-defined domains bundled with the software) that you would like to search. Click **Next**.

Figure 5: Searching Domains - step 3

- On Step 4 of the wizard:
  - Specify keywords for each selected domain. Click **Next**.

Figure 6: Searching Domains - step 4

- On Step 5 of the wizard specify the following parameters:
  - Minimum number of matches
  - Select **Apply Match Threshold %** checkbox to the change default matching threshold.
  - Check the appropriate checkbox to:
    - upload matches file(s) to server
    - rerun domain scan for previously scanned files
    - store offsets for matched values

  Click **Next**.

Figure 7: Searching Domains - step 5

- On Step 6 of the wizard:
  - Select the appropriate radio button to run the task now or schedule it for a later time.
  - Assign a priority level by choosing between five priority options **(Background, Low, Normal, High, Critical)**

- Select the **Run tasks in parallel** check box to enable running more than one task at the same time and specify **Maximum tasks to run in parallel**
- You can also add time restrictions by clicking on the **Add Time Restrictions** link

Click **Finish**.

Viewing searched domains

Domain Search Results tab at system level displays all files in which domains are found. The hierarchical table view also displays row count, column count, file owner's name, and the number of records found. **Metadata Discovered** check box is selected if the metadata has been discovered previously. You can select the **Is False Positive** checkbox if the domain found is not correct. If you select a file from this hierarchical table, you will be able to view its data. The values from the column that matched with a domain are highlighted in blue. *Please note: you will be able to view the data only if the selected file is located in your system*.

Figure 8: Viewing searched domains

Index

S

Scanning SharePoint Files
file types 6
Test Connection 6
Searching SharePoint files for domains
file types 8
Search Domains 8

V

Viewing scanned files
Scanned Files tab 7
Viewing searched domains
Domain Search Results 11
Is False Positive 11
Metadata Discovered 11