Pitney Bowes Software

File Scanner

User Guide

Spectrum™ Technology Platform

Data Discovery Module

Version 8

April 2013

Global IDs Product Suite® 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 File Scanner 6
Scanning Unstructured Files 6
Scanning PDF Files 7
Scanning Document Files 8
Scanning PPT Files 9
Scanning Plain Text Files 10
Scanning RTF Files 12
Index 13

List of Figures

Figure 1: Scanning unstructured files *6*
Figure 2: Scanning PDF files *7*
Figure 3: Viewing PDF files *8*
Figure 4: Scanning Word document files *9*
Figure 5: Scanning PowerPoint files *10*
Figure 6: Scanning Plain Text files *11*
Figure 7: Scanning RTF files *12*

File Scanner

What it includes:

- About File Scanner
- Scanning Unstructured Files
- Scanning PDF Files
- Scanning Document Files
- Scanning PPT Files
- Scanning Plain Text Files
- Scanning RTF Files

About File Scanner

Using File Scanner you can scan all unstructured files like PDF, Word Document, PPT, Plain Text and RTF files across your enterprise data landscape, load their data to an internal database and perform a number of profiling activities like profile columns, find patterns, profile domains.

Scanning Unstructured Files

This topic teaches you to scan unstructured files located in any system in your network.
To scan unstructured files:

- Right click on the desired system from the left hierarchy pane and select **Scan Unstructured Files** option. **Scan Unstructured Files** wizard appears.
- On Step 1 of the wizard:
  - Select a source
  - Specify a directory or select appropriate checkbox to scan the entire system.
  - Specify any Unstructured file type you want to scan. By default all types of unstructured files are scanned.
  - Select appropriate checkbox to:
    - scan compressed files
    - scan unstructured files in emails

Figure 1: Scanning unstructured files

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 unstructured files

You can select the system scanned for unstructured files from hierarchy pane and click on **Scanned Files** tab to view all scanned unstructured files. Expand the folders to view unstructured files.

Scanning PDF Files

This topic teaches you to scan PDF files located on any system in your network.
To scan PDF files:

- Right click on the desired system from the left hierarchy pane and select **Scan PDF Files** option. **Scan PDF Files** wizard appears.
- On Step 1 of the wizard:
  - Select a source.
  - Specify a directory or select appropriate checkbox to scan the entire system.
  - Select appropriate checkbox to:
    - scan compressed files
    - scan PDF files in emails

  Click **Next**.

Figure 2: Scanning PDF files

- 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 PDF files

You can select the system scanned for PDF files from hierarchy pane and click on **Scanned Files** tab to view all scanned PDF files along with other unstructured files. Select the folders to view files.

Figure 3: Viewing PDF files

Scanning Document Files

This topic teaches you to scan Word document files located on any system in your network.

To scan Word document files:

- Right click on the desired system from the left hierarchy pane and select **Scan Document Files** option. **Scan Document Files** wizard appears.
- On Step 1 of the wizard:
  - Select a system
  - Specify a directory or select appropriate checkbox to scan the entire system
  - Specify any document file format you want to scan. By default all types of document files are scanned.
  - Select appropriate checkbox to:
    - scan compressed files
    - scan document files in emails

Figure 4: Scanning Word document files

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 Word document files

You can select the system scanned for document files from hierarchy pane and click on **Scanned Files** tab to view all scanned document files along with other unstructured files. Select the folders to view files.

Scanning PPT Files

This topic teaches you to scan PPT files located on any system in your network.
To scan PPT files:

- Right click on the desired system from the left hierarchy pane and select **Scan PPT Files** option. **Scan PPT Files** wizard appears.
- On Step 1 of the wizard:
  - Select a source.

- Specify a directory or select appropriate checkbox to scan the entire system.
- Specify any PPT file format you want to scan. By default all types of PPT files are scanned
- Select appropriate checkbox to:
  - scan compressed files
  - scan PPT files in emails

Figure 5: Scanning PowerPoint files

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 PPT files

You can select the system scanned for PPT files from hierarchy pane and click on **Scanned Files** tab to view all scanned PPT files along with other unstructured files. Select the folders to view files.

Scanning Plain Text Files

This topic teaches you to scan Plain Text files located on any system in your network.
To scan plain text files:

- Right click on the desired system from the left hierarchy pane and select **Scan Plain Text Files** option. **Scan Plain Text Files** wizard appears.
- On Step 1 of the wizard:
  - Select a source.
  - Specify a directory or select appropriate checkbox to scan the entire system.
  - Select appropriate checkbox to:
    - scan compressed files
    - scan plain text files in emails

Figure 6: Scanning Plain Text files

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 plain text files

You can select the system scanned for plain text files from hierarchy pane and click on **Scanned Files** tab to view all scanned plain text files along with other unstructured files. You can select the folder to view all files in it or expand it to filter your scanned file list.

Scanning RTF Files

This topic teaches you to scan RTF files located on any system in your network.

To scan RTF files:

- Launch **Unstructured File Scanner**
- Right click on the desired system from the left hierarchy pane and select **Scan RTF Files.** Scan RTF Files wizard appears
- On Step 1 of the wizard:
  - Select a system
  - Specify a directory or select appropriate checkbox to scan the entire system
  - Specify any RTF file format you want to scan. By default all types of RTF files are scanned
  - Select appropriate checkbox to:
    - scan compressed files
    - scan RTF files in Emails

Figure 7: Scanning RTF files

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 RTF files

You can select the system scanned for RTF files from hierarchy pane and click on **Scanned Files** tab to view all scanned RTF files along with other unstructured files. Select the folders to view files.

Index

S

Scanning Document Files 8
Scanning PDF Files 7
Scanning Plain Text Files 10
Scanning PPT Files 9
Scanning RTF Files 12
Scanning Unstructured Files 6

V

Viewing scanned PDF files 8
Viewing scanned plain text files 11
Viewing scanned PPT files 10
Viewing scanned RTF files 12
Viewing scanned unstructured files 7
Viewing scanned Word document files 9