Pitney Bowes Software

PDF Scanner

User Guide

Spectrum™ Technology Platform

Data Discovery Module

Version 8

April 2013

Global IDs Product Suite® PDF 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 PDF Scanner 6
Scanning PDF Files 6
Working with scanned PDF files 7
Searching PDF for Domains 7
Searching Text in PDF files 9
Index 12

List of Figures

Figure 1: Scanning PDF files.....6
Figure 2: Viewing scanned PDF files.....7
Figure 3: Searching domains in PDF files - step 3.....8
Figure 4: Searching domains in PDF files - step 5.....9
Figure 5: Searching text in PDF files - step 1.....10
Figure 6: Searching text in PDF files - step 2.....10

PDF Scanner

What it includes:

- About PDF Scanner
- Scanning PDF Files
- Working with scanned PDF Files
    - Searching PDF files for Domains
    - Searching Text

About PDF Scanner

Using **PDF Scanner** you can scan PDF files across your enterprise data landscape and perform a number of activities like searching for domains and texts in them.

Scanning PDF Files

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

- Right click the desired system from the left hierarchy pane and select **Scan PDF File** option. **Scan PDF File** 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

Figure 1: Scanning PDF 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 PDF files

Scanned PDF files of the selected system appear in a separate explorer pane. You can view this explorer pane from the **Scanned Files** tab on the right display pane if a system from the left hierarchy pane is selected. In the explorer pane, you can select a drive or directory from the hierarchy on the left to view all scanned PDF files in a list view on the right.

Figure 2: Viewing scanned PDF files

Working with scanned PDF files

You can search for domains and text in the scanned PDF files.

Searching PDF 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 PDF File
- Search Domains

This topic teaches you to search domains in PDF files.

To search a PDF for domains:

- 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 the drive where you want to search the PDF file for domains. You may choose to search the entire drive by selecting the appropriate checkbox.

Click **Next**.

- On Step 2 of the wizard appears if you de-select Search Entire Drive checkbox:
  - Select the PDF 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.

Figure 3: Searching domains in PDF files - step 3

Click **Next**.

- On Step 4 of the wizard, you can:
  - Specify appropriate keywords for the selected domains to filter the search and reduce the number of false positives. This step is optional.

Click **Next**.

- 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.
  - To search a file for multiple domains select the appropriate checkbox and specify the number of domains to number of domains to search.
  - Check the appropriate checkbox to:
    - upload matched file(s) to server
    - rerun domain scan for previously scanned files

Figure 4: Searching domains in PDF files - step 5

Click **Next**.

- 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 PDF files in which domains are found. The hierarchical table view also displays file owners name and number of records found. You can select the **Is False Positive** checkbox if the domain found is not correct.

Searching Text in PDF files

This topic teaches you to search text in PDF files.

To search text:

- Select **Search** tab.
- Right click the **Search** root node from the left hierarchy pane and select **Search Text** option. **Search Text** wizard appears.
- On Step 1 of the wizard:
  - Select drives where the files in which you want to search for the text reside.

Figure 5: Searching text in PDF files - step 1

Click **Next**.

- On Step 2 of the wizard:
  - Click on **Basic** tab to
    - search an exact word or phrase
  - Click on **Advanced** tab to
    - search all the words, exact word or phrase, one or more of the given words, or to restrict viewing of files that have some specific words

Figure 6: Searching text in PDF files - step 2

Click **Next**.

- On Step 3 of the wizard:
  - Provide name and description

Click **Next**.

- On Step 4 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 files having the searched text

Search Results tab at the system level displays number of times a searched text is found in PDF files. You can add more field(s) in the pivot grid from the Field list for a detailed analysis. By default you can view the search name, location where the search is conducted, number of files searched and number of files in which the searched text is found.

Index

P

PDF files 6

S

Scanning PDF Files 6

Searching PDF for Domains 7

Searching Text in PDF files 9

V

Viewing files having the searched text 11

Viewing scanned PDF files 7

Viewing searched domains 9