AccuRead Automate

Version 2.5.4

Administrator's Guide

Contents

Change history

February 2022
- Updated information on training the application.

July 2020
- Updated information on the following:
  - Deployment readiness checklist
  - Configuring the application

April 2018
- Added information on the required minimum Embedded Solutions Framework version.

August 2017
- Updated information on the following:
  - Exporting configuration files
  - Configuring classifications

July 2016
- Added information on the following:
  - Configuring the OCR High Accuracy setting
  - Configuring the Location-Aware Classification setting
  - Printing a summary report after each job
  - Setting a confidence threshold for reviewing extracted values

January 2016
- Initial document release for multifunction products with a tablet-like touch-screen display.

Overview

Use the application to scan and classify documents, extract content, and then send them to a network or e-mail destination.

The application automates document processing and uses the recognition results from AccuRead^{TM} OCR to classify documents and extract content. For more information on AccuRead OCR, see the *AccuRead OCR Administrator's Guide*.

How it works

The application uses statistical methods, such as word frequency, to distinguish document classifications. Within a classification, the application uses the relative position of variable data to all fixed content (for example, headings, keys, or anchors) within the document. Since the extraction process analyzes all fixed content within a page, slight movement and scaling of data can be extracted without requiring an updated learnset.

Each time you train the application with a set of documents, it records statistical information from the documents in a learnset. Use various sample documents so that the engine can learn which content is fixed or variable.

This document provides instructions on how to configure, use, and troubleshoot the application.

Deployment readiness checklist

Make sure that:

☐ At least 1GB of RAM is installed.

☐ A 10GB printer hard disk is installed.

☐ At least Embedded Solutions Framework version 6 is installed.

☐ The AccuRead OCR license is valid.

☐ The date and time settings on the printer are updated.

☐ You have at least one of the following information to configure the application:
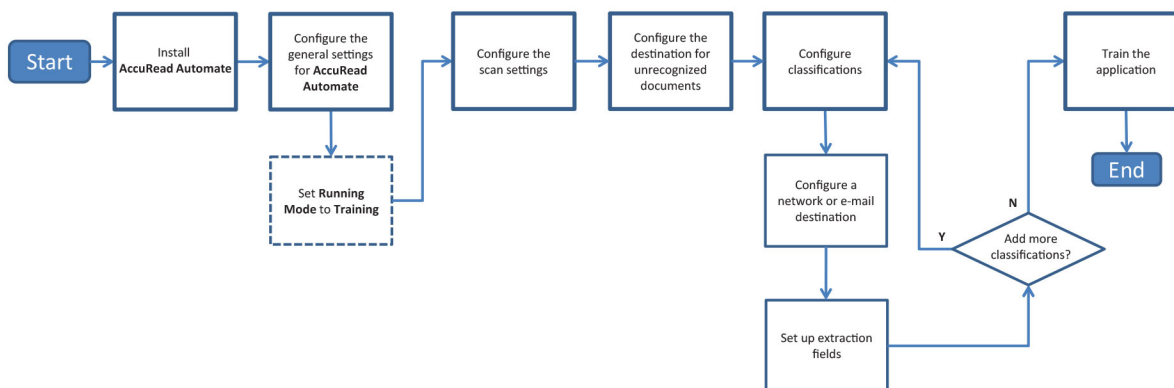 – Network folders

  _____

  – E-mail addresses

  _____

☐ You have at least five unique documents to train for each classification.

☐ You are using the latest version of Google Chrome^{TM} to access the configuration page for AccuRead Automate.

Configuring the application

Configuration workflow

You may need administrative rights to configure the application.

Accessing the Embedded Web Server

1 Obtain the printer IP address. Do either of the following:
  - Locate the IP address on the printer home screen.
  - View the IP address in the Network Overview section or in the TCP/IP section of the Network/Ports menu.

2 Open a web browser, and then type the printer IP address.

Configuring general settings for AccuRead Automate

These settings include changing the application mode, enabling manual review and job summary reports, and customizing the application icon.

1 From the Embedded Web Server, navigate to the configuration page for the application:

  Apps > **AccuRead Automate** > **Configure**

2 From the General Settings section, set Running Mode to either of the following:
  - **Training**—Use this mode when you train the application to classify documents and extract content.
  - **Production**—Use this mode when you scan and classify documents, extract content, and send them to a network or an e-mail destination.

    Note: To change Running Mode to Production, complete the training. For more information, see .

3 If necessary, enable the following options:
  - **Load Paper Reminder**—Prompt users to load paper.
  - **Blank Page Removal**—Remove blank pages automatically.

- **Turn On Manual Review**—Let users check the document classification and edit extracted values after scanning. You can also enable this option when the prompt for loading paper appears on the control panel.

  Note: If Manual Review is turned off, then you can review the extracted values in the XML audit file. For more information, see <u>"Reviewing extracted values in the audit file" on page 16</u>.

- **Turn On Summary Screen**—Show a summary report on the printer display after each job.
- **E-mail a Summary Report**—Send a summary report to one or more recipients after each job.
- **Print a Summary Report**—Print a summary report after each job. If an error occurs while a document is routed to its destination, then a summary report always prints regardless of the value of this setting.

4 If necessary, select one or more languages that you want the application to recognize on the document.

Note: Enabling several languages may reduce OCR accuracy. Make sure to select only the required languages.

Warning—Potential Damage: When the application is in Production mode, changing the language settings reverts the application to Training mode and removes all trained documents. To restore the learnset, train the documents again.

5 Specify the application name and icon that you want to appear on the home screen.

6 Click **Apply**.

Configuring scan settings

1 From the Embedded Web Server, navigate to the configuration page for the application:

Apps > **AccuRead Automate** > **Configure**

2 Click **Scan Settings**.

3 Configure the settings.

Notes:

- For more information on each setting, see the mouse-over help.
- Changing the Location-Aware Classification setting reverts the application to Training mode but retains all trained documents. To switch to Production mode, complete the training.

Warning—Potential Damage: When the application is in Production mode, changing the OCR High Accuracy setting reverts the application to Training mode and removes all trained documents. To restore the learnset, train the documents again.

4 Click **Apply**.

Configuring destinations for unrecognized documents

This feature helps you track the documents that the application cannot recognize. You can configure the application to route these documents to a specified destination.

Note: If these settings are not configured, then the AccuRead Automate icon does not appear on the home screen.

1 From the Embedded Web Server, navigate to the configuration page for the application:

Apps > **AccuRead Automate** > **Configure**

2 Click **Unrecognized Documents**.

3 Select and configure the destinations for unrecognized documents.

Network destination

a Select **Enable Network Folder Destination**.

b In the Network Folder Address field, type **\\\myServer\myFolder**, where **myServer** is the server host name or IP address and **myFolder** is the folder name.

Note: Make sure that the destination folder is shared and has a valid network address.

c Type the domain name and user credentials.

E-mail destination

a Select **Enable E-mail Destination**.

b Type the e-mail subject and e-mail address.

4 Select the output file format, and then type a name for the document.

5 If necessary, select **Audit XML Index File**.

Note: By default, this option is selected. For audit purposes, the application creates an XML file for each scanned document in a classification. To disable this feature, clear the check box.

6 Click **Apply**.

Configuring classifications

A classification is a grouping of documents that share common characteristics, such as word frequency.

Notes:

- Create at least two classifications before proceeding to training and production.
- For optimal performance, create no more than 100 classifications.

1 From the Embedded Web Server, access the configuration page for the application.

2 From the Classifications section, add or edit a classification.

3 Type a unique classification name.

Note: If necessary, specify the minimum confidence level. By default, the minimum confidence level is 90 percent.

4 Specify the number of pages for each document that you want to classify.

- **Single Page**—Each document has only one page, and classification is done only on that page.
- **Constant Page Length**—Each document contains the specified number of pages. Classification is done only on the first page, and the following pages are appended to the document.

- **Variable Page Length**—The number of pages vary for each document. In a page length, the first page is classified and the succeeding pages are appended. For example, if 3 is the page length, then the first page is classified and the succeeding two pages are appended. If the pages are not recognized, then those pages are appended to the previously classified page until the application classifies another page successfully.

5 Configure at least one classification destination.

Network destination

a Select **Network Folder Destination**.

b In the Network Folder Address field, type **\\myServer\myFolder**, where **myServer** is the server host name or IP address and **myFolder** is the folder name.

   Notes:

   - Make sure that the destination folder is shared and has a valid network address.
   - To set a dynamic folder name, in the address field, type **%** before and after the extraction field name. For example, **\\myServer\myFolder\%fieldname%**.

c Type the domain name and user credentials.

E-mail destination

a Select **E-mail Destination**.

b Type the e-mail subject and e-mail address.

   Note: To set a dynamic e-mail subject, in the address field, type **%** before and after the extraction field name. For example, **\\myServer\myFolder\%fieldname%**.

6 Select the output file format, and then type a name for the document.

Note: When you add an extraction field, the extraction field buttons appear following "Filename preview." Click these buttons to add a dynamic file name. For multiple extraction fields, type an underscore between each name. For example, **%fieldname1%_%fieldname2%**.

7 If necessary, select **Classification Summary E-mail**.

Note: If the subject field is blank, then the classification name is used.

8 If necessary, select **Audit XML Index File**.

Note: By default, this option is selected. For audit purposes, the application creates an XML file for each scanned document in a classification. To disable this feature, clear the check box.

9 Add up to four extraction fields, and then specify the appropriate field names, types, and confidence threshold value. For more information, see "Understanding extraction fields" on page 11.

Note: In each extraction field, you can enable and apply autocorrect rules. To set up more rules, see "Setting up Autocorrect rules" on page 12.

10 Click **Apply**.

Understanding extraction fields

Field type	Description	Training data format	Target data values
Amount	A number that can contain commas and periods as thousands and decimal separators. This field is often used for currencies. **Notes:** • There is no minimum number of leading or trailing digits. The application supports a single decimal separator followed by a maximum of two digits. • Do not include currency symbols or minus signs in the training data format. The currency symbols and minus signs in the target document are extracted automatically. • A target value of 0 is not supported.	*nnnn.nn* For example, **1234.56**	$1,234.56 -$1.234.56
Date	A number that represents a date value, depending on the configured country or region for your device.	Calendar interface	3-5-2015 3/5/2015 3/5/15 Mar 5, 2015 March 5, 2015 05 Mar 2015
Integer	A number up to 16 digits long. **Notes:** • The number must be a signed 64-bit integer (-9223372036854775807 to 9223372036854775807). Numbers smaller or larger than this range must use the String type. • Do not include minus signs in the training data format. The minus signs in the target document are extracted automatically. • A target value of 0 is not supported.	*nnnnnn* For example, **123456**	123456 -123456
Phrase	A series of alphanumeric characters or a group of words separated by spaces. **Note:** The field matching is case sensitive.	For example, **ABC Company**	ABC Company
String	A series of alphanumeric characters without spaces. **Note:** The field matching is not case sensitive.	For example, **ABC**	ABC

Setting up Autocorrect rules

This feature automatically corrects data for the common OCR errors. There are five default rules available that you can apply to each extraction field. You can also create up to 15 custom rules.

1 From the Embedded Web Server, navigate to the configuration page for the application:

Apps > **AccuRead Automate** > **Configure**

2 Click **Autocorrect Rules**.

3 Add or edit an Autocorrect rule.

Note: The default rules cannot be edited.

4 Type a unique name for the rule, and then select the correction options.

Note: Advanced users can use regular expressions (regex) to create more complex rules. For more information, see .

5 Click **Apply**.

Understanding regular expressions

Note: For the effective use of this function, a working knowledge of Perl 5 programming is helpful. For more information, see the online references for Perl 5 regex.

A regular expression is a pattern used for substituting characters. Use the Perl 5 regex format in defining Autocorrect rules.

For example, **s/pattern/replacement/modifier**

Where **pattern** is the search pattern, **replacement** is the replacement string, and **modifier** is the substitution option.

Use the following substitution options:

- **g**—Substitutes all occurrences of the search pattern with the replacement.
- **i**—Ignores the case of the character.
- **m**—Recognizes the string as multiple lines.
- **o**—Evaluates the pattern only once. When you use variable interpolation, this option uses only the initial values even if the values are changed.
- **s**—Recognizes the string as a single line.
- **x**—Enables extended regular expression that incorporates white space and comments to the pattern.

For example, **s/[^abc]/d/g**

This expression uses the modifier **g**, which replaces all characters other than **a**, **b**, or **c** with **d**. For example, **abce** results to **abcd**.

Exporting or importing a configuration file

Note: The configuration files created in earlier versions of AccuRead Automate are not supported in AccuRead Automate version 2.0 or later. For more information, see the *Readme* file.

1 From the Embedded Web Server, navigate to the configuration page for the application:

Apps > **AccuRead Automate** > **Configure**

2 Click **Save/Load Configuration**.

3 Do either of the following:

Export a configuration file

a In the Save tab, specify the network location where you want to save the settings and training data from the application.

b Specify the domain name and user credentials.

c Type a unique configuration package file name.

d Click **Export Configuration File**.

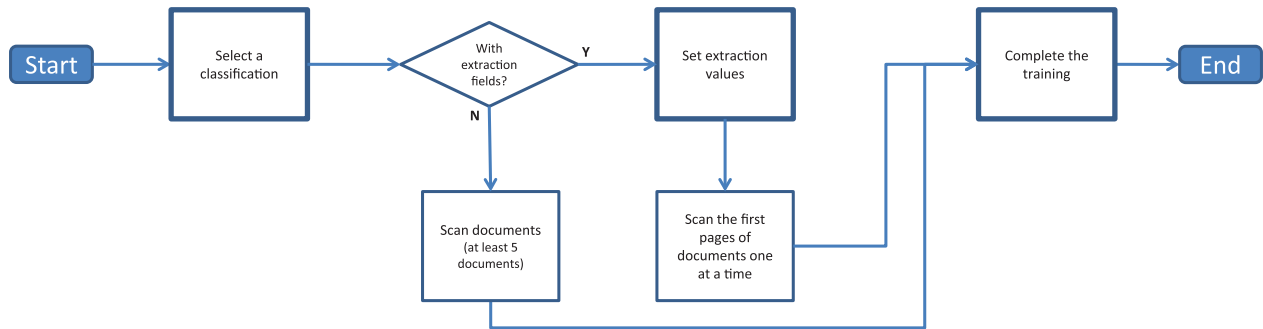## Import a configuration file

Note: When you import a configuration, the existing classifications are removed, and any ongoing activities, such as training and classification processes, are terminated.

a In the Load tab, specify the network location where you want to retrieve the configuration files.

b Specify the domain name and user credentials.

c Type the deployment package file name.

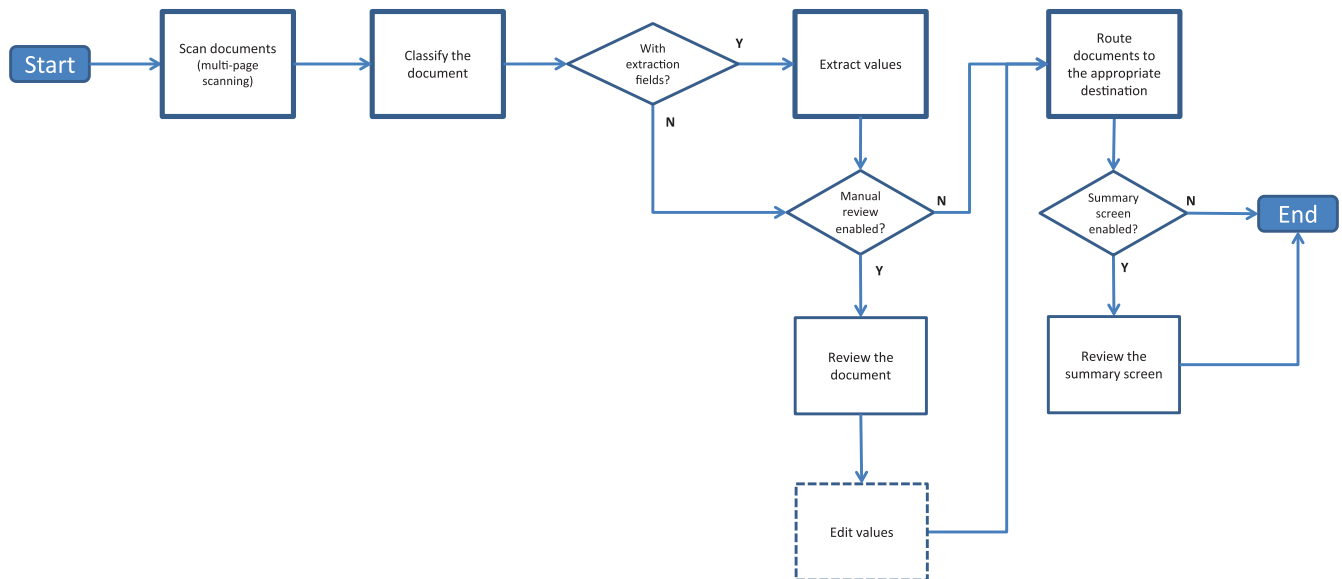**d** Click **Import Configuration File**.

4 Click **Apply**.

Using the application 14

Using the application

Training mode workflow

Start → Select a classification → With extraction fields? —Y→ Set extraction values → Complete the training → End

With extraction fields? —N→ Scan documents (at least 5 documents)

Set extraction values → Scan the first pages of documents one at a time

Production mode workflow

Start → Scan documents (multi-page scanning) → Classify the document → With extraction fields? —Y→ Extract values → Route documents to the appropriate destination

With extraction fields? —N→ Manual review enabled?

Extract values → Manual review enabled? —N→

Manual review enabled? —Y→ Review the document → Edit values

Route documents to the appropriate destination → Summary screen enabled? —N→ End

Summary screen enabled? —Y→ Review the summary screen → End

Training the application

Before you begin, make sure that Running Mode is set to **Training**. For more information, see "Configuring general settings for AccuRead Automate" on page 7.

Classification only

Note: Make sure that two-sided scanning is turned off.

1 From the home screen, touch the application icon.

2 Touch **Train** beside a classification.

3 Load the first page of a document into the automatic document feeder (ADF) tray or on the scanner glass, and then scan it.

> **Note:** When training the application with multiple-page documents, scan only the first page of each document, and load at least five documents on the ADF tray. For more information on scanning multiple pages, see "Frequently asked questions" on page 21.

4 Follow the instructions on the display.

Classification with extraction

1 From the home screen, touch the application icon.

2 Touch **Train** beside a classification.

3 Set the extraction values, and then touch **Next**.

4 Load the first page of a document into the ADF tray or on the scanner glass, and then scan it.

> **Notes:**
>
> - When training the application with multiple-page documents, scan the first page of each document one at a time. For more information on scanning multiple pages, see "Frequently asked questions" on page 21.
> - When extracting a page that contains two or more identical fields, find the most legible field and cover the other fields before scanning.

5 Follow the instructions on the display.

Completing the training

When you complete the training, the application exits Training mode and changes Running Mode to Production automatically. To train the application again, change Running Mode to Training manually from the Embedded Web Server. For more information, see "Configuring general settings for AccuRead Automate" on page 7.

1 From the home screen, touch the application icon.

2 Touch **Complete Training**.

3 Follow the instructions on the display.

Notes:

- After completing the training, you can modify all other fields except the classification name and the extraction field.
- To modify the value in the classification name and the extraction field, you must create a classification and reset the values.

Classifying documents

Before you begin, make sure that Running Mode is set to **Production**. For more information, see "Configuring general settings for AccuRead Automate" on page 7.

1 From the home screen, touch the application icon.

2 Load a document into the ADF tray or on the scanner glass, and then scan it.

3 If Manual Review is enabled, then check if the classification and extracted values are correct, modify them
if necessary, and then do either of the following:

- Touch **Approve**.

- Touch ⋮ > **Approve All and Exit**.

Note: Touching the home icon during the manual review cancels the job.

4 Follow the instructions on the display.

Reviewing extracted values in the audit file

Note: If Manual Review is turned off, then you can review the extracted values in the XML audit file.

This feature helps you check whether the extraction fields are found with high confidence. If the confidence
value of an extraction field is below the threshold, then the **review** prefix is appended to the document file
name. The default value for Review Confidence Threshold is 75 percent. You can change this value in the
extraction field settings for the classification. Autocorrect rules are not applied to documents that are marked
for review.

This feature applies only when **Audit XML Index File** is enabled in the classification settings. For more
information, see "Configuring classifications" on page 9.

Troubleshooting

Application error

Try one or more of the following:

Check the diagnostic log

1 From the Embedded Web Server, navigate to the configuration page for the application:

   Apps > **AccuRead Automate** > **Configure**

2 Click **Logging**.

3 Select **Enable Logging**, and then select one or more logging options.

4 Click **Apply**.

5 Open a Web browser, type the printer IP address followed by **/se**, and then click **Embedded Solutions**.

6 Analyze the log, and then resolve the problem.

Adjust the scan settings

1 From the Embedded Web Server, navigate to the configuration page for the application:

   Apps > **AccuRead Automate** > **Configure**

2 From the scanning options, select a lower scan resolution or turn off color scanning.

3 Click **Apply**.

Contact your Lexmark representative

Application icon does not appear on the home screen

Try one or more of the following:

Make sure to configure the destination settings for unrecognized documents

For more information, see <u>"Configuring destinations for unrecognized documents" on page 8</u>.

Make sure to create at least two classifications

For more information, see <u>"Configuring classifications" on page 9</u>.

Error while processing a configuration

Try one or more of the following:

Make sure that the AccuRead OCR license is valid

Make sure that a printer hard disk is installed

For more information, see the printer *User's Guide*.

Make sure that the available printer hard disk space is at least 10GB

For more information, contact your system administrator.

Restart AccuRead Automate

1 Open a web browser, and then type the printer IP address.

   Note: Locate the IP address on the printer home screen.

2 Navigate to **Apps** > **AccuRead Automate**, and then click **Stop**.

3 Click **Start** to restore the application.

Contact your Lexmark representative

Insufficient memory

Try one or more of the following:

Make sure that the available printer hard disk space is at least 10GB

For more information, contact your system administrator.

Make sure that the printer memory is at least 1GB

For more information, contact your system administrator.

Contact your Lexmark representative

Connection error while testing network location

Try one or more of the following:

Make sure that the destination folder is shared and has a valid network address

Make sure that the user credentials configured for the network destination are correct

Make sure that the server and the printer are connected to the same network

At the server command prompt, type **ping IP_address**, where **IP_address** is the IP address of the printer. If the ping is unsuccessful, then make sure that all appropriate network cables are securely connected and that the printer network settings are correctly configured.

Contact your system administrator

Document is misrouted to the folder for unrecognized documents

Try one or more of the following:

Make sure that the network path, domain, and credentials for the classification destination are correct

For more information, contact your system administrator.

Make sure that the document is clean and legible

Train the application to classify the document

For more information, see .

Cannot export a configuration file

Try one or more of the following:

Make sure that there is available space on the hard disk

When you export a configuration file, the application creates a deployment package that requires available space on the hard disk. If your hard disk is full, then remove unused files.

Contact your Lexmark representative

Loading configuration failed

When the error occurs, the application becomes unusable. Try one or more of the following:

Reinstall AccuRead Automate

Note: Do not remove the AccuRead OCR license.

Contact your Lexmark representative

License error

Restart AccuRead Automate

The application comes with a bundled feature license. Restart the application to restore the license.

Contact your Lexmark representative

Frequently asked questions

Can I scan multiple-page documents?

Yes, you can scan documents containing more than one page, depending on the document length specified in the classification settings. For more information, see "Configuring classifications" on page 9.

Can I edit the extraction field names?

Yes, but only before you start the training.

What happens when I scan to a paper size that is not supported?

Document sizes that are not supported are scanned to legal (8.5 x 14 inches), which may cause cropping of images. The following paper size options are not supported:

- A3
- Auto Size Sense
- Book Original
- Custom Scan Size [x]
- Ledger
- Mixed Sizes
- Tabloid
- Universal

What should I do if the document structure changes?

If the structure of your document changes significantly and you are extracting content, then create another classification. If you are only classifying the document, then no action is required.

If there are minor changes, such as the header or footer of the document, then no additional training is required. If the classification or extraction is not successful, then train the application with several documents. This process improves the accuracy of the classification or extraction process.

What happens if the document does not match a classification?

If a document does not belong to any classification, then it is routed to the specified destination for unrecognized documents.

A document is likely to be unrecognized if it is new, has been changed significantly, or is not legible.

Why do I have an XML file in the classification folder or e-mail?

If **Audit XML Index file** is enabled in the classification settings, then an XML file is automatically created with every scan job. This file contains information about the scanned document, including extracted fields that helps create indices in the Enterprise Content Management systems or core business systems.

Notices

Edition notice

February 2022

The following paragraph does not apply to any country where such provisions are inconsistent with local law: LEXMARK INTERNATIONAL, INC., PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

This publication could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in later editions. Improvements or changes in the products or the programs described may be made at any time.

References in this publication to products, programs, or services do not imply that the manufacturer intends to make these available in all countries in which it operates. Any reference to a product, program, or service is not intended to state or imply that only that product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any existing intellectual property right may be used instead. Evaluation and verification of operation in conjunction with other products, programs, or services, except those expressly designated by the manufacturer, are the user's responsibility.

For Lexmark technical support, go to **http://support.lexmark.com**.

For information on Lexmark's privacy policy governing the use of this product, go to **www.lexmark.com/privacy**.

For information on supplies and downloads, go to **www.lexmark.com**.

© **2016 Lexmark International, Inc.**

All rights reserved.

GOVERNMENT END USERS

The Software Program and any related documentation are "Commercial Items," as that term is defined in 48 C.F.R. 2.101, "Computer Software" and "Commercial Computer Software Documentation," as such terms are used in 48 C.F.R. 12.212 or 48 C.F.R. 227.7202, as applicable. Consistent with 48 C.F.R. 12.212 or 48 C.F.R. 227.7202-1 through 227.7207-4, as applicable, the Commercial Computer Software and Commercial Software Documentation are licensed to the U.S. Government end users (a) only as Commercial Items and (b) with only those rights as are granted to all other end users pursuant to the terms and conditions herein.

Trademarks

Lexmark, the Lexmark logo, and AccuRead are trademarks or registered trademarks of Lexmark International, Inc. in the United States and/or other countries.

Google Chrome is a trademark of Google LLC.

All other trademarks are the property of their respective owners.

Glossary

classification	A grouping of documents that share common characteristics, such as page structure and frequency of certain words.
classify	Categorizing and organizing scanned documents.
dynamic file name	A placeholder for a file name containing plain text and macros (for example, %macro%), which is substituted with the field values extracted from a document.
extraction	The capturing of values from fields in scanned documents.
extraction field	Predefined data that will be extracted from a document.
learnset	A collection of information that uniquely identifies a document.
manual review	The checking and editing of values for extraction fields and deleting documents from the printer control panel.
production	The process in which the application classifies scanned documents, captures values from the extraction fields, and sends them to a specified destination.
regular expression	A pattern used to perform find-and-replace functions.
route	Sending scanned documents to a specified destination.
running mode	The mode in which the application is running, whether Training or Production.
training	Teaching the application to classify scanned documents and set extraction fields.
unrecognized documents	Documents that are not classified during a scan.

Index