

AccuRead Automate

Administrator's Guide

July 2017

www.lexmark.com

Contents 2

Contents

Overview	3
Configuring the application	4
Accessing the configuration page for the application	
Configuring AccuRead Messenger	
Configuring general settings for AccuRead Automate	
Configuring scan settings	
Configuring destinations for unrecognized documents	
Configuring classifications	7
Understanding extraction fields	9
Setting up Autocorrect rules	10
Exporting or importing a configuration file	11
Jsing the application	12
Training the application	12
Classifying documents	13
Reviewing extracted values in the audit file	13
Troubleshooting	15
Application error	15
Application icon does not appear on the home screen	15
An internal error has occurred while processing a configuration	16
Insufficient memory	17
Document does not match a classification	17
Cannot export a configuration file	17
Loading configuration failed	17
Frequently asked questions	18
Notices	20
Glossary	
ndex	
ngex	

Overview 3

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 $^{\text{TM}}$ 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. Because 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.

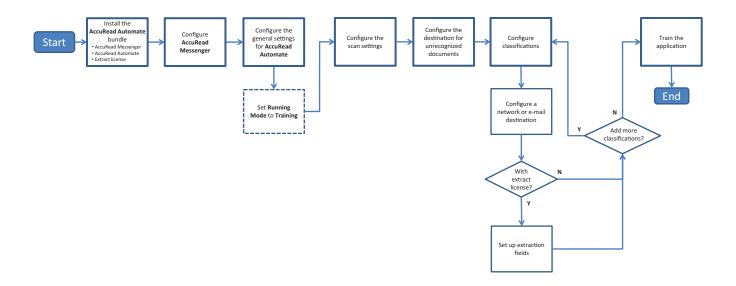
System requirements

- 7- or 10-inch multifunction products (MFPs)
- 1GB printer memory or DIMM
- 10GB printer hard disk
- AccuRead Automate bundle components and licenses:
 - AccuRead OCR
 - AccuRead Messenger
 - AccuRead OCR License
 - AccuRead Classify license
 - AccuRead Extract license
- Network folders or e-mail addresses
- Google ChromeTM browser

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

Configuring the application

Configuration workflow



Accessing the configuration page for the application

Note: Make sure to disable the pop-up blockers on your browser.

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

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

- 2 Click Settings or Configuration.
- **3** Depending on your printer model, do one of the following:
 - Click Apps > Apps Management.
 - Click Device Solutions > Solutions (eSF).
 - Click Embedded Solutions.
- **4** Select the application from the list, and then click **Configure**.

Configuring AccuRead Messenger

AccuRead Messenger lets you track the number of documents processed by AccuRead Automate, and notifies you of the remaining document volume before the license expires. The number of documents may vary depending on the license purchased. For more information, contact your solution provider.

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

Note: From the Usage section, you can view the license installation and renewal date. You can also check the number of documents successfully classified and routed, and the remaining number of documents you can use until the renewal date.

2 Enable the notifications, and then set the usage threshold at which you want to be notified of the remaining document volume.

Notes:

- By default, notifications are sent when you have used 90% and 100% of the document volume allowed by your license.
- You can specify up to two custom thresholds.
- For more information on each setting, see the mouse-over help.
- **3** From the E-mail Notifications section, specify the e-mail information.

Note: Use a comma to separate multiple e-mail addresses.

4 Click Apply.

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, access the configuration page for AccuRead Automate.
- 2 From the General Settings section, change the 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.
- **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.

Note: You can also enable this option when the prompt for loading paper appears on the control panel.

- Turn On Summary Screen—Show a summary on the printer display after each job.
- Job Summary E-mail—Send a summary to one or more recipients after each job.
- 4 Select the language 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.

- 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, access the configuration page for AccuRead Automate.
- **2** From the Scan Settings section, configure the settings.

Note: For more information on each setting, see the mouse-over help.

3 Click Apply.

Configuring destinations for unrecognized documents

Note: Configure these settings for the AccuRead Automate icon to appear on the home screen.

- 1 From the Embedded Web Server, access the configuration page for AccuRead Automate.
- 2 From the Unrecognized Documents section, configure the destinations for unrecognized documents.

Network destination

- a Select Share for Unrecognized Documents.
- **b** In the Network Share 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 E-mail for Unrecognized Documents.
- **b** Type the e-mail subject and e-mail address.
- **3** Type a name for the document, and then select the output file format.
- 4 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.

5 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.

Notes:

- This feature is available only if an extract license is installed. For more information, see <u>"Understanding extraction fields" on page 9.</u>
- In each extraction field, you can enable and apply autocorrect rules. To set up more rules, see <u>"Setting up Autocorrect rules" on page 10</u>.

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.	NNNN.NN For example, 1234.56	\$1,234.56 \$1.234,56
	Notes:		
	 There is no minimum number of leading or trailing digits. 		
	 Do not include currency symbols in the "Training data format." 		
	 Currency symbols in the target document will be extracted. 		
	 A target value of "0" is not supported. 		
Date	A number that represents a date value,	MMDDYYYY or DDMMYYYY	D.M.YY
	depending on the configured country or region for your device.	For example, 03052015	M.D.YY
	region for your device.		YY.M.D
			Sample values:
			3-5-2015
			3/5/2015 3/5/15
			DD MMM YYYY
			MMM DD, YYYY
			Sample values:
			Mar 5, 2015
			March 5, 2015
			05 Mar 2015
			Notes:
			• D can also be DD .
			• M can also be MM or MMM, where MMM is a month name or abbreviation.
			• YY can also be YYYY.
			A hyphen (-) or forward slash (/) can be used as a separator.

Field type	Description	Training data format	Target data values
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 should use the String type. A target value of "0" is not	NNNNNN For example, 123456	123456
	supported.		
Phrase	A series of alphanumeric characters or group of words separated by spaces. Note: The field matching is case-sensitive.	For example, Lexmark International, Inc.	Lexmark International, Inc.
String	A series of alphanumeric characters without spaces. Note: The field matching is not case-sensitive.	For example, Lexmark	Lexmark

Setting up Autocorrect rules

This feature automatically corrects data for the most common OCR errors. There are 5 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, access the configuration page for AccuRead Automate.
- **2** From the Autocorrect Rules section, add or edit an Autocorrect rule.

Note: The default rules cannot be edited.

3 Type a unique name for the rule, and then select the correction options that you want.

Note: Advanced users can use *regular expressions* (regex) to create more complex rules. For more information, see <u>"Understanding regular expressions" on page 10</u>.

4 Click Apply.

Understanding regular expressions

Note: A working knowledge of Perl 5 programming is helpful for the effective use of this function. 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 whitespace and comments to the pattern.

Sample regular expression

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

- 1 From the Embedded Web Server, access the configuration page for AccuRead Automate.
- **2** From the Save/Load Configuration section, 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** Type a new configuration package file name.
- c Click Export Configuration File.

Import a configuration file

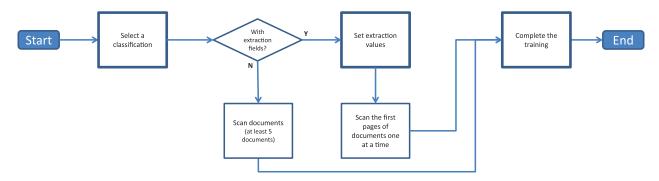
Note: When you import a configuration, the existing classifications are removed, and any occurring activity such as training and classification processes are terminated.

- a In the Load tab, specify the network location from where you want to retrieve the configuration file.
- **b** Type the deployment package file name.
- c Click Import Configuration File.
- 3 Click Apply.

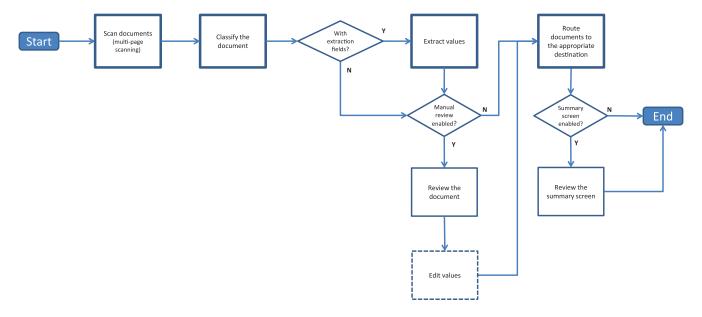
Using the application 12

Using the application

Training mode workflow



Production mode workflow



Training the application

Before you begin, make sure that Running Mode is set to **Training**. For more information, see <u>"Configuring general settings for AccuRead Automate" on page 5</u>.

Using classification only

- **1** From the home screen, touch the application icon.
- 2 Select a classification, and then touch Next.
- **3** Load the first page of an original document into the automatic document feeder (ADF) tray or on the scanner glass, and then scan it.

Using the application 13

Notes:

• If you are using the ADF tray, load at least 5 documents. For more information on scanning multiple pages, see <u>"Frequently asked questions" on page 18.</u>

- If two-sided scanning is turned on, load the first page faceup on the ADF tray, depending on your printer.
- **4** Follow the instructions on the display.

Using classification with extraction

- **1** From the home screen, touch the application icon.
- 2 Select a classification, and then touch Next.
- **3** Set the extraction values, and then touch **Next**.
- 4 Load the first page of an original document into the ADF tray or on the scanner glass, and then scan it.

Notes:

- Train the application to classify and extract the first pages of the documents one at a time.
- When extracting a page that contains two or more identical fields, use redaction by finding the most legible field and covering the other ones before scanning.
- For more information on scanning multiple pages, see "Frequently asked questions" on page 18.
- **5** Follow the instructions on the display.

Classifying documents

Before you begin, make sure that Running Mode is set to **Production**. For more information, see <u>"Configuring general settings for AccuRead Automate" on page 5</u>.

- **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.

Note: For more information on the paper orientation, see the printer *User's Guide*.

3 If Manual Review is enabled, then check if the document classification and extracted values are correct. If necessary, modify the extracted values.

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

To enable this feature, select **Audit XML Index File** from the classification settings. For more information, see <u>"Configuring classifications" on page 7.</u>

Note: This feature applies only when manual review is turned off.

Using the application 14

Documents are marked for review when the confidence value of an extraction field is below the review threshold. The default review threshold is 75 percent. The following conditions apply to documents that are marked for review:

- The **review**_ prefix is added to the document file name.
- The **review** text is added to the subject line on the e-mail destination.
- The number of documents marked for review are listed in the Classification Summary e-mail and Job Summary e-mail.
- In the document audit file that is created, the extraction fields that fall below the review threshold are indicated by **extractedConfidence**, **configuredConfidence**, and **reviewRequired** XML tags.
- Autocorrect rules are not applied to the documents.

Troubleshooting 15

Troubleshooting

Application error

Try one or more of the following:

Check the diagnostic log

- 1 From the Embedded Web Server, access the configuration page for AccuRead Automate.
- 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, access the configuration page for AccuRead Automate.
- 2 In the scan settings, 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 "Configuring destinations for unrecognized documents" on page 6.

Make sure to create at least two classifications

For more information, see "Configuring classifications" on page 7.

Troubleshooting 16

An internal error has occurred while processing a configuration

Try one or more of the following:

Make sure that AccuRead OCR and AccuRead Messenger are installed

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

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

- 2 Click Settings or Configuration.
- **3** Depending on your printer model, do one of the following:
 - Click Apps > Apps Management.
 - Click Device Solutions > Solutions (eSF).
 - Click Embedded Solutions.
- 4 Click Install a New App.
- 5 From the App Install File section, browse to the flash file (.fls), and then click Start.

Make sure that the application is licensed and that the license is up-to-date

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

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

- 2 Click Settings or Configuration.
- **3** Depending on your printer model, do one of the following:
 - Click Apps > Apps Management.
 - Click Device Solutions > Solutions (eSF).
 - Click Embedded Solutions.
- 4 Click the license status of the appropriate application, and then install or update the license.

Make sure that the printer hard disk is installed

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

Make sure that the installed printer hard disk has a minimum of 10GB available space

For more information, contact your system administrator.

Contact your system administrator

If you still cannot resolve the problem, then contact your system administrator.

Troubleshooting 17

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

Document does not match a classification

Make sure that the document is clean and legible

Train the application to classify the document

For more information, see "Training the application" on page 12.

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 licenses and other applications included in the bundle.

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 <u>"Configuring classifications" on page 7</u>.

Can I edit the extraction field names?

Yes, but only before you start the training.

What paper size options are not supported?

Document sizes that are not supported are scanned to legal $(8.5 \times 14 \text{ inches})$, 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

Why are the usage values not updated in the AccuRead Messenger application?

The information under the Usage section is updated only when the application classifies and routes the document successfully.

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 20

Notices

Edition notice

July 2017

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, visit http://support.lexmark.com.

For information on supplies and downloads, visit www.lexmark.com.

© 2014 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. or its subsidiaries in the United States and/or other countries.

Google Chrome is a trademark of Google Inc.

All other trademarks are the property of their respective owners.

Glossary 21

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 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 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 of scanned documents to a specified destination.

running modeThe 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 22

Index

Α

accessing the configuration page 4
AccuRead Messenger configuring 5 updating usage values 18 application configuration 4 application error 15 application icon does not appear 15 application mode setting 5 audit file reviewing extracted values in 13
Autocorrect rules 10

В

blank page removal enabling 5

C

cannot classify document 17 cannot export a configuration file 17 character recognition errors automatic correction 10 classification only usina 12 classification with extraction using 12 classifying and extracting documents 12 classifying documents 12, 13 configuration page for the application accessing 4 configuring AccuRead Messenger 5 unrecognized documents 6 configuring scan settings 6

D

destinations
unrecognized documents 6
document does not match a
classification 17

documents classifying 12, 13 classifying and extracting 12

E

errors in character recognition automatic correction 10 extraction field types 9 extraction fields editing 18 understanding 9

F

frequently asked questions 18

information on extraction field types 9 insufficient memory 17 internal error 16

J

job summary e-mail enabling 5

L

loading configuration failed 17

M

manual review
enabling 5
missing application icon 15, 18

R

regular expressions
understanding 10
reviewing extracted values in the audit file 13
running mode
Production 5
setting 5
Training 5

S

sample regular expressions 10

scan settings
configuring 6
scanning
multiple-page documents 18
setting up Autocorrect rules 10
summary screen
enabling 5
supported copy settings 18

T

training the application 12
troubleshooting
application error 15
application icon does not
appear 15
cannot classify document 17
cannot export a configuration
file 17
document does not match a
classification 17
insufficient memory 17
internal error during
configuration 16
loading configuration failed 17

U

unrecognized documents 18 configuring 6

X

XML file in the classification folder 18