



Automated Indexing

COPYRIGHT

Information in this document is subject to change without notice. The OnBase® Information Management System software described in this document is furnished only under a separate license agreement and may be used or copied only according to the terms of such agreement. It is against the law to copy the software except as specifically allowed in the license agreement, or without the expressed written consent of Hyland Software, Inc. If Hyland Software, Inc. and you have entered into a nondisclosure agreement, then this document or accompanying materials provided by Hyland Software, Inc. contains certain information which is confidential information of Hyland Software, Inc. and which may be used or copied only according to the terms of such nondisclosure agreement. All data, names, and formats used in this document's examples are fictitious unless noted otherwise. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Hyland Software, Inc.

©2011 Hyland Software, Inc. All rights reserved.

Depending on the modules licensed, The OnBase® Information Management System software may contain portions of: Imaging technology, Copyright © Snowbound Software Corporation; CD-R technology, Copyright © Sonic Solutions; CD-R technology, Copyright © Rimage Corporation; OCR technology, Copyright © Nuance Corporation; Mail interface technology, Copyright © Intuitive Data Solutions; Electronic signature technology, Copyright © Silanis Technology, Inc.; Full text search technology, Office core assembly, ASP.NET extensions, application blocks, smart client architecture, Object Builder, and WPF controls, Copyright © Microsoft Corporation; Full Text Indexing technology, Copyright © Verity, Inc.; SYBASE Adaptive Server Anywhere Desktop Runtime, Copyright © SYBASE, Inc., portions Copyright © Rational Systems, Inc.; ISIS technology, Copyright © EMC Corporation; JLex technology, Copyright © 1996-2003 by Elliot Joel Berk and C. Scott Ananian; A2iA CheckReader, Copyright © A2iA; Terminal emulation technology, Copyright © Attachmate; User interface controls, Copyright © Infragistics; Terminal emulation technology, Copyright © NetManage; CAD document technology, Copyright © Open Text Corporation; ISIS scanning interface, Copyright © Pegasus Imaging Corporation; CD/DVD burner technology, Copyright © Prassi Software Incorporated; Code obfuscation technology, Copyright © PreEmptive Solutions; Icon library, Copyright © Professional Icons; OSA dlls, Copyright © Sharp Electronics Corp.; JAVA components, Copyright © Sun Microsystems; Signature pad technology, Copyright © Topaz Systems Incorporated; and User interface tools, Copyright © Xceed Software, Incorporated.

Portions of the OnBase® software modules may be covered by one or more of the following U.S. Patents: 7,644,091 and 7,765,271. Portions contained within OnBase® are licensed by U.S. Patent Nos. 6,094,505; 5,768,416; 5,625,465 and 5,258,855.

Hyland Software® and OnBase® are registered trademarks of Hyland Software, Inc. Application Enabler™ is an unregistered trademark of Hyland Software, Inc. EMC Centera® is a registered trademark of EMC Corporation. All other trademarks, service marks, trade names and products of other companies are the property of their respective owners.

Attribute	Detail
Document Name	Automated Indexing
Department/Group	Documentation
Revision Number	11.0.0
Part Number	AutoIndexRM-11.0.0-SP 1 -OB

Exposure

OVERVIEW	1
Automated Indexing.....	2
Point and Shoot Indexing.....	3
APPLICATIONS	3
SYSTEM PREREQUISITES.....	4
LICENSING.....	4
Optional Licensing	5

Usage

AUTOMATED INDEXING	7
Performing Ad-Hoc Automated Indexing.....	7
<i>From the Document Search Results List</i>	<i>9</i>
<i>From an Open Document</i>	<i>11</i>
<i>From the Document Imaging Window</i>	<i>14</i>
<i>From the Import Document Dialog Box</i>	<i>17</i>
<i>From a Folder</i>	<i>19</i>
Performing Batch Automated Indexing.....	22
<i>On a Newly-Imported Batch</i>	<i>24</i>
Index Verification	26
<i>On Documents Queued for Ad-Hoc Automated Indexing</i>	<i>29</i>
POINT AND SHOOT INDEXING.....	32
Performing Point and Shoot Indexing.....	33
A Note About Single and Double-Byte Characters	38
SYSTEM INTERACTION	38
AutoFill Keyword Sets.....	38
<i>Ad Hoc Automated Indexing</i>	<i>39</i>
<i>Batch Automated Indexing</i>	<i>39</i>
<i>Point and Shoot Indexing</i>	<i>39</i>
<i>Creating AutoFill Keyword Sets via Automated Indexing</i>	<i>39</i>
Automated Redaction	40
Document Imaging.....	40
<i>Orientation Detection & Automatic Page Rotation</i>	<i>40</i>
<i>Appending Documents to Existing Documents</i>	<i>41</i>
EDM Services.....	41
Intelligent Character Recognition (ICR)	42
<i>Limitations of Intelligent Character Recognition</i>	<i>42</i>
<i>Improving the Accuracy of Intelligent Character Recognition</i>	<i>42</i>
Web Client	43

Workflow	44
Virtual Print Driver	44

Configuration

CONFIGURING DOCUMENT TYPES	47
Creating an OCR Format	47
<i>Preprocessor Options</i>	49
<i>Recognizer Options</i>	51
<i>Languages</i>	52
<i>Filter Options</i>	53
<i>Spelling</i>	53
<i>Codepage</i>	54
<i>Output format</i>	54
Converting PDF Image Only to Searchable PDFs	55
<i>Format Retention</i>	55
Configuring Document Types for OCR	56
<i>Configuring a Document Type to Use Renditions</i>	56
<i>Assigning an OCR Format to a Document Type</i>	57
CONFIGURING POINT AND SHOOT INDEXING	57
<i>A Note About Single and Double-Byte Characters</i>	57
CONFIGURING AUTOMATED INDEXING	57
Configuring Automated Indexing Forms for Institutional Databases	58
<i>Using Automated Indexing Forms with Institutional Databases</i>	60
Testing Automated Indexing Forms	61
Performing Automated Indexing Processes	61
Granting Users the Rights to Configure Automated Indexing Forms	61
Configuring Automated Indexing Forms	64
<i>Accessing the Automated Index Configuration Window</i>	64
Opening the Automated Index Configuration Window from the Client	66
Opening the Automated Index Configuration Window from the Document Imaging Window	67
The Automated Index Configuration Window	68
<i>Creating a New Automated Indexing Form</i>	73
Manually Matching Existing Automated Indexing Forms to Documents in the Batch	76
Configuring Document Breaks	79
Specifying Automated Indexing Form Availability by Document Type	81
Configuring Scan Queue or Keyword Value Rules for a Form	85
Configuring Keyword Lookup/Replace Dictionaries	93
Modifying or Deleting a Form	105
<i>Configuring A Form Identification Zone</i>	109
Configuring a Text Value Form Identifier	115
Configuring an Image Match Form Identifier	121

Modifying, Deleting or Resizing/Moving a Form Identification Zone	125
Configuring A Data Field Zone	126
Configuring a Text Data Field Zone	131
Configuring a Line Item Extraction Field Zone	144
Configuring a Grouped Line Item Extraction Data Field Zone	153
Configuring an Optical Mark Data Field Zone	167
Configuring a Signature Detection Data Field Zone	175
Modifying, Deleting, Resizing/Moving, or Copying a Data Field Zone	179
Configuring a Page Registration Zone	181
Configuring a Page Registration Zone to Use a Text Value	186
Configuring a Page Registration Zone to Use an Image Match	191
Modifying, Deleting, or Resizing/Moving a Page Registration Zone	195
Modifying, Deleting, Resizing/Moving, or Copying Overlapping Zones	195
Importing/Exporting Automated Indexing Forms	196
Importing Automated Indexing Forms	197
Reconciling Configuration References	199
Exporting Automated Indexing Forms	201
Automated Indexing Form Order Analysis	204
Configuring XML Renditions	208
Testing Automated Indexing Forms	208
Configuring Automated Indexing to Use Intelligent Character Recognition	211
Limitations of Intelligent Character Recognition	213
Improving the Accuracy of Intelligent Character Recognition	214
Configuring a Scan Queue for Automated Indexing	215
Disabling Images of Keyword Values from Displaying During Indexing	218
Disabling Automatic Keyword Value Zoom	219
Navigating Only Between Suspect Keyword Values	219
Configuring Automatic Automated Indexing Processes	220
Automating Automated Indexing at the Scan Queue Level	220
Automating Automated Indexing via a Scheduled Process	223
Automating Automated Indexing via a Command Line Switch	223
Exiting the Automated Indexing Form Configuration Interface	224

Installation

REQUIREMENTS	225
OCR Processing Workstation Supported Operating Systems	225
Operating System Requirements	225
OCR Processing Workstation Hardware Requirements	226
Databases Supported	226
Microsoft SQL Server	227
Oracle	227

<i>Sybase SQL Anywhere</i>	228
Database Client / Server Version Compatibility	228
About Virtual Environments	229
64-Bit Support Statement	230
Licensing	230
PRE-INSTALLATION	230
Hyland OCR Engine for OnBase	230
<i>Additional Information</i>	231
INSTALLATION	231
COMMAND LINE SWITCHES	233
Auto-Processing Switches	233
<i>-SCANAUTOINDEX</i>	233
<i>-SCANAUTOQUEUE</i>	233
<i>AutoProcessDelay INI Setting</i>	233
INI OPTIONS	234
<i>Scan</i>	234
<i>AutomatedIndex</i>	234
INI File	235
<i>Previous File Location/File Name</i>	235
<i>Location</i>	236
<i>INI Considerations in a Citrix and Microsoft Windows Remote Desktop Environment</i>	236
<i>Editing the INI File</i>	237
BACKUP/RECOVERY	237
Backup	237
<i>Configuration</i>	237
<i>Registry Settings</i>	237
<i>External Files</i>	237
<i>Module-Related INI Options</i>	238
<i>Additional Steps</i>	238
Recovery	238
<i>Configuration</i>	238
<i>Registry Settings</i>	238
<i>External Files</i>	238
<i>Module-Related INI Options</i>	238
<i>Registration</i>	239
<i>Additional Steps</i>	239
TROUBLESHOOTING	239
Common Issues	239
<i>Process Locks</i>	240
Automated Indexing & AutoFill Keyword Sets	241
CONTACTING SUPPORT	242

Scheduling

SCHEDULING OVERVIEW	245
CONFIGURING & USING THE SCHEDULER	245
Requirements for Configuring/Running a Scheduled Process	245
<i>Using the -SCHED and -SCHEDINST Switches</i>	<i>246</i>
<i>-SCHED</i>	<i>246</i>
<i>-SCHEDINST</i>	<i>246</i>
Verifying the Scheduler is Running	246
Running Multiple Scheduled Processes	247
Scheduled Process Configuration Reports	248
Working With Process Formats	248
<i>Creating a Scheduled Process Format</i>	<i>248</i>
Schedule Configuration	251
Processing Options	256
<i>Viewing Scheduled Processes</i>	<i>260</i>
<i>Modifying a Scheduled Process Format</i>	<i>260</i>
<i>Deleting a Scheduled Process Format</i>	<i>261</i>
<i>Running/Suspending a Scheduled Process Format</i>	<i>261</i>
Working With Process Jobs	262
<i>Creating a Job</i>	<i>263</i>
<i>Configuring a Job</i>	<i>264</i>
<i>Scheduling a Job</i>	<i>266</i>
Schedule Configuration	266
Processing Options	271
<i>Viewing a Job</i>	<i>273</i>
<i>Modifying a Job</i>	<i>273</i>
Renaming a Job	274
<i>Deleting a Job</i>	<i>274</i>
<i>Running/Suspending a Job</i>	<i>275</i>
Creating Schedule Templates	275
Creating Schedule Templates	276

Automated Indexing Best Practices

GENERAL INFORMATION	281
Recommended Standards for OCR Processing	281
Recommended Standards for ICR Processing	282
USAGE	283
Automated Indexing	283
<i>Ad-Hoc Automated Indexing</i>	<i>283</i>
<i>Batch Automated Indexing</i>	<i>283</i>

Performing Point and Shoot Indexing	283
CONFIGURATION	284
Automated Indexing Form Configuration.....	284

OVERVIEW

Automated Indexing allows documents to be automatically indexed by extracting indexing values (i.e., a Document Type identifier, the Document Date and Keyword Values) directly from documents via optical character recognition (OCR) technology.

Automated Indexing consists of two distinct, but equally powerful, functionalities:

- **Automated Indexing.** Automated Indexing automatically compares documents to pre-defined Automated Indexing forms, or templates, that define the areas of the document that are to be processed. Once a document is matched to a form, it is processed and the extracted indexing values are assigned to it.

Automated Indexing can be performed on image or PDF documents.

- **Point and Shoot Indexing.** Point and Shoot Indexing allows you to select an area of a document to be processed so a Document Date or Keyword Values can be extracted from that area.

Point and Shoot Indexing can be performed on image documents only.

Both Automated Indexing and Point and Shoot Indexing support for Keyword data type recognition and Keyword Masking.

Your Automated Indexing solution is able to identify and process machine-printed data via optical character recognition (OCR). If your solution is also licensed for **ICR Support for Full-Page OCR & Automated Index**, your solution is also able to identify and process hand-written text via intelligent character recognition (ICR).

Note: ICR is not supported for Point and Shoot Indexing.

Unlike other forms recognition products, Automated Indexing has no per-page or per-document restrictions or costs, and all Automated Indexing is performed within, and is seamlessly integrated into, OnBase's Document Imaging interface – no external databases or additional hardware or software is required.

Automated Indexing

1

Like other document processing options, such as Bar Code Processing or Full-Page OCR, Automated Indexing can be performed on batches of documents within a scan queue. It can also be performed on a single document from several places in the OnBase Client module, including the Document Viewer, the **Import** dialog box, or, with additional licensing, from within a Workflow queue.

Automated Indexing can help to streamline most indexing tasks. Automated Indexing can do much more than identify and extract basic indexing data; depending on your configuration, Automated Indexing can also perform these advanced indexing tasks:

- Automatically score and evaluate indexing data read by the Automated Indexing engine to determine if a document needs to be manually reviewed by a user for accuracy.

Batches of documents that are determined to have been accurately indexed can be marked as fully indexed and committed without any user intervention.

- Group and store line-item values using Multi-Instance Keyword Type Groups.
- Perform Optical Mark Recognition (OMR) and signature detection, and assign appropriate Keyword Values based on these results.
- Read MICR line data from check documents.

Automated Indexing is easily configured from using a visual, point-and-click interface. Users can create, modify and delete Automated Indexing forms -- templates that allow a user to define the areas of the document that are processed and the type of indexing data (i.e., a Document Type identifier, the Document Date or Keyword Value) that is extracted from each area.

A user with the proper rights can create and edit Automated Indexing forms from within the OnBase Client or the Document Imaging interface. Access to the OnBase Configuration module is not needed.

During processing, a document is classified and indexed based on the form matched to it by the Automated Indexing engine. Multiple forms can be created and associated with a single OnBase Document Type, allowing documents with different physical layouts to be processed and assigned to the same Document Type.

For example, an organization might scan and archive all of its shipping receipts, from any number of vendors, in one Document Type. Even though all of these documents have a different physical layout, as long as Automated Indexing is able to match an Automated Indexing form to each one, they can all be properly processed, indexed and archived in the same Document Type.

If Automated Indexing cannot match a form to a document, or if its built-in quality controls determine that a value returned by the Automated Indexing engine is suspect, the batch containing the document is routed for user review. Documents are reviewed and corrected in the familiar environment of OnBase Document Imaging; no external programs or additional software is required.

Point and Shoot Indexing

Point and Shoot Indexing offers organizations an automated and accurate alternative to manual indexing without requiring a major change to existing scanning and indexing processes. Users can perform Point and Shoot Indexing when indexing individual image documents from within Document Imaging without needing pre-configured Automated Indexing forms.

Point and Shoot Indexing can be performed in any of Document Imaging's indexing batch status queues (e.g., **Awaiting Indexing, Index in Progress, Awaiting Reindex, Awaiting QA Reindex**, etc.). Users can capture a value from the scanned image and automatically assign it to the selected Keyword Type simply by selecting the area of the image containing the desired value with the pointer. This area is automatically processed and the selected Keyword field is populated with the extracted data.

APPLICATIONS

Automated Indexing and Point and Shoot Indexing can be used by any organization that is looking for ways to reduce the cost and chance for error that exist when manually indexing documents.

- An Accounts Payable department can use Automated Indexing to automatically index all incoming vendor invoices. Even though no two of the invoices are physically alike, an Automated Indexing form can be created for each different layout and all of the documents can still be archived under the same Document Type. By automating the indexing process, the clerk or clerks that used to manually index these invoices now have additional time to work on more critical tasks, and the capacity of the department to process invoices can be greatly increased without any additional staff.
- A Human Resources department can take advantage of Point and Shoot Indexing to improve its hiring process. As cover letters and resumes are received from applicants, the HR specialists can scan the paper documents, or sweep batches of electronically-submitted images, and use Point and Shoot Indexing to extract Keyword Values from the images more accurately and efficiently than if they had to manually index them.

- A bank with several regional offices can use Disconnected Scanning to scan documents into a central OnBase system. One or more centralized Automated Indexing processing workstations can be designated to perform Automated Indexing on all batches uploaded to the system from each of the regional offices.

Additionally, to conserve network capacity and bandwidth during business hours, Disconnected Scanning's scheduling functionality can be used to schedule batch uploads and the OnBase Scheduler can be used to schedule Automated Indexing processes to take place at night.

SYSTEM PREREQUISITES

In order to take advantage of batch Automated Indexing processing, your OnBase solution must be properly configured to import batches of images or PDF documents into a scan queue.

A number of OnBase modules, such as Document Imaging, Disconnected Scanning, or the Document Import Processor (DIP), can be used to import documents into batches via scan queues.

Additionally, the Hyland OCR Engine for OnBase software must be installed on each workstation that is to perform Automated Indexing. This software is available from your solution provider, and does not need separate licensing or registration.

LICENSING

Each Automated Indexing workstation requires a valid OnBase Client license, in addition to one of the following Automated Indexing licenses:

- **Ad-Hoc Automated Indexing.** An Ad-Hoc Automated Indexing license allows you to perform Point and Shoot Indexing and Automated Indexing on an open document from the right-click menu.
- **Automated Indexing.** An Automated Indexing license allows you to perform Point and Shoot Indexing and Automated Indexing on an open document as well as on batches residing in the **Awaiting Automated Indexing** or **Awaiting Ad Hoc Automated Index** batch status queues.

Note: If your solution is licensed for **Batch Automated Indexing** (a license available for OnBase solutions prior to the release of OnBase 10.0), contact your solution provider for information on obtaining a current **Automated Indexing** license.

If your solution is licensed for at least one Ad-Hoc Automated Indexing or Automated Indexing license, users of workstations not registered for Automated Indexing can route batches to the **Awaiting Ad Hoc Automated Index** batch status queue for processing by a centralized workstation that is licensed for Automated Indexing.

Optional Licensing

Additional licensing is required in order to use the following features in conjunction with Automated Indexing:

- A valid OnBase Workflow license is required to initiate Automated Indexing from within a Workflow Life Cycle. For more information, see the Workflow documentation.
- In order to mark areas of a document that have been processed via Automated Indexing for redaction via Automated Redaction, your solution must also be licensed for Automated Redaction. For more information, see the Automated Redaction.
- To identify handwritten text in an Automated Indexing Data Field Zone using Intelligent Character Recognition (ICR), your solution must be licensed for ICR Support for Full-Page OCR & Automated Index. This site license allows all of your Automated Indexing OCR processing workstations to perform ICR processing.

Check your current licensing status by selecting **Utils | Product Licenses** from the Configuration module.

1

Automated Indexing allows you to automatically extract indexing values, such as Document Type identifiers, the Document Date and Keyword Values, directly from an image or PDF document.

Point and Shoot Indexing allows you to extract the Document Date and Keyword Values directly from any image document displayed in the Document Imaging Working window.

AUTOMATED INDEXING

Both ad-hoc and batch Automated Indexing allow you to index a document by automatically matching it to Automated Indexing forms. Automated Indexing forms are pre-configured templates that have been created with certain zones, or fields, that are analyzed for a specified type of indexing information (i.e., a Document Type identifier, the Document Date or Keyword Values).

The Automated Indexing engine compares all available Automated Indexing forms to a document until it is able to match a form to the document's physical layout. Once a match is determined, the document is processed and the indexing information is automatically applied to the document.

You can perform ad-hoc Automated Indexing from a number of places within the Client module, and, if your workstation is not registered for Automated Indexing, you can queue documents for processing at a later time or on another workstation. Depending on your licensing, you may also be able to perform batch Automated Indexing on a batch of documents from within the Document Imaging interface.

Performing Ad-Hoc Automated Indexing

Ad-hoc Automated Indexing allows you to perform Automated Indexing on a document while working with the document. Depending on how you are interacting with the document, ad-hoc Automated Indexing can be initiated via right-click option or by pressing a button on a dialog box.

Ad-hoc Automated Indexing can be performed from the following areas of the OnBase Client:

- **From the Document Search Results list.** See From the Document Search Results List on page 9 for more information.
- **From an open document.** See From an Open Document on page 11 for more information.
- **From the Document Imaging window.** See From the Document Imaging Window on page 14 for more information.
- **From the Import Document dialog box.** See From the Import Document Dialog Box on page 17 for more information.
- **From a Folder.** See From a Folder on page 19 for more information.
- **From a Workflow Queue.** See the Workflow entry in the System Interaction section on page 44 for more information.
- **From a document being created and automatically imported by the Virtual Print Driver.** See the Virtual Print Driver entry in the System Interaction section on page 44 for more information.

If you are at a workstation that has either an Automated Indexing or Ad-Hoc Automated Indexing license registered to it, the document will immediately undergo Automated Indexing.

If your workstation is not registered for Automated Indexing or Ad-Hoc Automated Indexing, but your OnBase solution is licensed for one or both of these products, the document will be routed to the **Awaiting Automated Indexing** batch status queue and will be processed later, when a license becomes available, or by another workstation licensed for Automated Indexing.

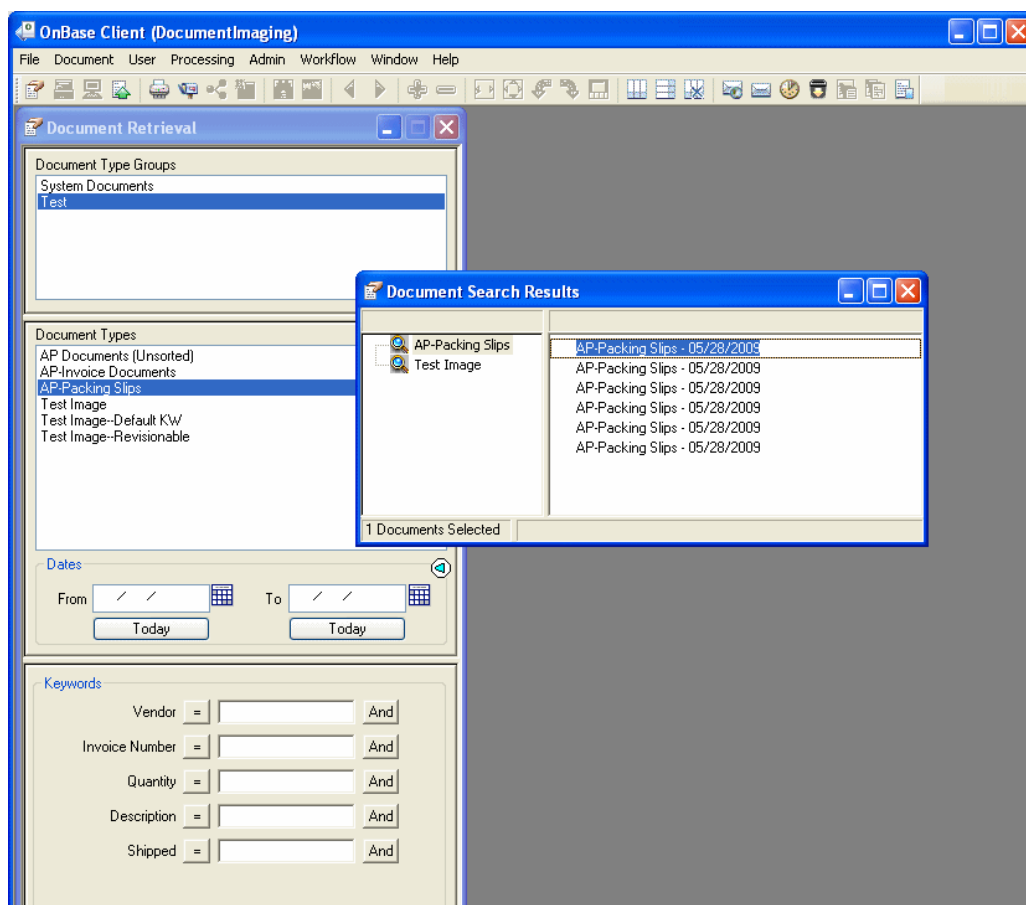
Tip: Unlike batch Automated Indexing, ad-hoc Automated Indexing offers no indication of values that were not confidently identified by the Automated Indexing engine (i.e., values that did not meet their configured Suspect Level). It is considered a best practice to always compare the indexing information assigned to the document by the ad-hoc Automated Indexing process to the actual information displayed on the document to ensure that it is accurate.

From the Document Search Results List

To perform ad-hoc Automated Indexing from the **Document Search Results** list:

1. From the OnBase Client, search for a document. The **Document Search Results** list is displayed.

2



2. Right-click on the document that you would like to perform ad-hoc Automated Indexing on and select **Perform Document Automated Index...**

The Automated Indexing process is performed.

Note: If you are using Ad Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document. For more information on how AutoFill Keyword Sets function with Automated Indexing, see AutoFill Keyword Sets on page 38.

Note: If your workstation is not registered for Ad Hoc Automated Indexing or Automated Indexing, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing. In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

From an Open Document

To perform ad-hoc Automated Indexing on an open document:

1. From the OnBase Client, retrieve and open the document you wish to perform ad-hoc Automated Indexing on. The document is displayed in the Document Viewer.

AP-Packing Slips - 05/28/2009

Packing Slip

Invoice No: 3330
Date: March 3, 2008

RECEIVED
FD 12-5-4-08

Phone Shack
5546 South Street
Johnstown, PA 15901

To: 9-Second Foods, Inc.
19000 Quickly Circle
Columbia, MD 21041

Ship To: Same

Salesperson	P.O. Number	Date Shipped	Shipped Via	F.O.B. Point	Terms
Joe Smith	1888	3/3/08	UPS		2% 15 Net 30

Quantity	Description	Shipped	Received
40	722588 Siemens Black Phones	40	

Quantity Shipped	40
Quantity Received	
Accepted By	
Date Received	

Make all checks payable to: Phone Shack
If you have any questions concerning this Packing Slip, call:
Joe S., 814-555-2234

Thank you for your business!

Test... **Page 1** **1 Page(s)**

- Right-click on the document and select **Perform Document Automated Index....**
The Automated Indexing process is performed.

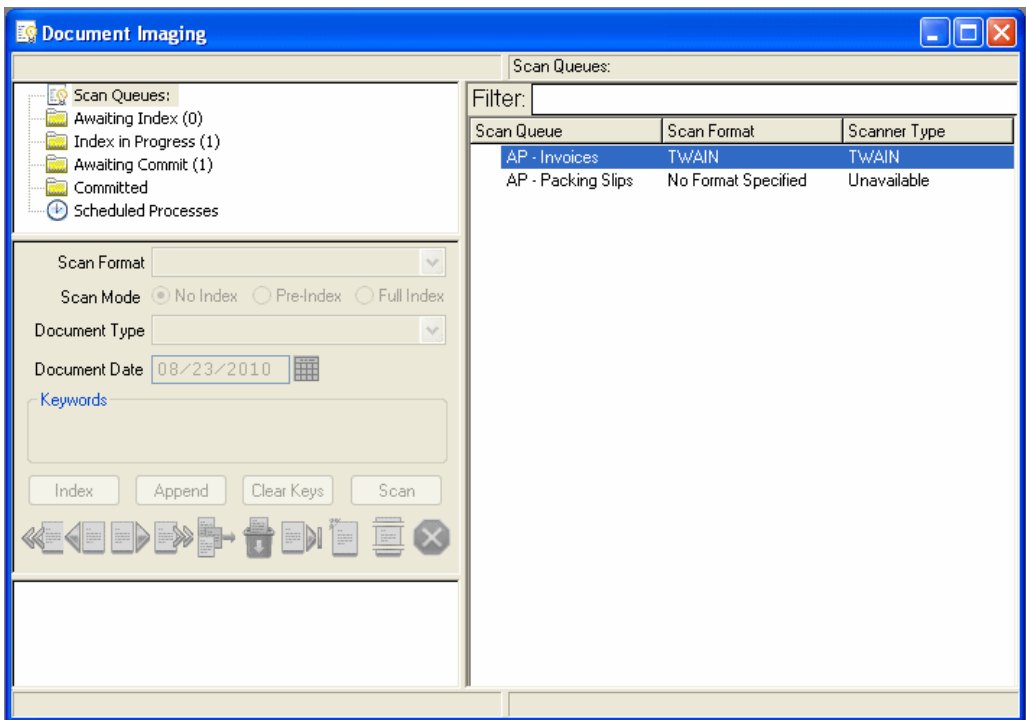
Note: If you are using Ad Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document. For more information on how AutoFill Keyword Sets function with Automated Indexing, see AutoFill Keyword Sets on page 38.

Note: If your workstation is not registered for Ad Hoc Automated Indexing or Automated Indexing, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing. In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

From the Document Imaging Window

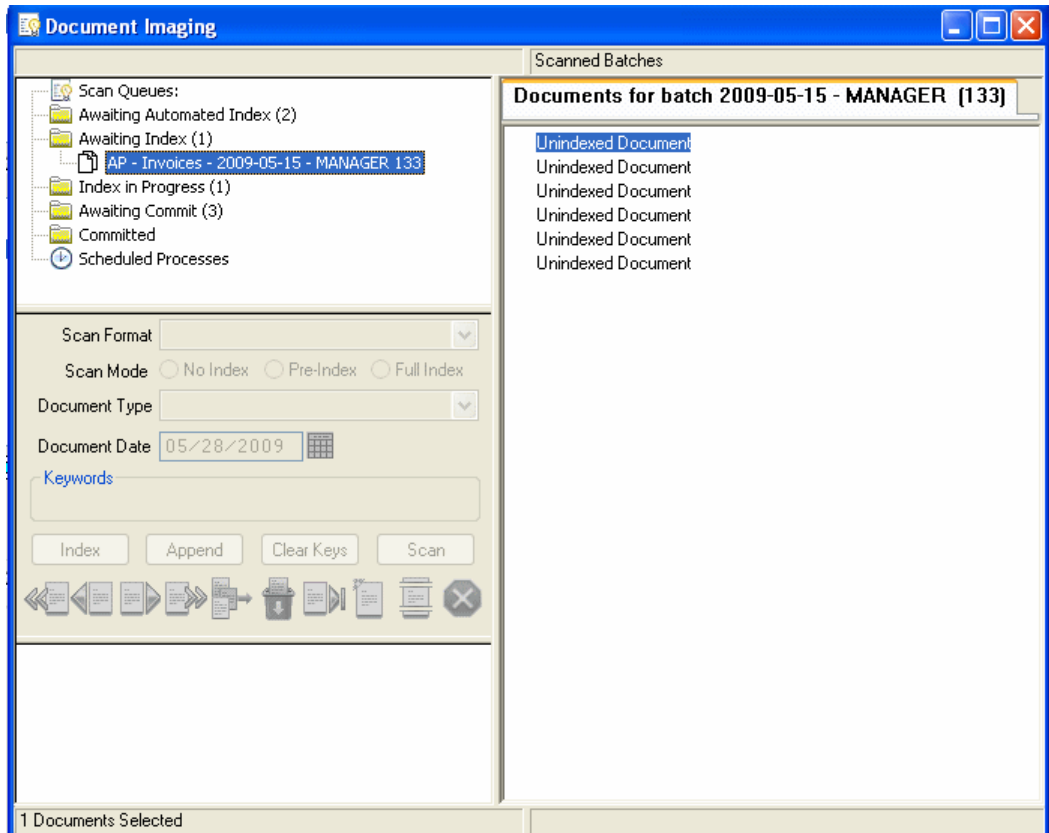
To perform ad-hoc Automated Indexing on a document from within the Document Imaging interface:

1. From the OnBase Client, click **Processing | Scan/Index**. The Document Imaging window is displayed.



2. In the Queue List window, select the batch status queue in which the document you wish to index resides. All batches residing in that batch status queue are listed in the working window.

3. In the Working window, double-click the batch containing the document you wish to perform ad-hoc Automated Indexing on. All documents residing in that batch are listed in the Working window.



4. In the Working window, select the document you wish to perform ad-hoc Automated Indexing on. Right-click and select **Perform Automated Index....**

The Automated Indexing process is performed. The document and its associated batch remain in the **Awaiting Index** batch status queue.

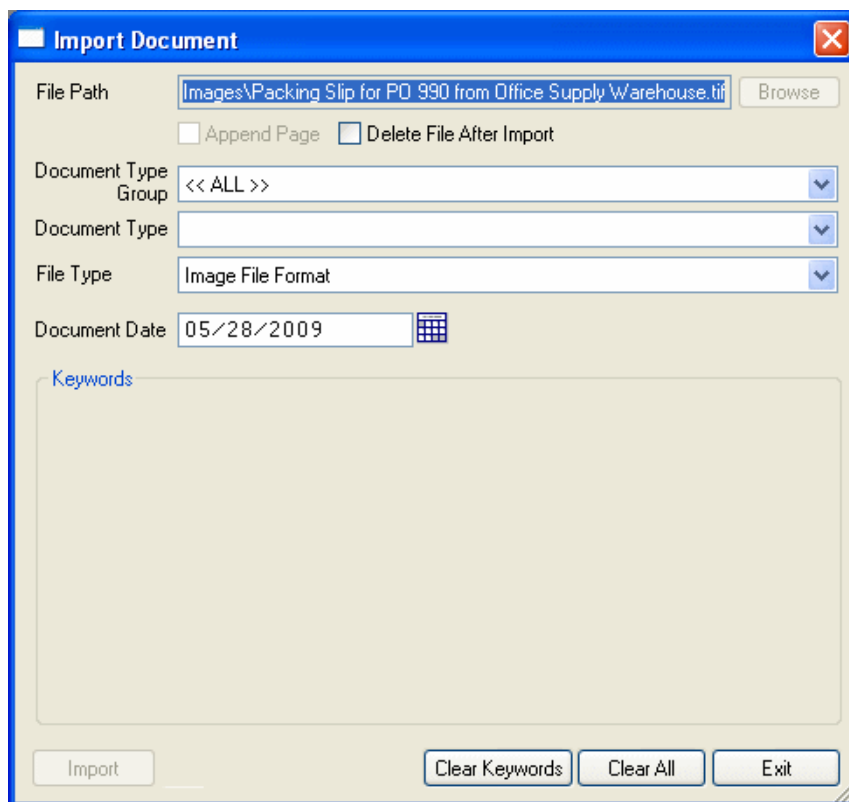
Note: If you are using Ad Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document. For more information on how AutoFill Keyword Sets function with Automated Indexing, see AutoFill Keyword Sets on page 38.

Note: If your workstation is not registered for Ad Hoc Automated Indexing or Automated Indexing, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing. In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

From the Import Document Dialog Box

To perform ad-hoc Automated Indexing from the **Import Document** dialog box:

1. Import a document into OnBase. The **Import Document** dialog box is displayed.



Note: For information on the various methods to ad-hoc import documents, see the Client documentation.

- Using the Document Type drop-down, select the Document Type for the document being imported. Once the Document Type is selected, the **Queue for Automated Index** check box is displayed.

2

Import Document

File Path: C:\SharpFTP\Packing Slip for PO 990.tif

☐ Append Page ☐ Delete File After Import

Document Type Group: Accounts Payable

Document Type: AP - Packing Slip

File Type: Image File Format

Document Date: 12/01/2010

☐ Queue for Automated Index

Keywords

PO #:

Vendor #:

Vendor Name:

Invoice #:

- Select the **Queue for Automated Index** check box and, if desired, enter any other indexing information for the document.
- Click **Import**.

The Automated Indexing process is performed.

Note: If you are using Ad Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document. For more information on how AutoFill Keyword Sets function with Automated Indexing, see AutoFill Keyword Sets on page 38.

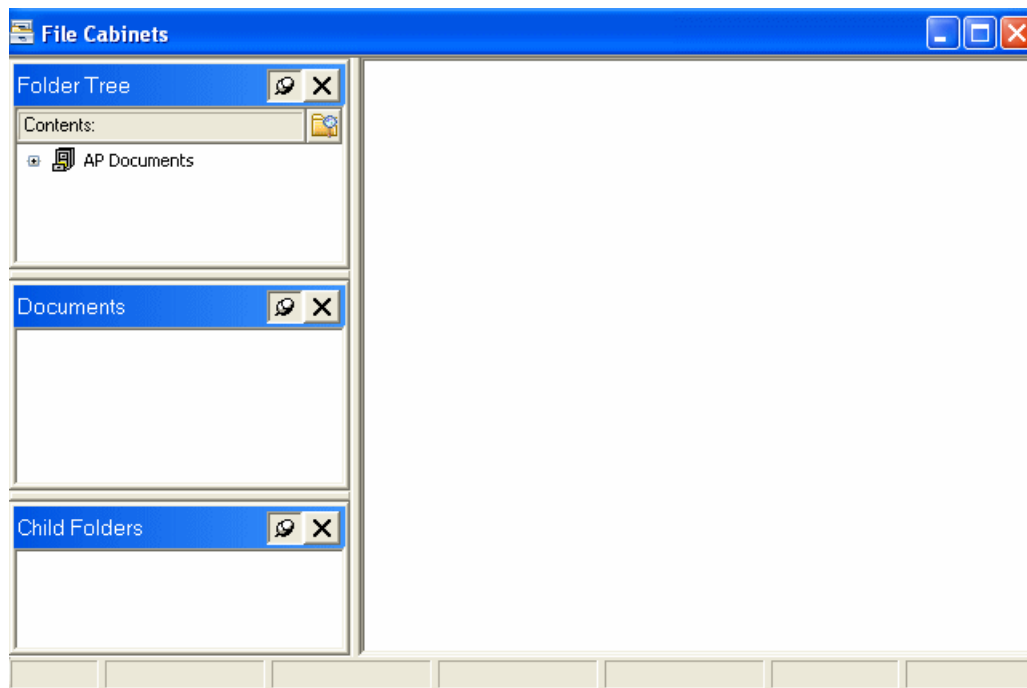
Note: If your workstation is not registered for Ad Hoc Automated Indexing or Automated Indexing, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing. In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

From a Folder

Note: Ensure that Folders have been properly configured for your OnBase solution. For more information, see the Folders documentation.

To perform ad-hoc Automated Indexing from a Folder:

1. From the OnBase Client, open the File Cabinets window by clicking **File | File Cabinets** or by clicking the **Open File Cabinets** toolbar button.



2

- Expand the folder until you reach the document that you would like to perform ad-hoc Automated Indexing on. The document is displayed in the Document Viewer.

File Cabinets - AP-Packing Slips - 06/02/2009

Folder Tree

Contents: Static / Dynamic

- AP Documents
 - AP Documents - 06/02/2009
 - Packing Slips - 06/02/2009

Documents

AP-Packing Slips - 06/02/2009

Child Folders

Packing Slip

INVOICE NO: 1330
DATE: March 3, 2008

RECEIVED
3/3/08

Computers Are Us
12345 Lincoln Blvd., Suite 225
Lincoln, NE 68501

To: 9-Second Foods, Inc.
19000 Quickly Circle
Columbia, MO 65201

Ship To: Same

Salesperson	P.O. Number	Date Shipped	Shipped via	F.O.B. Point	Terms
Bill Smith	881	3/3/08	UPS		2% 10; net 30

Quantity	Description	Shipped	Received
2	4 GB MEM Upgrade	2	
2	Video Card	2	
		Quantity Shipped	4
		Quantity Received	
		Accepted By	
		Date Received	

Make all checks payable to: Computers Are Us
If you have any questions concerning this Packing Slip, call:
Bill Smith, 402-555-2234

Thank you for your business!

0 Note(s) 0 Highlight(s) Page 1 1 Page(s)

3. To perform Automated Indexing on the displayed document, perform one of the following:
 - With the document selected in the Documents pane, right-click within the Documents pane and selected **Perform Document Automated Index....**
 - Right-click on the document in the Document Viewer and select **Perform Document Automated Index....**

The Automated Indexing process is performed.

Note: If you are using Ad Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document. For more information on how AutoFill Keyword Sets function with Automated Indexing, see AutoFill Keyword Sets on page 38.

Note: If your workstation is not registered for Ad Hoc Automated Indexing or Automated Indexing, a message is displayed notifying you that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing. In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

Performing Batch Automated Indexing

Automated Indexing can be performed on batches of documents within the Document Imaging interface.

Batch Automated Indexing allows you to immediately and automatically perform Automated Indexing on a large number of documents imported into a scan queue via a batch import module, such as Document Imaging, Disconnected Scanning, Document Import Processor, etc., after import.

Tip: When scanning a batch of documents into a scan queue, it is considered a best practice to scan the batch as **Unindexed** or **Pre-Indexed** so that any documents that undergo batch Automated Indexing and are identified as having suspect Keyword Values are sent for user review in the **Index in Progress** batch status queue. If the batch is scanned as **Fully Indexed**, batches containing documents that are associated with suspect Keyword Values are not routed for manual review in the **Index in Progress** batch status queue because the batch is already marked as indexed.

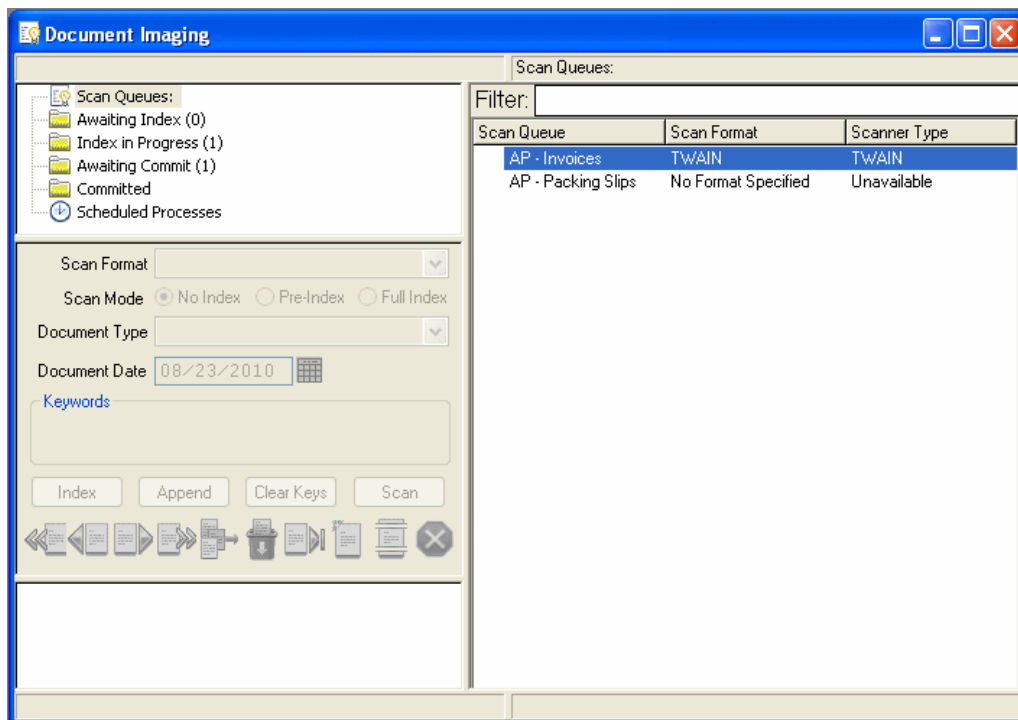
One of the benefits of performing batch Automated Indexing is the identification of values that are not confidently identified by the Automated Indexing engine. Each indexing value identified by the Automated Indexing engine is given a score, and if this score does not meet the specified Suspect Level, the value is highlighted and the document is brought to the user's attention.

Batch Automated Indexing also allows you to perform Automated Indexing on documents that have been queued for ad-hoc Automated Indexing by a workstation not registered for Automated Indexing.

On a Newly-Imported Batch

To perform Batch Automated Indexing on a newly-imported batch of documents:

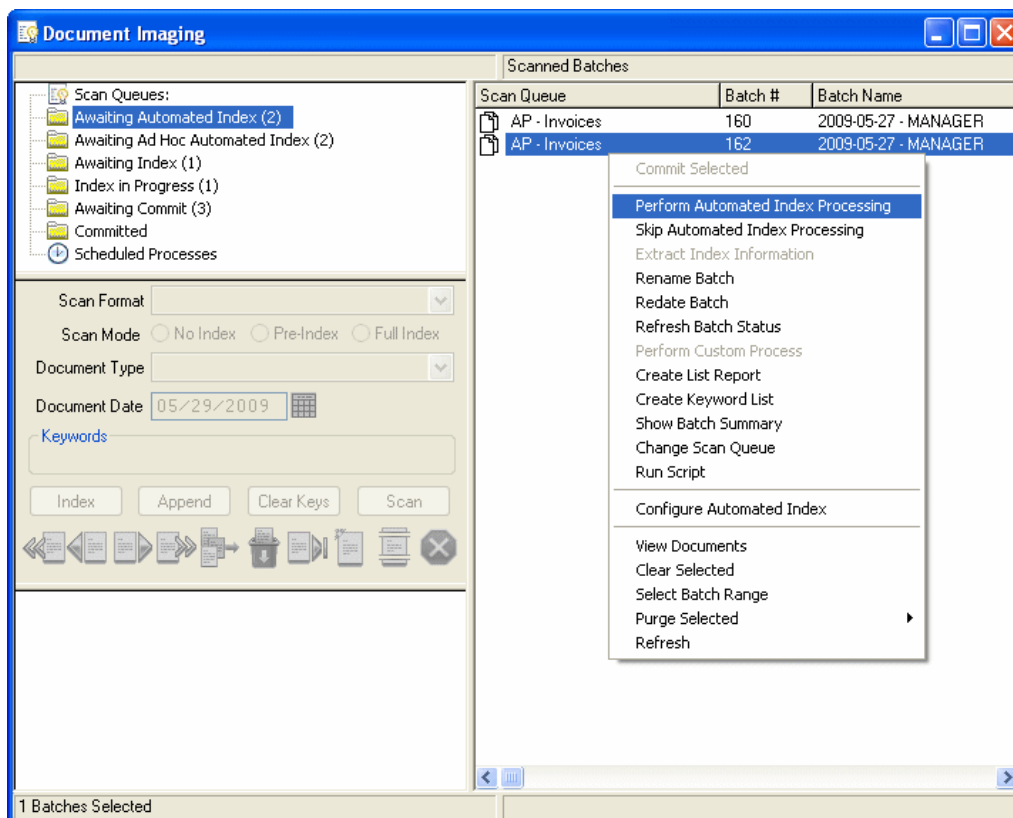
1. Import a batch of documents into OnBase via a scan queue configured to use Automated Indexing.
2. From the OnBase Client, click **Processing | Scan/Index**. The Document Imaging window is displayed.



Note: Depending on your scan queue configuration, the batch you imported in Step 1 may have been routed to a batch status queue prior to the **Awaiting Automated Index** queue. If necessary, perform any necessary activities to route the batch to the **Awaiting Automated Index** batch status queue prior to continuing to Step 3.

3. In the Queue List window, select the **Awaiting Automated Index** batch status queue. All batches residing in the queue, including the batch you imported in Step 1, are displayed in the Working window.

4. In the Working window, select the batch that you would like to perform batch Automated Indexing on. Right-click and select **Perform Automated Index Processing**.



5. A message is displayed asking you to confirm that you would like to perform batch Automated Indexing on the selected batch. Click **Yes**.

The batch Automated Indexing process is performed.

Note: If you are using Batch Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either all instances of the AutoFill Keyword Set are assigned to the document or the batch is routed to the **Index in Progress** batch status queue so that you can select the instance, or instances, of the AutoFill Keyword Set to assign to the document. For more information on AutoFill Keyword Sets and Automated Indexing, see AutoFill Keyword Sets on page 38.

6. Once the Automated Indexing process is complete, the batch is transitioned out of the **Awaiting Automated Index** batch status queue.
 - If the Automated Indexing process was successful, and all values processed by the Automated Indexing engine meet the acceptable Suspect Level, then the batch is transitioned to the next batch status queue configured for the scan queue.
 - If one or more values processed by the Automated Indexing engine is marked as suspect, or if any forms matched to documents in the batch are configured to also require documents to undergo manual indexing, the batch is transitioned to the **Index in Progress** batch status queue for additional indexing or index verification.

INDEX VERIFICATION

Index verification must be performed if indexing values for documents that undergo batch Automated Indexing do not meet their configured Suspect Levels during processing.

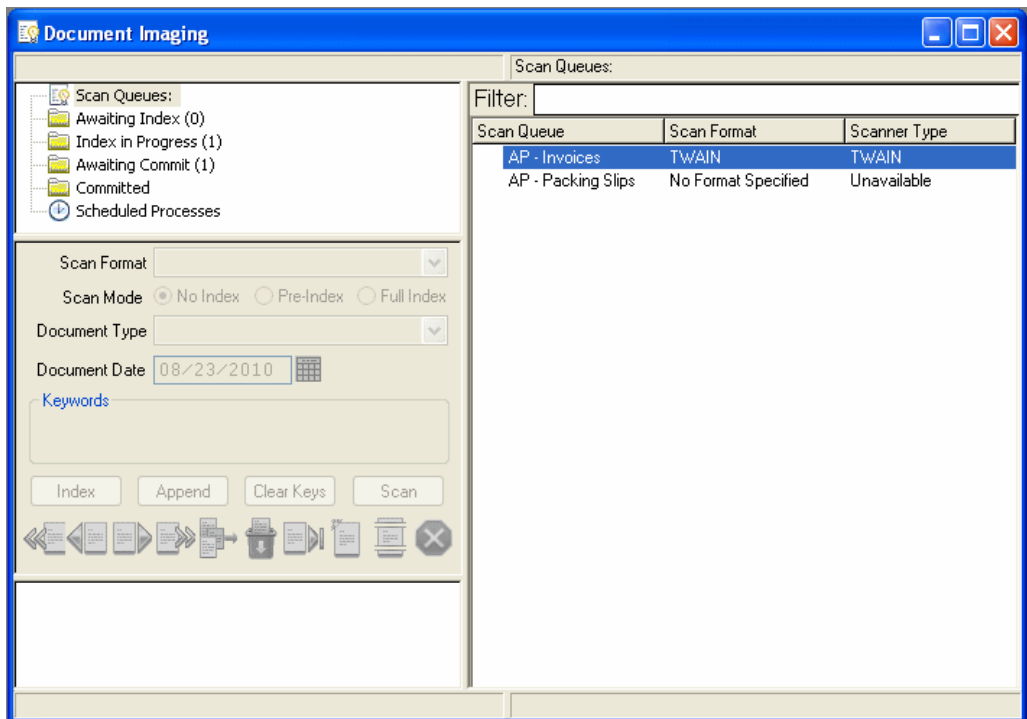
If at least one value on one document in a batch is determined to be suspect, then the entire batch is moved to the **Index in Progress** batch status queue for index verification.

Note: Index verification is only available for batch Automated Indexing. When performing ad-hoc Automated Indexing on a document, no indication is given if any detected values are suspect.

Note: Be aware that while a batch that has undergone batch Automated Indexing can be reviewed, corrected, or undergo further indexing in the OnBase Web Client, the index verification features available in the OnBase Client described below (e.g., suspect Keyword Values being highlighted in red) are not available in the OnBase Web Client.

To perform index verification:

1. From the OnBase Client, click **Processing | Scan/Index**. The Document Imaging window is displayed.



2. In the Queue List window, select the **Index in Progress** batch status queue. All batches residing in the queue are displayed in the Working window.

3. In the Queue List window, select the batch that you would like to perform index verification on. Right-click and select **Index Documents**.

2

Document Imaging

Documents for batch #166 (2009-05-29 - MANAGER)

Office Supply Warehouse
 8879 Blue Street
 Denver, CO 80002

To:
 9-Second Foods, Inc.
 19000 Quickly Circle
 Columbia, MO 65201

Ship To:
 Same

PO	Terms	Date Shipped	Ship
990	2% 10; Net 30	4/7/08	UPS

Keywords

Vendor	OFFICE SUPPLY	[Icon]
Invoice Number	2440	[Icon]
Quantity	10	[Icon]
Quantity	40	[Icon]
Quantity	100	[Icon]

Index Append Clear Keys Scan

Discarded Batch #163 (2009-05-29 - MANAGER)
 Barcode disabled

1 Batches Selected

Page 1 of 1 of document 1 of 4 (remaining) and 6 (total)

Note: When performing index verification, only documents in the batch that have a suspect value or that were left unindexed will be displayed.

The first document in the batch is displayed in the Working window and the **Indexing** dialog box is populated with its indexing information. Indexing values that were marked as suspect are highlighted in red in the **Indexing** dialog box.

Note: If a Keyword Value is displayed for review, but is not highlighted in red, the Keyword Value is the primary Keyword Value of more than one instance of an AutoFill Keyword Set. Once you review and accept the Keyword Value, you are asked to select the correct instance, or instances, of the AutoFill Keyword Set to assign to the document.

4. Click in the first suspect Document Date or Keyword Value field in the **Indexing** dialog box. Depending on your configuration, the area of the document that the value was extracted from is displayed in detail in the Working window and is highlighted in yellow, and the value recognized by the Automated Indexing engine is displayed in detail above the Keyword Type field.
5. Compare the Document Date value or Keyword Value to the value displayed on the document.

If the value is correct, navigate to the next suspect value on the document, and repeat Step 5. If no suspect values remain on the current document, click **Index** to display the next document in the batch that has at least one suspect value.

If the Document Date or Keyword Value is incorrect, manually enter the correct value in the field or extract it from the document using Point and Shoot Indexing.

Once the value has been corrected, navigate to the next suspect value on the document and repeat Step 5. If no suspect values remain on the current document, click **Index** to display the next document in the batch that has at least one suspect value.

6. Once all documents in the batch requiring index verification have been verified, the batch is automatically routed to the next configured batch status queue for the scan queue.

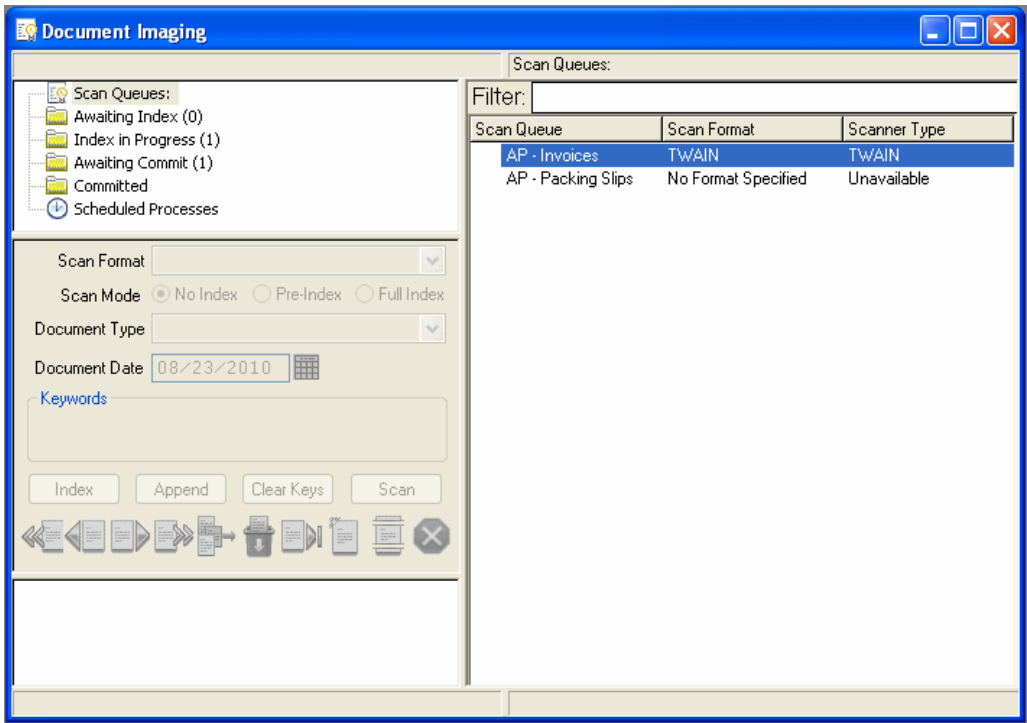
On Documents Queued for Ad-Hoc Automated Indexing

A workstation connected to an OnBase solution licensed for Automated Indexing or Ad-Hoc Automated Indexing has the ability to queue individual documents for Automated Indexing if the workstation is not registered for Ad-Hoc Automated Indexing or Automated Indexing.

These queued documents reside in the **Awaiting Ad Hoc Automated Index** batch status queue. They can be processed by a workstation registered for Automated Indexing.

To perform Automated Indexing on documents that have been queued for ad-hoc Automated Indexing:

1. From the OnBase Client, ensure that the workstation is registered for Automated Indexing and then click **Processing | Scan/Index**. The Document Imaging window is displayed.

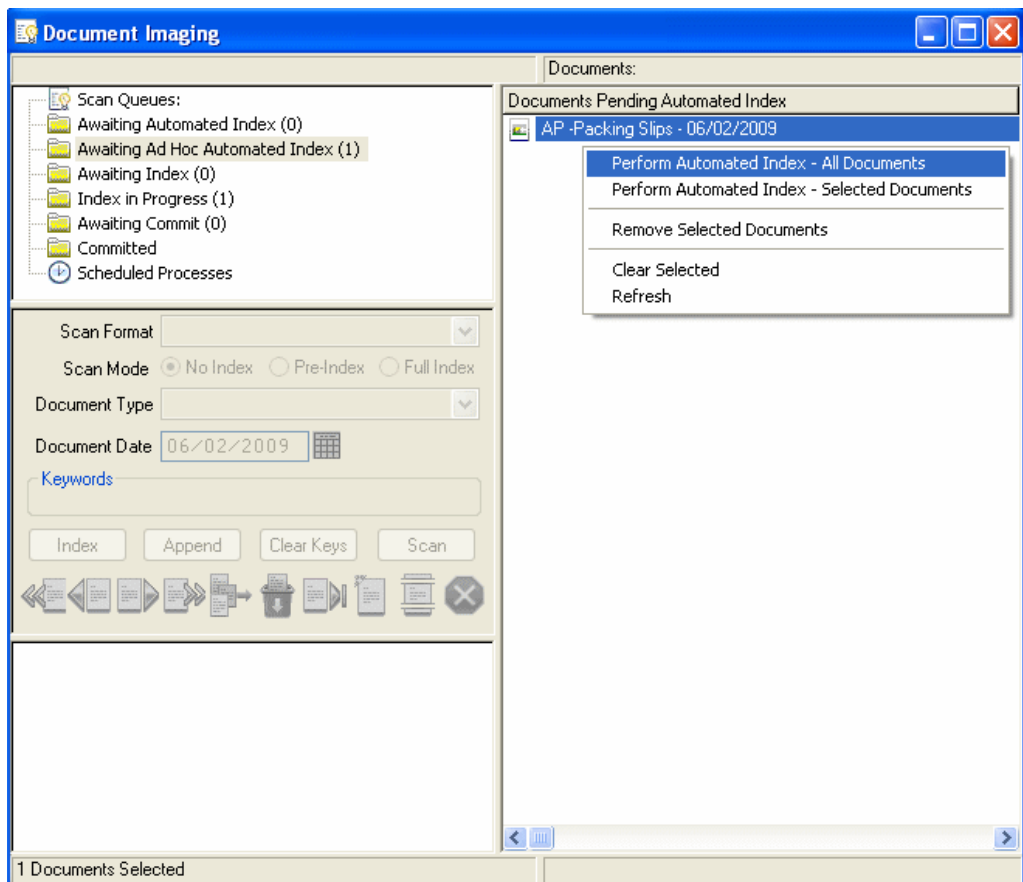


2. In the Queue List window, select the **Awaiting Ad Hoc Automated Index** batch status queue. All documents residing in the queue are displayed in the Working window.

Note: Despite the fact that they reside in Document Imaging batch status queue, be aware that these are individual documents that are not affiliated with a scan queue. No scan queue processing options will apply to the documents that undergo Ad-Hoc Automated Indexing.

3. In the Working window, select the document that you would like to perform Automated Indexing on. Right-click and select **Perform Automated Index Processing - Selected Document**.

To perform Automated Indexing on all documents residing in the batch status queue, right-click and select **Perform Automated Index Processing - All Documents**.



2

4. A message is displayed asking you to confirm that you would like to perform Automated Indexing on the selected document(s). Click **Yes**.

The Automated Indexing process is performed.

Note: If you are performing Automated Indexing on documents queued for Ad-Hoc Automated Indexing, be aware that when extracting a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, depending on the configuration of the AutoFill Keyword Set, either no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document) or all instances of the AutoFill Keyword Set are assigned to the document. You are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document. For more information on AutoFill Keyword Sets and Automated Indexing, see AutoFill Keyword Sets on page 38.

5. Once the process is complete, the documents are removed from the **Awaiting Ad Hoc Automated Index** batch status queue.

Tip: Processing documents in the **Awaiting Ad Hoc Automated Index** batch status queue is considered ad hoc Automated Indexing, despite the fact that an Automated Indexing license is required. Ad-hoc Automated Indexing offers no indication for values that did not meet the configured Suspect Level. It is considered a best practice to always compare the indexing information assigned to the document by the ad-hoc Automated Indexing process to the actual information displayed on the document to ensure that it is accurate.

POINT AND SHOOT INDEXING

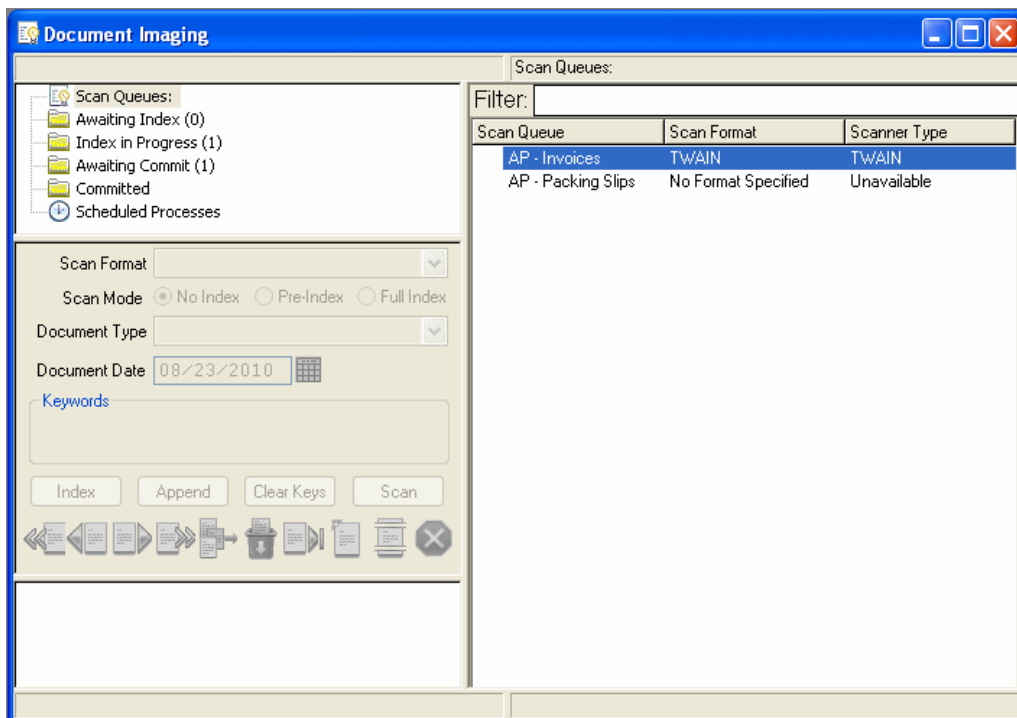
Point and Shoot Indexing allows you to extract Keyword Values from any image document during indexing or re-indexing from within the Document Imaging interface.

Point and Shoot Indexing can be performed by any user with rights to index documents on a workstation registered with an Automated Indexing or Ad-Hoc Automated Indexing license.

Performing Point and Shoot Indexing

To perform Point and Shoot Indexing:

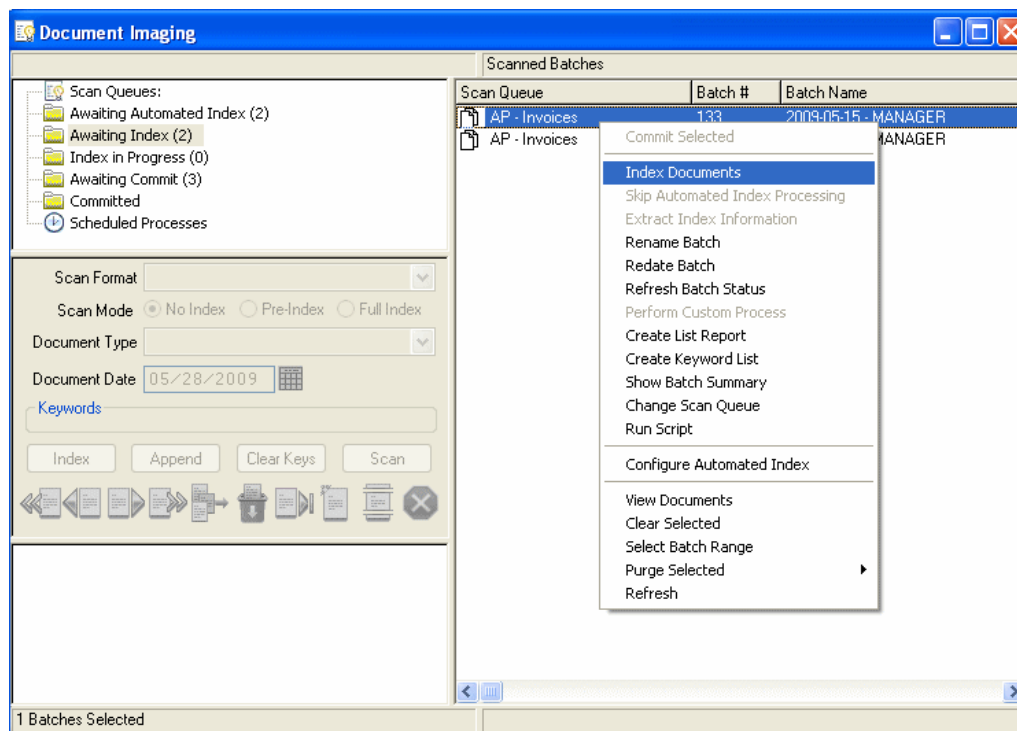
1. From the OnBase Client module, click **Processing | Scan/Index**. The Document Imaging window is displayed.



2. In the Queue List window, select an indexing or re-indexing batch status queue (i.e., **Awaiting Index**, **Index in Progress**, **Awaiting Reindex**, **Reindex in Progress**, **Awaiting QA Reindex**, etc.). All batches residing in the selected batch status queue are displayed in the Working window.

Note: The example shown below demonstrates on how to perform Point and Shoot Indexing on a document residing in the **Awaiting Indexing** batch status queue. Depending on the batch status queue you are using, some menu options and button names may be different. Consult the Document Imaging documentation for more information.

3. Right-click on a batch in the Working window and select **Index Documents**.



The first document in the batch is displayed in the Working window.

- If a Document Type has not already been selected for the displayed document, select one from the Document Type drop-down for the displayed document.

Once a Document Type is selected, the **Indexing** dialog box is populated with the Keyword Types associated with the Document Type.

The screenshot shows the Document Imaging application window. The left pane displays a tree view of Scan Queues, with 'Unindexed Document' selected. The right pane shows a document titled 'Packing Slip' for batch #133 (2009-05-15 - MANAGER). The document content includes a 'RECEIVED' stamp, a 'Phone Shack' logo, and a table of salesperson data. The Indexing dialog box is open, showing fields for Scan Format (Kofax), Scan Mode (No Index), Document Type (AP-Packing Slips), Document Date (05/28/2009), and Keyword fields (Vendor, Invoice Number, Quantity, Description, Shipped). The bottom status bar indicates '1 Batches Selected' and 'Page 1 of 1 of document 1 of 6 (remaining) and 6 (total)'.

Document Imaging

Documents for batch #133 (2009-05-15 - MANAGER)

Packing Slip

Invoice No: 3330
Date: March 3, 2008

RECEIVED
MAR 3 4 08

Phone Shack
5546 South Street
Johnstown, PA 15901

To: 9-Second Foods, Inc.
19000 Quickly Circle
Columbia, MO 65201

Ship To: Same

Salesperson	P.O. Number	Date Shipped	Shipped Via	F.O.B. Point	Terms
Joe Smith	888	3/3/08	UPS		2% 10, Net 30

Quantity	Description	Shipped	Received
40	722985 Siemens Black Phones	40	
			40
		Quantity Shipped	
		Quantity Received	
		Accepted By	
		Date Received	

Make all checks payable to: Phone Shack
If you have any questions concerning this Packing Slip, call:
Joe S., 814-555-2234

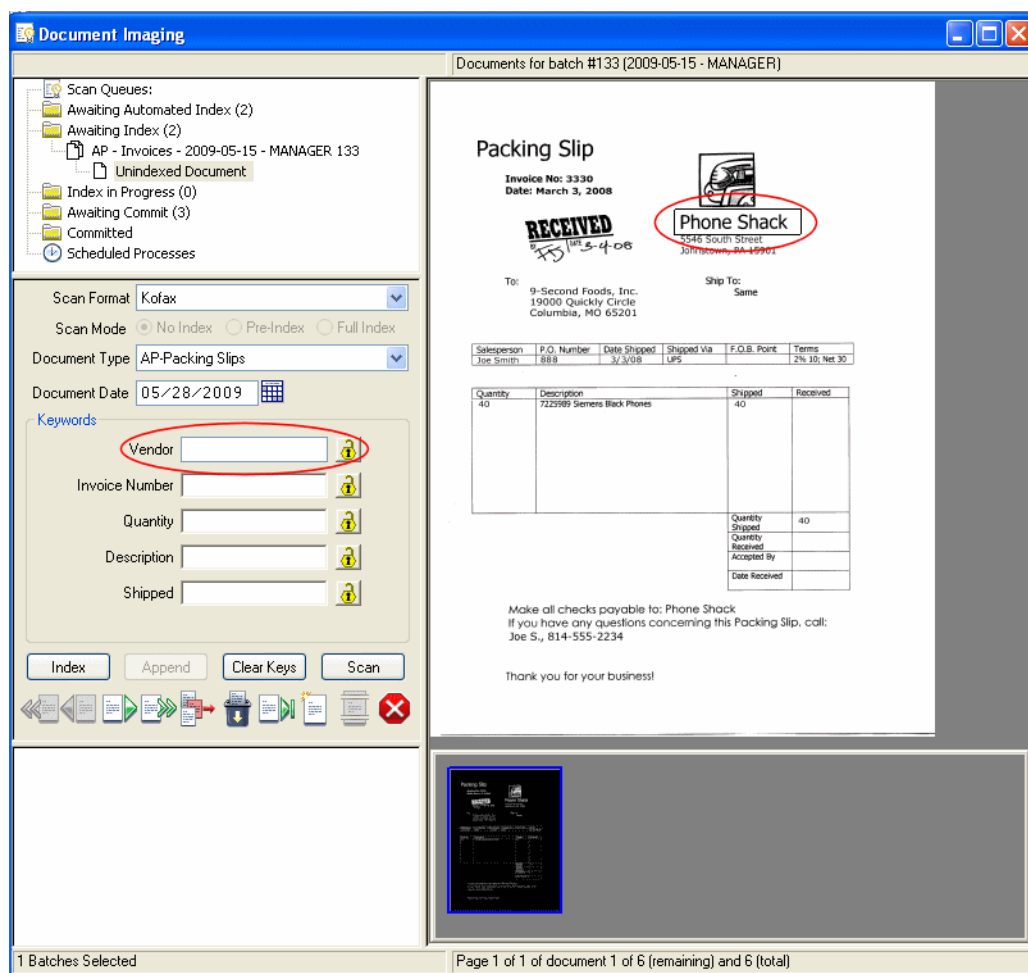
Thank you for your business!

1 Batches Selected

Page 1 of 1 of document 1 of 6 (remaining) and 6 (total)

- In the **Indexing** dialog box, set the focus to the field, either the Document Date field or one of the Keyword Type fields, that you would like to extract a value for.

- In the Working window, press and hold the **Shift** key and click and hold the left mouse button while using the pointer to draw a box around the value that is to be extracted.



- Release the **Shift** key and the left mouse button.

There may be a brief pause while the Automated Indexing engine processes the area you selected.

8. The value returned from the selected area is automatically populated in the selected field. Note that the focus is still set to the field after the value is populated.

Document Imaging

Documents for batch #133 (2009-05-15 - MANAGER)

Scan Queues:

- Awaiting Automated Index (2)
- Awaiting Index (2)
- AP - Invoices - 2009-05-15 - MANAGER 133
 - Unindexed Document
- Index in Progress (0)
- Awaiting Commit (3)
- Committed
- Scheduled Processes

Scan Format: Kofax

Scan Mode: ☒ No Index ☐ Pre-Index ☐ Full Index

Document Type: AP-Packing Slips

Document Date: 05/28/2009

Keywords

Vendor: **PHONE SHACK**

Invoice Number:

Quantity:

Description:

Shipped:

Index Append Clear Keys Scan

Packing Slip

Invoice No: 3330
Date: March 3, 2008

RECEIVED
AP-5-4-08

Phone Shack
5546 South Street
Johnstown, PA 15901

To: 9-Second Foods, Inc.
19000 Quickly Circle
Columbia, MO 65201

Ship To: Same

Salesperson	P.O. Number	Date Shipped	Shipped Via	F.O.B. Point	Terms
Joe Smith	888	3/3/08	LPS		2% 10, Net 30

Quantity	Description	Shipped	Received
40	722989 Siemens Black Phones	40	

Quantity Shipped	40
Quantity Received	
Accepted By	
Date Received	

Make all checks payable to: Phone Shack
If you have any questions concerning this Packing Slip, call:
Joe S., 814-555-2234

Thank you for your business!

1 Batches Selected

Page 1 of 1 of document 1 of 6 (remaining) and 6 (total)

Note: If you are using Point and Shoot Indexing to extract a Keyword Value that is the primary value for an AutoFill Keyword Set, be aware that the AutoFill Keyword Set is not triggered until you set the focus to another Keyword Type field. For additional information on using AutoFill Keyword Sets in conjunction with Point and Shoot Indexing, see AutoFill Keyword Sets on page 38.

9. Repeat Steps 5 through 8 for each value that you would like to extract for the currently-displayed document.

Tip: It is considered a best practice to always compare the values returned from the Automated Indexing engine to the actual values displayed on the document when performing Point and Shoot Indexing.

10. Once you are finished indexing the current document, click **Index**. The next document in the batch is displayed.

Repeat Steps 4 through 9 for the remaining documents in the batch. Once all documents in the batch have been indexed, the batch is automatically routed to the next batch status queue configured for the scan queue.

2

A Note About Single and Double-Byte Characters

If your Automated Indexing solution is configured to identify Latin characters and/or Arabic numerals (i.e., single-byte characters such as those found in English, French, German, etc.) and you wish to use Point and Shoot Indexing to identify Asian characters on an image, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the point to draw a box around the value to be extracted.

Likewise, if your Automated Indexing solution is configured to identify Asian characters (i.e., double-byte characters such as those found in Chinese and Japanese) and you wish to use Point and Shoot Indexing to identify Latin characters and/or Arabic numerals, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the point to draw a box around the value to be extracted.

The language, or languages, that are identified by your Automated Indexing solution are set when configuring an OCR format. Multiple OCR Formats can be created, but only one can be assigned per Document Type.

SYSTEM INTERACTION

The following OnBase features and modules interact with Automated Indexing.

AutoFill Keyword Sets

Although AutoFill Keyword Sets are part of the OnBase Client, and are not a separately-licensed OnBase module, give additional consideration when using Automated Indexing to index documents that are associated with an AutoFill Keyword Set.

Ad Hoc Automated Indexing

- If you are using Ad-Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:
 - If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document). This is the default behavior.
 - If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

Regardless of the setting of the **Expand All Matching Instances** option you are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document.

Batch Automated Indexing

- If you are using Batch Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:
 - If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, the batch is routed to the **Index in Progress** batch status queue so that you can select the instance, or instances, of the AutoFill Keyword Set to assign to the document. This is the default behavior.
 - If the **Expand All Matching Instances** option was selected the for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

Point and Shoot Indexing

- When using Point and Shoot Indexing, be aware that the focus remains on the field containing the just-extracted Keyword Value.

If you are using Point and Shoot Indexing to extract a Keyword Value that is the primary value for an AutoFill Keyword Set, the AutoFill Keyword Set is not triggered until you set the focus to another Keyword Type field.

Creating AutoFill Keyword Sets via Automated Indexing

Instances of AutoFill Keyword Sets can automatically be created using data applied to a document from a Batch Automated Indexing process if the associated scan queue has the **Store Keyword Set Data** option enabled. AutoFill Keyword Set data is only stored for documents that are indexed correctly and marked as fully-indexed; AutoFill Keyword Set data is not be stored for any document that is unidentified or has any values marked as suspect. AutoFill Keyword Set information cannot be stored for any document indexed via an Ad-Hoc Automated Indexing process because the document does not have a connection to a scan queue and, therefore, the **Store Keyword Set Data** scan queue option.

Automated Redaction

2

If your OnBase solution is also licensed for Automated Redaction, you can configure your Automated Indexing solution to mark Data Field Zones for Automated Redaction. Once the Data Field Zone has been processed and its Keyword Value captured and extracted, the zone is marked for redaction and the document is sent for Automated Redaction review and approval.

Automated Redaction is only available for Text Field Data Field Zones.

Automated Redaction is configured at the Data Field Zone level. This allows you to redact some, but not all, Keyword Values from a document, or a Keyword Value (e.g., a Social Security Number) from one Document Type but not another. Depending on the documents, you may be able to redact a Keyword Values from some documents in a Document Type, but not others (depending on the configuration of the Automated Indexing forms matched to the documents).

For more information on Automated Redaction, see the Automated Redaction documentation.

Document Imaging

Orientation Detection & Automatic Page Rotation

If the scanning workstation is licensed and registered for Batch Automated Indexing, it can be used to perform the Auto-Detect Orientation/Rotate Document Pages image process on batches of documents scanned via a scan queue configured to use image processing.

The Auto-Detect Orientation/Rotate Document Pages process uses the Automated Indexing engine to automatically detect if a page has been upside down or sideways and corrects them so that they are displayed at the correct orientation.

For more information on image processing and/or the Auto-Detect Orientation/Rotate Document Pages image process, see the Document Imaging documentation.

Appending Documents to Existing Documents

Depending on your scan queue configuration, documents that are indexed via Automated Indexing with the same Document Type and Keyword Values as existing OnBase documents can be appended to the existing documents.

See the Document Imaging documentation for more information on the **Append to Existing Documents Matched on Keyword(s)** scan queue setting.

EDM Services

The EDM Services module allows you to configure Document Types to allow revisions. Depending on your scan queue configuration, documents indexed via Automated Indexing can automatically be checked to determine if they are revisions of an existing document.

If a scan queue configured to perform Automated Indexing is also configured to use the **Check for Revisions** option on the Indexing tab of the **Assign Process Options for <Scan Queue Name>** dialog box, OnBase will check the Document Type and Keyword Values of the indexed document against existing documents. If the newly-indexed document is a match, then it may be treated as a revision (or rendition, if the original document is a non-image document and the Document Type is renditionable) of the existing document (see the Document Imaging documentation for more information on the behavior of the **Check for Revisions** setting).

However, because Automated Indexing is an automated process and no user interaction is needed to index documents, the behavior of documents identified as possible revisions of existing documents is different than if documents are manually indexed:

- When the scan queue's **Check for Revision** setting is used and more than one existing document with a matching Document Type and Keyword Values is found, the new document is routed to the **Awaiting Index** or **Index in Progress** queue, even though a Document Type has been selected for it and Keyword Values have been applied to it, where it can be reviewed and assigned as a revision of an existing document.
- When the scan queue's **Check for Revision** setting is used and a document indexed via Automated Indexing is determined to be a revision of an existing document and a revision comment is required for its Document Type, a revision comment is automatically generated indicating that the revision was identified via Automated Indexing.
- The scan queue's **Force Revision Prompt** setting is ignored when documents are indexed via Automated Indexing.

Intelligent Character Recognition (ICR)

Your Automated Indexing solution is able to identify and process machine-printed data via optical character recognition (OCR).

If your solution is also licensed for **ICR Support for Full-Page OCR & Automated Index**, your solution is also able to identify and process hand-written text via intelligent character recognition (ICR).

In order to recognize handwritten-text, the Data Field Zone must be configured to recognize hand-written text instead of machine-printed text.

Limitations of Intelligent Character Recognition

Automated Indexing has the following limitations when attempting to perform ICR:

- ICR cannot be performed using Point and Shoot Indexing.
- A single Data Field Zone cannot be configured to perform both OCR and ICR (i.e., both hand-written text and machine-printed text cannot be identified from with the same Data Field Zone).
- Only characters belonging to single byte character-based alphabets (e.g., English, French, Spanish, German, etc.) can be recognized via ICR processing. Characters belonging to double-byte character-based alphabets (e.g., Traditional Chinese, Japanese, Korean, etc.) cannot be recognized.
- ICR processing is available for only Text Field Data Field Zones. It is not available for any of the other types of Data Field Zones (Line Item Extraction, Group Line Item Extraction, etc.).

Improving the Accuracy of Intelligent Character Recognition

It is difficult to predict the accuracy of an ICR process due to the variability in the quality of the hand-written text. The following tips will help to improve the accuracy of ICR processing.

- The hand-written characters should be clearly separate from one another (i.e., they should not touch or intersect).
- The hand-written text should be unconstrained. If the characters are being entered into constraint boxes, you should configure your OCR process format to automatically drop them by selecting the **Remove vertical/horizontal lines from zone before processing** check box.

Constrained Text:**Name****Unconstrained Text:**

A handwritten signature 'John Adams' in black ink, written in a cursive style.

Name

- The ICR numeric text processor (**Handwriting/ICR - numerals/grouping punctuation (North American style)**, **Handwriting/ICR - numerals/grouping punctuation (European style)**) is significantly more accurate than the ICR alphanumeric text processor (**Handwriting/ICR - alphanumeric/punctuation**) because the character set is much smaller.

If you are attempting to process numeric data only from a Data Field Zone, you will have more accurate results if you use the numeric text processor.

Web Client

While batches that have undergone batch Automated Indexing can undergo additional indexing or verification in the OnBase Web Client, be aware that the index verification features available in the OnBase Client (e.g., suspect Keyword Values being highlighted in red, the Working window zooming in on the area of the document that the selected Keyword Value was extracted from, etc.) are not available in the OnBase Web Client.

Additionally, Ad-Hoc Automated Indexing cannot be initiated from the OnBase Web Client.

Workflow

Ad-Hoc Automated Indexing can be performed as an action within a Workflow queue if your Workflow Life Cycle is properly configured and your OnBase solution is licensed for both Workflow and Automated Indexing.

If a user performs ad-hoc Automated Indexing on a document from within a Workflow queue and an Automated Indexing or Ad-Hoc Automated Indexing license is registered on the workstation, a status bar is displayed and the document will undergo Automated Indexing.

The progress of the process is displayed in the status bar, and once the Automated Indexing process is complete, the status bar is closed. All configured indexing information has been extracted and assigned to the document.

If a user performs ad-hoc Automated Indexing on a document from within a Workflow queue and the workstation is not registered for Ad Hoc Automated Indexing or Automated Indexing, a message is displayed notifying him/her that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing.

Note: In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

Tip: Unlike batch Automated Indexing, ad-hoc Automated Indexing offers no indication for values that did not meet the configured Suspect Level. It is considered a best practice to always compare the indexing information assigned to the document by the ad-hoc Automated Indexing process to the actual information displayed on the document to ensure that it is accurate.

For more information, see the Workflow documentation.

Virtual Print Driver

Ad-Hoc Automated Indexing can be performed when printing and importing an image document created via the Virtual Print Driver if your OnBase solution is licensed for the Virtual Print Driver and either Automated Indexing or Ad-Hoc Automated Indexing, and your OnBase Client module is configured to use the Print Monitor.

When the Print Monitor detects that an image document has been printed by the Virtual Print Driver, the **Import from Virtual Printer** dialog box is displayed. Once the Document Type of the document has been selected via the **Document Type** drop-down, the **Queue Document for Automated Index** check box is displayed.

To perform ad-hoc Automated Indexing, select the **Queue Document for Automated Index** check box, enter any additional Keyword Values and click **Import**.

If the workstation is registered for Automated Indexing, the document undergoes ad-hoc Automated Indexing. The progress of the Automated Indexing process is displayed in the status bar.

Once the process is complete, the status bar is closed. All configured indexing information has been extracted and assigned to the document.

If the workstation is not registered for Ad-Hoc Automated Indexing or Automated Indexing, a message is displayed notifying the user that the document has been routed to the **Awaiting Ad Hoc Automated Index** batch status queue for processing.

Note: In order to see the **Awaiting Ad Hoc Automated Index** batch status queue, as well as process batches residing within it, the workstation must be registered for Automated Indexing.

Tip: Unlike batch Automated Indexing, ad-hoc Automated Indexing offers no indication for values that did not meet the configured Suspect Level. It is considered a best practice to always compare the indexing information assigned to the document by the ad-hoc Automated Indexing process to the actual information displayed on the document to ensure that it is accurate.

For more information, see the Virtual Print Driver documentation.

CONFIGURATION

The Automated Indexing module is easily configured. Prior to configuring Automated Indexing, you must configure Document Types to be able to undergo OCR processing. Once the Document Types have been configured, Automated Indexing configuration is performed in the OnBase Client using a point-and-shoot visual configuration interface. Point and Shoot Indexing requires no additional configuration.

3

CONFIGURING DOCUMENT TYPES

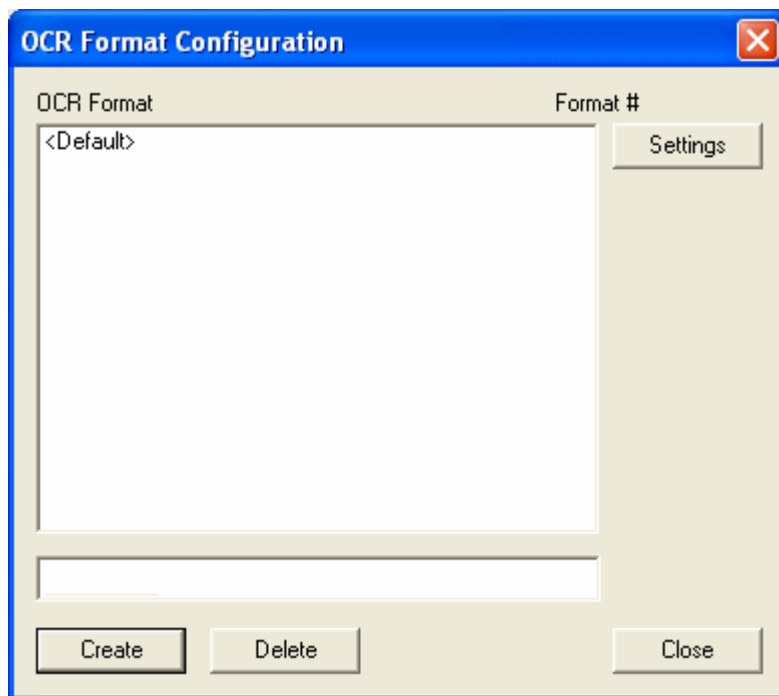
Prior to configuring Automated Indexing, you must create one or more OCR formats and then configure the Document Types that will undergo Automated Indexing for OCR processing.

Creating an OCR Format

An OCR format defines the parameters of the character recognition process. OCR formats are assigned at the Document Type level. Multiple OCR formats can be configured for your OnBase solution, but only one OCR format can be assigned per Document Type.

To create an OCR format:

1. From the OnBase Configuration module, click **Document | OCR Formats**. The **OCR Format Configuration** dialog box is displayed.



2. Enter a name for the OCR format that you are creating in the OCR Format field and click **Create**. The newly-created format is displayed in the OCR Format list.

3. Select the format from the OCR Format list and click **Settings**. The **OCR Format** dialog box is displayed.

3

Preprocessor Options

The options in the Preprocessor Options section allow you to modify the image file before it undergoes OCR processing.

Preprocessor Option	Description
No color conversion	Does not convert original colors. OCRing color images dramatically reduces the speed of the OCR engine. Average speed for color OCR processes is 13 seconds per page. This option does not modify the original image. The conversion is used only for the OCR process.

Preprocessor Option	Description
Convert color to grayscale	<p>Converts color images to grayscale images (composed of discrete shades of gray).</p> <p>This option does not modify the original image. The conversion is used only for the OCR process.</p>
Convert all to bitonal	<p>Converts images to bitonal images (images composed of a single background color and a single foreground color).</p> <p>This option does not modify the original image. The conversion is used only for the OCR process.</p>
Automatic inversion	<p>Inverts the entire image if the OCR engine perceives it to be inverted (having a black background and white foreground). Black pixels are converted to white. White pixels are converted to black.</p>
Never invert	<p>The image is never inverted.</p>
Always invert	<p>The image is always inverted.</p>
Automatically deskew	<p>Deskews the image if the OCR engine perceives it to be skewed (placed at an angle).</p>
Automatic rotation	<p>Text orientation is automatically detected on the page and rotated to the correct, upright rotation.</p> <hr/> <p>Note:If your solution is licensed for Automated Indexing, this setting is ignored when performing Point and Shoot Indexing. When performing Point and Shooting Indexing, the orientation of the page is always left as-is.</p> <hr/>
Never rotate	<p>Text orientation remains as in the original document.</p> <hr/> <p>Note:If your solution is licensed for Automated Indexing, this setting is ignored when performing Point and Shoot Indexing. When performing Point and Shooting Indexing, the orientation of the page is always left as-is.</p> <hr/>
Always rotate left 90 degrees	<p>Rotates the text orientation 90 degrees to the left.</p> <hr/> <p>Note:If your solution is licensed for Automated Indexing, this setting is ignored when performing Point and Shoot Indexing. When performing Point and Shooting Indexing, the orientation of the page is always left as-is.</p> <hr/>

Preprocessor Option	Description
Always rotate right 90 degrees	Rotates the text orientation 90 degrees to the right. Note: If your solution is licensed for Automated Indexing, this setting is ignored when performing Point and Shoot Indexing. When performing Point and Shooting Indexing, the orientation of the page is always left as-is.
Always rotate 180 degrees	Rotates the text orientation 180 degrees. Note: If your solution is licensed for Automated Indexing, this setting is ignored when performing Point and Shoot Indexing. When performing Point and Shooting Indexing, the orientation of the page is always left as-is.
Automatically despeckle	Automatically removes dots and speckles that commonly show up in faxed documents.
Automatically enhance resolution	Automatically adds pixels to enhance resolution if the OCR engine perceives it to be necessary.
Never enhance resolution	Pixels are never added to enhance resolution.
Always enhance resolution	Pixels are always added to enhance resolution.

Report per-page timing Statistics: Select this option to include detailed per-page OCR engine statistics in the Verification report.

Recognizer Options

The options in the Recognizer Options section allow you to specify the processing options used by the OCR engine.

Recognizer Option	Description
Automatic zoning	Slows process time. Most accurate page parser.
Standard zoning	Uses the same zoning as ScanSoft Omnipage 12.
Fast zoning	For fastest process times. Assumes few graphics and mostly text.
Legacy zoning	For faster processing times. This setting uses the original zoning first implemented with the OCR engine.
Nongridded table detect	Preserves nongridded formatting. Slows process speed.
Force single column	Converts multiple columns on a single page into a single column.

Recognizer Option	Description
Document-at-once-method	<p>Sends the document to the OCR engine all at once instead of page by page.</p> <hr/> <p>Note: This option must be deselected in order to OCR color images.</p> <hr/> <p>Tip: Select this option if you are getting an error when attempting to output to a searchable PDF format.</p> <hr/>
Direct output (faster)	<p>Select this option to speed up an OCR process that is configured to create an ASCII text (standard or formatted) output format.</p> <hr/> <p>Note: This option should not be used with OCR processes configured to use non-ASCII text output formats.</p> <hr/>
Most accurate	Most accurate recognition
Balance accuracy / speed	Mid level accuracy/speed recognition.
Fastest	Fastest recognition.

Tip: Due to the updated, more accurate OCR engine used in OnBase 7.2 and later, OCR processing times may be slightly longer than those experienced in previous versions of OnBase. To increase the speed of the process configured to use an ASCII text output format, select the **Direct output (faster)** check box. Or, to increase the speed of any OCR process, select the **Fast Zoning** and/or the **Fastest** radio buttons. Be aware, however, that the accuracy of the process may be reduced if the **Fast Zoning** and/or **Fastest** settings are selected.

Languages

These options represent the language(s) used for character recognition. Select the check box associated with each language you would like the OCR engine to consider when performing character recognition.

Filter Options

These options allow you to specify the type of characters that you would like the OCR engine to recognize. This can help the engine be more accurate and efficient by helping it to determine the type of data it is looking for.

Filter Option	Description
Uppercase	The OCR engine only recognizes uppercase characters.
Lowercase	The OCR engine only recognizes lowercase characters.
Digits	Only recognize 0-9.
Punctuation	The OCR engine only recognizes the following characters: (?!:,,"")
Misc.	The OCR engine only recognizes the following characters: (/ % # \$ + * @ etc.)

Spelling

Spelling Option	Description
Verify	Displays unknown or misspelled words in a bold font. Note: This option may not produce the expected results for all output file formats. For more information, contact your solution provider.
Correct	Note: This option is only available if the Verify check box is selected. Corrects words if a more accurate spelling is detected in the dictionary.

Spelling Option	Description
Vertical Dictionary	Enables the selection of a vertical dictionary from the drop-down list below. A vertical dictionary contains terms specific to an industry.

Codepage

A code page is a coded character set. For example, Windows ANSI (American National Standards Institute) is the Microsoft collective name for all Windows code pages. Select the appropriate Code Page to be used during the character recognition process.

Output format

Select the format to be used for the non-image rendition of the document that is created by the OCR process. The following are available:

:

Output Format	Additional Information
ASCII Text (Standard)	Non-formatted plain ASCII text.
ASCII Text (Formatted)	Plain ASCII text with formatting preserved.
PDF (Standard)	Normal Portable Document Format (PDF) document.
PDF (Image with Searchable text)	Original image overlayed on text.
PDF (Image Substitutes)	PDF includes embedded image snippets of text or image areas of the document that cannot be read by the OCR engine.
HTML 3.2	HTML compliant 3.2.
HTML 4.0	HTML compliant 4.0.
Microsoft Word 2003 (DOC)	Microsoft Word 2003 format (.doc)
Rich Text Format	Rich Text format.
Microsoft Word 2007 (DOCX)	Microsoft Word 2007 format (.docx)

CONVERTING PDF IMAGE ONLY TO SEARCHABLE PDFs

The **PDF Image Only** format is a scanned, non-searchable image of a document that is saved using the PDF file format and can be viewed using a PDF viewer. To convert **PDF Image Only** documents to searchable PDFs, sweep them into a scan queue, assign them to a Document Type configured to use an OCR Format with the **PDF (Image with Searchable Text)** output format and have the batch undergo OCR processing.

Because both the image-only PDF document and the searchable PDF document have a **PDF** file format, the searchable PDF document cannot be saved as a rendition of the image-only PDF (the image-only PDF is replaced by the searchable PDF document; the image-only PDF document is deleted from OnBase). However, if the Document Type is configured to be revisionable, then the searchable PDF document is saved as the latest revision of the original, image-only PDF document.

Note: In order to configure Document Types to be revisionable, additional licensing is required. See the EDMS documentation for more information.

Format Retention

Format retention is the level of formatting retained in the document. The **Format Retention** drop-down is active for all documents except ASCII Text (Standard) documents.

Retention Option	Description
True Page	Select the True Page option to keep the original layout of the pages.
Formatted Text	Select the Formatted Text option to retain the formatting information for fonts and paragraphs, but ignore layout formatting.
Plain Text	Select the Plain Text option to ignore all formatting information.
Flowing Page	Select the Flowing Page option to preserve the original layout of the pages, including column layouts.

Retention Option	Description
Respect lines per page	When this option is selected, if the number of lines of text on a page exceeds the Document Type's Lines per page setting, the excess lines are displayed in the right margin of the current page instead of being moved to the next page of the text document. This ensures a one-to-one correspondence between information stored on the page of the image and text documents. If this option is not selected, text documents adhere to the Lines per page setting, pushing the excess lines of text to the next page of the text document.
Do not OCR PDF documents	When selected, PDFs in batches are skipped by the OCR process.
Create PDF/A compatible output	When a PDF output format is selected, select this check box to create PDF documents that conform to the PDF/A standard.

3

Configuring Document Types for OCR

In order to configure a Document Type to be able to undergo Automated Indexing, you must configure the following Document Type options.

Configuring a Document Type to Use Renditions

Renditions are versions of a document that are generated in a file format that differs from the original. Documents that undergo OCR processing are originally images or PDFs, and the resulting file format is text-based (e.g., text, rich format text, Microsoft Word, etc.).

Documents that undergo OCR processing must belong to a Document Type that is renditionable in order for the OCR engine to process them into text-based documents.

To configure a Document Type to be renditionable:

1. From the OnBase Configuration module, select **Document | Document Types**.
2. Select the Document Type to be configured and click **Rendition/Revision**. The **Rendition/Revision Control Settings** dialog box is displayed.
3. Select the **Allow Multiple Renditions** option.
4. Click **Save & Close**.

Assigning an OCR Format to a Document Type

You must assign an OCR Format to the Document Type in order to instruct the OCR engine how to process documents.

1. From the OnBase Configuration module, click **Document | Document Types**.
2. Select the Document Type to be configured and click **Settings**. The **<Document Type Name> Configuration** dialog box is displayed.
3. Using the **OCR Format** drop-down, select the OCR format that the Document Type should use.
4. Click **Save**.

3

CONFIGURING POINT AND SHOOT INDEXING

Point and Shoot Indexing can be performed on image documents displayed in the Document Imaging Working window. It does not require any special configuration.

A Note About Single and Double-Byte Characters

If your Automated Indexing solution is configured to identify Latin characters and/or Arabic numerals (i.e., single-byte characters such as those found in English, French, German, etc.) and you wish to use Point and Shoot Indexing to identify Asian characters on an image, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the pointer to draw a box around the value to be extracted.

Likewise, if your Automated Indexing solution is configured to identify Asian characters (i.e., double-byte characters such as those found in Chinese and Japanese) and you wish to use Point and Shoot Indexing to identify Latin characters and/or Arabic numerals, you must press and hold **Ctrl+Shift** and then click and hold the left mouse button while using the point to draw a box around the value to be extracted.

CONFIGURING AUTOMATED INDEXING

In order to configure Automated Indexing, you must perform the following configuration steps:

1. **Give users rights to configure Automated Indexing forms.** See Granting Users the Rights to Configure Automated Indexing Forms on page 61 for more information.
2. **Configure Automated Indexing Forms for Documents in the Batch.** See Configuring Automated Indexing Forms on page 64 for more information.

3. **Test the Automated Indexing forms.** See Testing Automated Indexing Forms on page 208 for more information.
4. **Configure scan queues to route batches to the Awaiting Automated Indexing batch status queue.** See Configuring a Scan Queue for Automated Indexing on page 215 for more information.
5. **Configure batches to automatically undergo Automated Indexing.** See Configuring Automatic Automated Indexing Processes on page 220 for more information on this optional configuration step.

Note: Beginning in OnBase 10.0, access to batches of documents within the Document Imaging interface is no longer needed to configure Automated Indexing forms; forms can be configured either from the OnBase Client or from within the Document Imaging interface. See For information on importing batches of documents into a scan queue, consult the documentation of the module you are using to import the documents into OnBase. on page 65 for more information.

3

Configuring Automated Indexing Forms for Institutional Databases

If your OnBase solution is configured to use Institutional Databases, Automated Indexing forms can be created at the institution level or at a global level.

- Institution-level forms are available only to the institution that they were created for.
- Global forms are available to all institutions that are part of your OnBase solution.

Non-super users can only create and configure Automated Indexing forms for their own institutions.

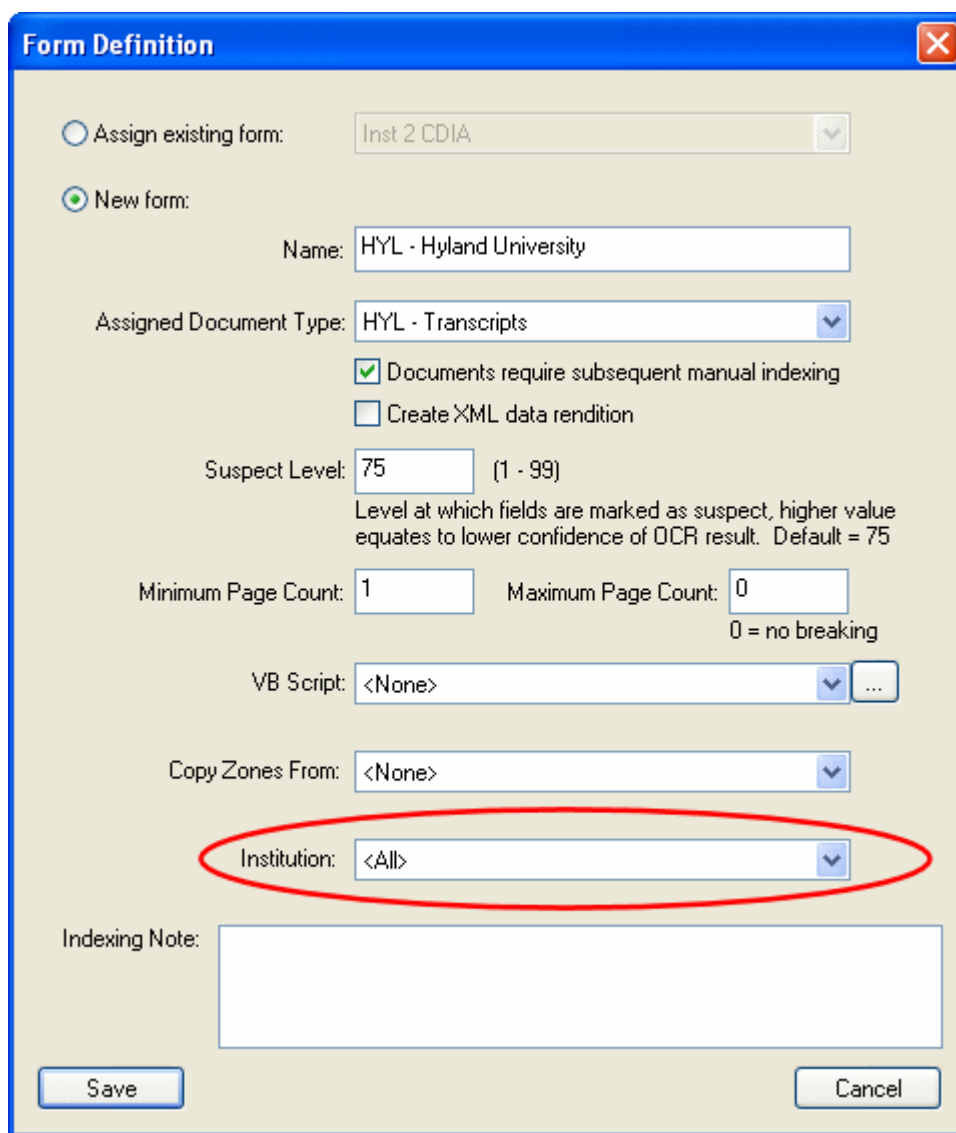
Super users can configure Automated Indexing forms for any institution, provided that the institutions using the forms have access to the Document Types and Keyword Types associated with the form. Super users may also configure global Automated Indexing forms.

Tip: It is considered a best practice for super users to create only global Automated Indexing forms. Non-super users should create all institution-level Automated Indexing forms.

Automated Indexing forms are assigned to institutions using the **Institution** drop-down select list on the **Form Definition** dialog box.

- For non-super users, the current institution is pre-selected in the **Institution** drop-down select list and the **Institution** drop-down select list is disabled.
- For super users, the **Institution** drop-down select list is enabled, allowing them to select any institution or the **<All>** option (used to create global Automated Indexing forms).

3



The image shows a 'Form Definition' dialog box with a blue title bar and a close button. It contains two radio buttons: 'Assign existing form' (selected) and 'New form'. The 'Assign existing form' dropdown shows 'Inst 2 CDIA'. The 'New form' section includes a 'Name' field with 'HYL - Hyland University', an 'Assigned Document Type' dropdown with 'HYL - Transcripts', and two checkboxes: 'Documents require subsequent manual indexing' (checked) and 'Create XML data rendition' (unchecked). The 'Suspect Level' is a text box with '75' and a range '(1 - 99)', with a note below it. The 'Minimum Page Count' is '1' and the 'Maximum Page Count' is '0', with a note '0 = no breaking'. The 'VB Script' dropdown is '<None>' with a browse button. The 'Copy Zones From' dropdown is '<None>'. The 'Institution' dropdown is '<All>' and is circled in red. There is an 'Indexing Note' text area at the bottom. 'Save' and 'Cancel' buttons are at the bottom.

Form Definition

☐ Assign existing form: Inst 2 CDIA

☒ New form:

Name: HYL - Hyland University

Assigned Document Type: HYL - Transcripts

☒ Documents require subsequent manual indexing

☐ Create XML data rendition

Suspect Level: 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: 1 Maximum Page Count: 0
0 = no breaking

VB Script: <None> ...

Copy Zones From: <None>

Institution: <All>

Indexing Note:

Save Cancel

Using Automated Indexing Forms with Institutional Databases

Be aware that documents that are to undergo Automated Indexing must be pre-indexed with their **Institution #** Keyword Value.

Note: Documents cannot be swept into OnBase in pre-index mode. If you are sweeping in batches of documents that you want to undergo Automated Indexing, the **Institution #** Keyword Value must be added to the documents after they are swept but before they undergo Automated Indexing (e.g., via a batch bar code process using the Barcode Recognition Server, via the **Add/Modify Keywords** dialog box before the batch undergoes Automated Indexing processing, etc.).

TESTING AUTOMATED INDEXING FORMS

When testing Automated Indexing form configuration in the **Automated Index Configuration** window, the documents being processed are treated as if they belong to the current institution, even if they have not yet been assigned to that institution.

Tip: When a super user is performing a test Automated Indexing process (i.e., after a user has clicked **Process Current Document** or **Process Batch** in the Tools window), ensure that the current institution is the institution you want to be testing the form(s) for.

During the test Automated Indexing process, information about the institution is displayed in the Results Verification window.

PERFORMING AUTOMATED INDEXING PROCESSES

When a batch undergoes Automated Indexing, the OCR engine only attempt to match documents in the batch against Automated Indexing forms available to the documents' assigned Document Type and Automated Indexing forms available to the current institution.

During processing, information about the current institution is displayed in the Indexing Status window of the **Document Imaging** window.

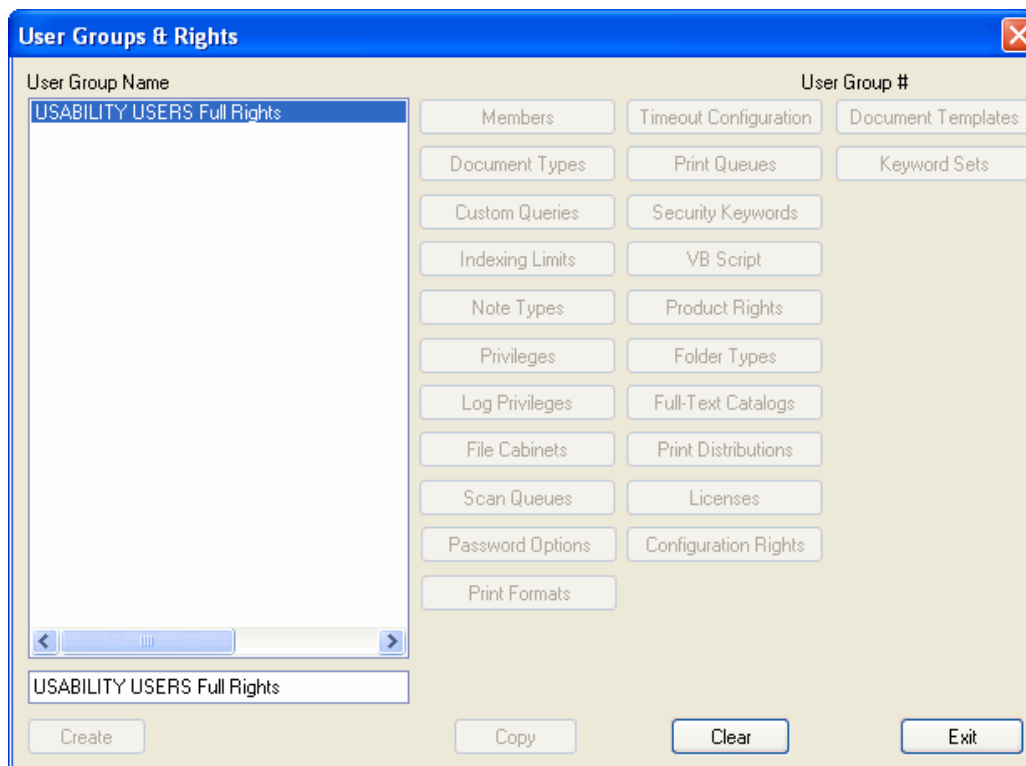
Granting Users the Rights to Configure Automated Indexing Forms

In order to create, modify, or delete Automated Indexing forms and associate Automated Indexing forms to existing documents, a user must belong to a User Group that has been given the Document Imaging administrative processing privilege.

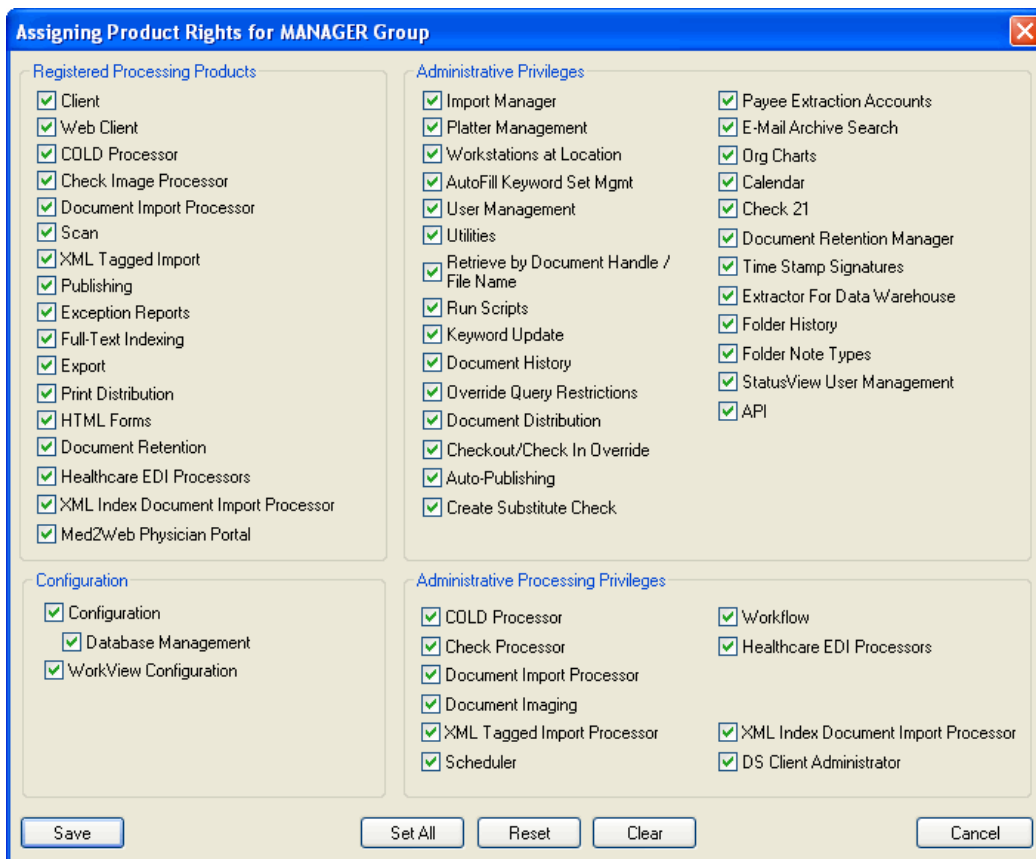
Caution: Careful consideration should be taken when granting or restricting privileges or product rights to/from a User Group. For more information on User Group configuration, see the System Administration documentation.

To assign the Document Imaging administrative processing privilege to a User Group:

1. From the OnBase Configuration module, click **Users | User Groups /Rights**. The **User Groups & Rights** dialog box is displayed.



2. Select the User Group that is to be granted the Document Imaging administrative processing privilege from the User Group Name list and click **Product Rights**. The **Assigning Product Rights for <User Group Name> Group** dialog box is displayed.

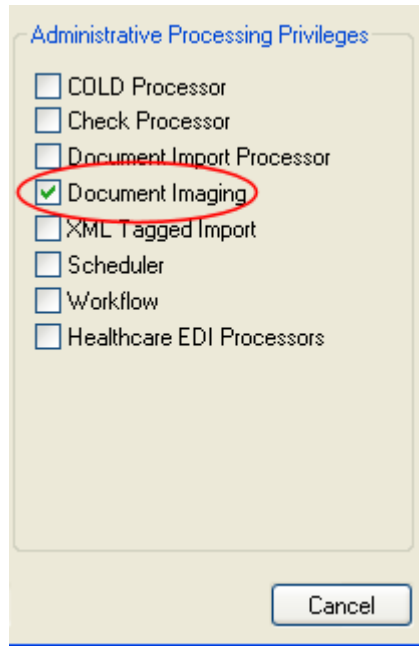


The dialog box titled "Assigning Product Rights for MANAGER Group" contains four sections of checkboxes, all of which are checked:

- Registered Processing Products:**
 - Client
 - Web Client
 - COLD Processor
 - Check Image Processor
 - Document Import Processor
 - Scan
 - XML Tagged Import
 - Publishing
 - Exception Reports
 - Full-Text Indexing
 - Export
 - Print Distribution
 - HTML Forms
 - Document Retention
 - Healthcare EDI Processors
 - XML Index Document Import Processor
 - MedZWeb Physician Portal
- Administrative Privileges:**
 - Import Manager
 - Platter Management
 - Workstations at Location
 - AutoFill Keyword Set Mgmt
 - User Management
 - Utilities
 - Retrieve by Document Handle / File Name
 - Run Scripts
 - Keyword Update
 - Document History
 - Override Query Restrictions
 - Document Distribution
 - Checkout/Check In Override
 - Auto-Publishing
 - Create Substitute Check
 - Payee Extraction Accounts
 - E-Mail Archive Search
 - Org Charts
 - Calendar
 - Check 21
 - Document Retention Manager
 - Time Stamp Signatures
 - Extractor For Data Warehouse
 - Folder History
 - Folder Note Types
 - StatusView User Management
 - API
- Configuration:**
 - Configuration
 - Database Management
 - WorkView Configuration
- Administrative Processing Privileges:**
 - COLD Processor
 - Check Processor
 - Document Import Processor
 - Document Imaging
 - XML Tagged Import Processor
 - Scheduler
 - Workflow
 - Healthcare EDI Processors
 - XML Index Document Import Processor
 - DS Client Administrator

At the bottom of the dialog box are five buttons: Save, Set All, Reset, Clear, and Cancel.

3. Ensure that the **Document Imaging** check box is selected in the Administrative Processing Privileges section.



4. Click **Save**.

Configuring Automated Indexing Forms

Automated Indexing forms are the pre-configured templates that are matched to incoming documents to determine how the documents will be classified and indexed.

In order to configure Automated Indexing forms (e.g., create, modify or delete Automated Indexing forms or manually associate Automated Indexing forms to documents in a batch):

- You must belong to a User Group with the **Document Imaging** Administrative Processing Privilege Product Right (see Granting Users the Rights to Configure Automated Indexing Forms on page 61 for more information).
- Your OnBase solution and workstation must be licensed and registered for Automated Indexing or Ad-Hoc Automated Indexing.

Accessing the Automated Index Configuration Window

Automated Indexing forms are created and configured in the **Automated Index Configuration** window.

The **Automated Index Configuration** window can be accessed in two ways:

- **From the Client.** The Automated Index Configuration window can be opened from the Client's main menu (outside of the Document Imaging window). This allows you to create Automated Indexing forms from documents currently stored by your Automated Indexing solution.
- **From the Document Imaging Interface.** The Automated Index Configuration window can be opened from the Document Imaging window. This allows users to create forms from documents that have been scanned and currently reside in the batch status queue.

Note: For information on importing batches of documents into a scan queue, consult the documentation of the module you are using to import the documents into OnBase.

OPENING THE AUTOMATED INDEX CONFIGURATION WINDOW FROM THE CLIENT

This option exists to allow you to create Automated Indexing forms from documents that currently available in the Client. Prior to creating Automated Indexing forms using this method, ensure that you have one or more documents that can be used as a basis for creating Automated Indexing forms.

1. From the Client, click **Admin | Automated Index Configuration**. The **Automated Index Configuration Documents** dialog box is displayed.

Automated Index Configuration Documents

☐ Limit document search results to only those document(s) that failed automated index recognition

☒ Maximum number of document(s) returned in result set: 10

Document Type Groups

Student Information

Document Types

STDNT - Transcripts

Dates

January 2011
December 2010
November 2010
October 2010
September 2010
August 2010
July 2010

From / /

To / /

Keywords

Find Clear All Clear Keywords Cancel

2. Using the Document Type Group and Document Type lists, select the Document Type that the document you want to use as a basis for the Automated Indexing form belongs to.

Once a Document Type is selected, the Keyword Type fields associated with the selected Document Type are displayed in the Keywords section.

3. Using the Date Range fields (i.e., **From** and **To**) and the Keyword Type fields, enter a date range and/or Keyword Values to retrieve the document you want to use to create Automated Indexing form.

Tip: To limit the documents returned by the search to only those documents that have already been processed and were not successfully matched to a Automated Indexing form, select the **Limit document search results to only those document(s) that failed automated index recognition** check box.

4. To configure a maximum number of documents that can be returned by the search, select the **Maximum number of document(s) returned in result set** check box and enter the maximum number of documents allowed to be returned in the associated field.
5. Click **Find** to perform the search.

All documents returned by the search can be used to configure Automated Indexing forms in the **Automated Index Configuration** window.

OPENING THE AUTOMATED INDEX CONFIGURATION WINDOW FROM THE DOCUMENT IMAGING WINDOW

This option exists to allow you to create Automated Indexing forms from documents in a batch. Prior to creating Automated Indexing forms using this method, ensure that you have scanned a batch of documents that can be used as a basis for creating Automated Indexing forms.

1. From the **Document Imaging** window, in the Queue List window, select the batch status queue (e.g., **Awaiting Automated Index**, **Awaiting Commit**, **Committed**, etc.) containing the batch of documents that are to be used as a basis for creating Automated Indexing forms.
2. All batches residing in the batch status queue are displayed in the Working window. Right-click on the desired batch in the Working window and select **Configure Automated Index**.

All documents in the batch are available for Automated Indexing form configuration in the **Automated Index Configuration** window.

THE AUTOMATED INDEX CONFIGURATION WINDOW

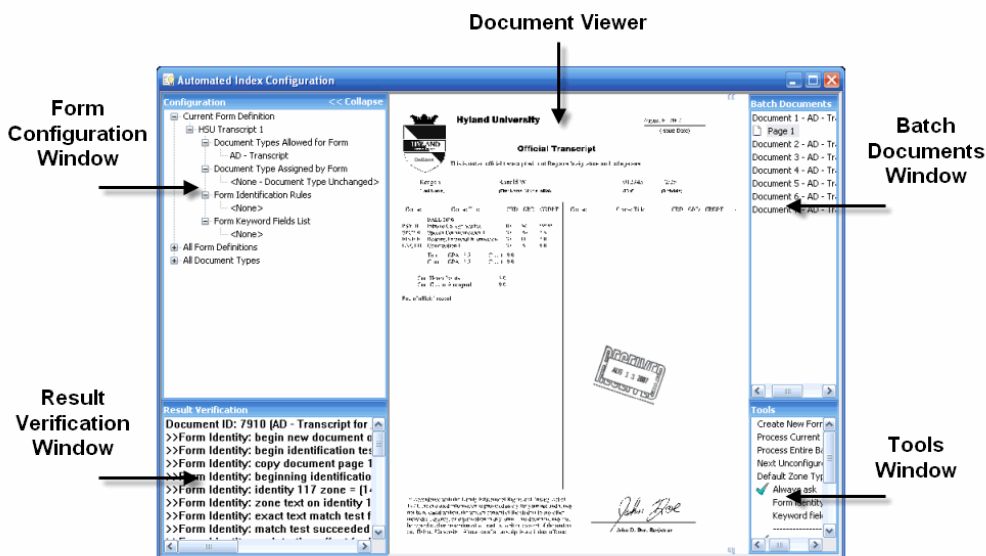
Before the **Automated Index Configuration** window is displayed, you are prompted with a message asking if you would like to pre-process the documents in the selected batch to attempt to match them to an existing Automated Indexing form.

- Click **Yes** to pre-process the documents.
 - If pre-processing results in documents in the batch being matched to Automated Indexing forms, then you will have the chance to modify the forms from within the **Automated Index Configuration** window.
 - If pre-processing fails to match Automated Indexing forms to documents in the batch, if pre-processing is cancelled before a document in the batch is processed, or if pre-processing is skipped altogether, then the documents in the batch are unclassified and you will be able to configure new forms for the documents' layouts.
- Click **No** to skip pre-processing

Note: If pre-processing is skipped or cancelled before a document is pre-processed, that document is as unclassified, even if a matching Automated Indexing form exists, because the OCR engine was not able to attempt to match the document to the matching Automated Indexing form.

Tip: Documents can always be processed, or re-processed, later from within the **Automated Index Configuration** window by clicking **Process Entire Batch** in the **Tools** window.

The **Automated Index Configuration** window is displayed.



Window	Description
Form Configuration Window	The Form Configuration window displays all form configuration information in a tree view.
Result Verification Window	The Result Verification window displays information on the fields processed and the values read by the OCR engine. Tip: Double-clicking on the Result Verification window exports the contents of the Result Verification window to a new text document. This does not export the preview image displayed when configuring a Form Identification Zone or Data Field Zone. If the Result Verification window is empty, a new text document is not opened.

Window	Description
Document Viewer	<p>The Document Viewer displays the currently-selected document with the Automated Indexing form applied to it.</p> <p>Form Identification Zones are highlighted in magenta, Data Field Zones are highlighted in green and Page Registration Zones are highlighted in yellow.</p>
Batch Documents Window	<p>The Document Select window displays a list of all documents available for Automated Indexing form configuration (i.e., depending on how the Automated Index Configuration window was accessed, either the documents returned by the retrieval operation or the documents in the current batch).</p> <p>A green check mark is displayed next to documents that have been matched to Automated Indexing forms and a red X is displayed next to documents that are unclassified.</p> <p>Selecting a document opens it in the Document Viewer.</p>

Window	Description
Tools Window	<p>The Tools window contains links to all of the tools necessary to configure Automated Indexing forms.</p> <p>The available tools are:</p> <ul style="list-style-type: none"> • Create a New Form Definition. Click to open the Form Definition dialog box. • Process Current Document. Click to run a test Automated Indexing process on the document currently displayed in the Document Viewer. • Process Entire Batch. Click to run a test Automated Indexing process on all documents listed in the Batch Documents window. • Next Unconfigured Document. Click to display the next document that has not yet been matched to a form. • Default New Template/Zone Settings. Click one of the Default Zone Type sub-options to set the default behavior when a new Automated Indexing form zone is created. <p>Assign Current Scan Queue Only. <i>This option is selected by default.</i> When selected, this option makes any Automated Indexing forms you create available for only the scan queue that the currently-displayed document is associated with and for ad-hoc Automated Indexing processes.</p> <p>When not selected, any Automated Indexing forms you create are automatically made available to all scan queues.</p> <p>Always Ask New Zone Type. Select this option to always prompt the user to select the type of zone being created.</p> <p>Form identity or registration. Select this option to automatically set each newly-created zone as a Form Identity or Registration Zone.</p> <p>Keyword field. Select this option to automatically set each newly-created zone as a Data Field Zone. When this option is selecting, you can also specify the type of Data Field Zone being configured by selecting one of the following options: Keyword Value Field, Line Item Extraction, Group Line Item Extraction, Optical Mark Field, or Signature Detection Field.</p> <ul style="list-style-type: none"> • Import/Export Configuration. Click to import or export Automated Indexing forms from/to another Automated Indexing solution.

Window	Description
Tools Window (cont.)	<ul style="list-style-type: none"> • Global Keyword Lookup/Replace. Click to configure a global Keyword Lookup/Replace dictionary to replace Keyword Values identified by the Automated Indexing process with different values. For example, if you require a Term Keyword Value to contain a Season and a Four Digit Year (e.g., Fall 2010) and some incoming documents display a Season and a Two Digit Year (e.g., Fall 10), you could configure a lookup/replace action to automatically replace all Keywords Values identified as Fall 10 with Fall 2010. • Analyze Form Template Precedence. Click to perform an analysis of how frequently your forms are being matched to documents. Your forms can be re-ordered based on the results, allowing you to increase the efficiency of your indexing processes. • Save Configuration. Click to save the current configuration, including any configured Automated Indexing forms. • Discard Configuration Changes. Click to discard any unsaved Automated Indexing form changes and revert to the last saved form configuration. • Exit Configuration. Click to close the Automated Index Configuration window and return to the Document Imaging window.

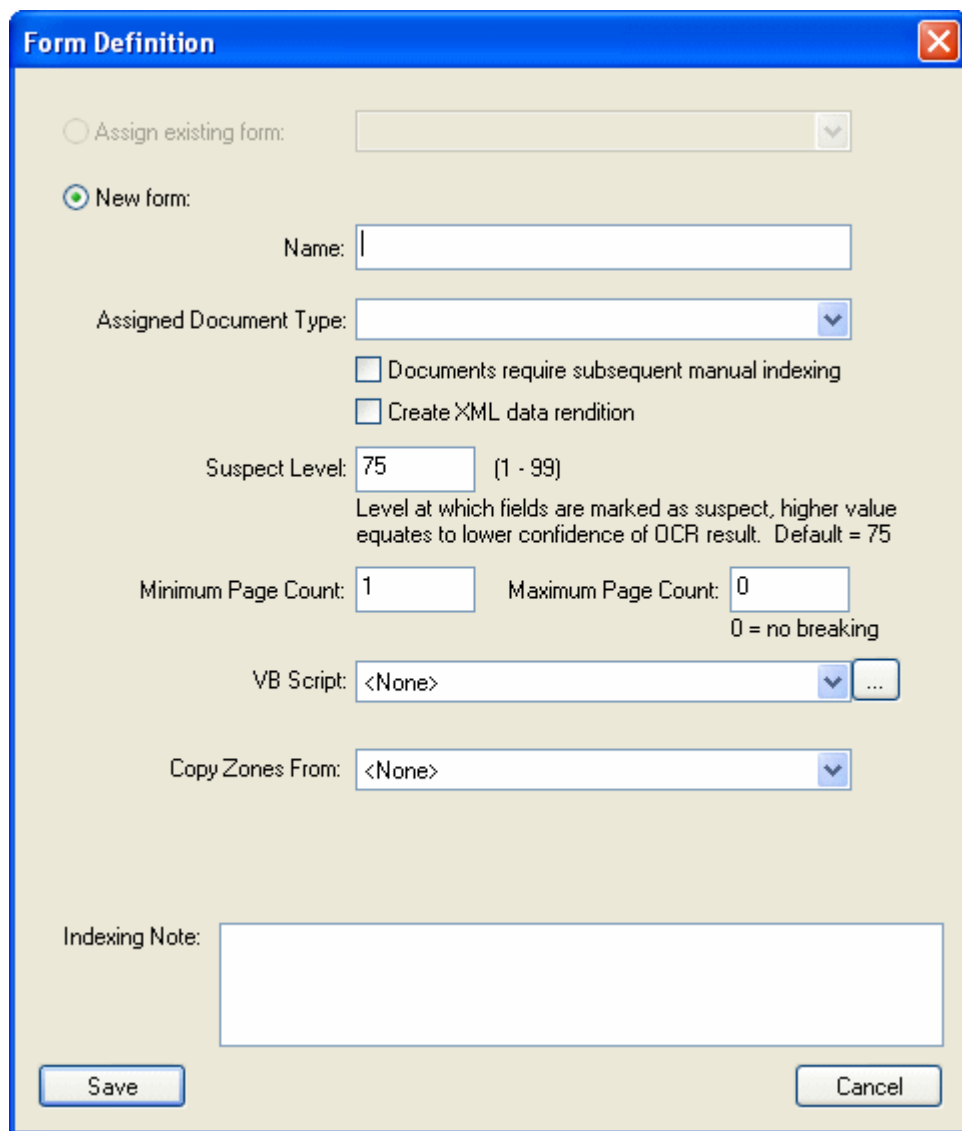
Once the Automated Index Configuration window is displayed, the first page (depending on how you accessed the Automated Index Configuration window, either the first page in the batch or the first page of the first document returned by your query) is displayed in the Document Viewer and the **Form Definition** dialog box is automatically displayed.

- To select a different page or document, click the proper page/document from the **Batch Documents** window.
- If the **Form Definition** dialog box is not automatically displayed, click **Create New Form Definition** in the Tools window.

Creating a New Automated Indexing Form

To create a new Automated Indexing form:

1. From the **Form Definition** dialog box, click the **New Form** radio button to enable the new Automated Indexing form options on the dialog box.



The **Form Definition** dialog box is shown with the **New form:** radio button selected. The **Name:** field is empty. The **Assigned Document Type:** dropdown is set to **<None>**. The **Documents require subsequent manual indexing** and **Create XML data rendition** checkboxes are unchecked. The **Suspect Level:** is set to **75** (range 1 - 99). The **Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75** is displayed below. The **Minimum Page Count:** is **1** and the **Maximum Page Count:** is **0** (0 = no breaking). The **VB Script:** dropdown is set to **<None>** with a **...** button. The **Copy Zones From:** dropdown is set to **<None>**. The **Indexing Note:** text area is empty. The **Save** and **Cancel** buttons are at the bottom.

Form Definition

☐ Assign existing form:

☒ New form:

Name:

Assigned Document Type:

☐ Documents require subsequent manual indexing

☐ Create XML data rendition

Suspect Level: (1 - 99)

Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: Maximum Page Count:

0 = no breaking

VB Script:

Copy Zones From:

Indexing Note:

Option	Description
Assign existing form/ New form	<p>Select the New form radio button to create a new Automated Indexing form for the document currently displayed in the Document Viewer.</p> <p>Once selected, the remaining options in the dialog box, other than the Assign existing form option, are enabled.</p> <hr/> <p>Note:The Assign existing form radio button is used to assign an existing Automated Indexing form to a document. See <i>Manually Matching Existing Automated Indexing Forms to Documents in the Batch</i> on page 76 for more information.</p> <hr/>
Name	Enter the name of the Automated Indexing form that you are creating.
Assigned Document Type	<p>Select the Document Type that documents will be assigned to once they are associated with this Automated Indexing form.</p> <p>To leave the Document Type of documents undergoing Automated Indexing processing unchanged (e.g., to ensure that the document remains unindexed or that the document still belongs to the Document Type it was assigned to at the time of import), select <Document Type Unchanged>.</p>
Documents require subsequent manual indexing	Select this check box to automatically route documents matched to this Automated Indexing form for index verification and review after processing.
Create XML data rendition	<p>Select this check box to create XML renditions of the documents matched to this Automated Indexing form.</p> <p>The XML renditions contain indexing values (e.g., the Document Date and Keyword Values) identified during the Automated Indexing process. The structure of the XML file (i.e., the element that the Keyword Value is contained in) is defined when configuring the Data Field Zone for that Keyword Value.</p>

Option	Description
Suspect Level	<p>Enter the Suspect Level threshold, 1-99, in this field.</p> <p>The Suspect Level is the level of confidence placed in the data values captured using this form.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold causes the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Suspect Levels can also be set at the zone level for individual zones if more or less stringent standards are needed for their values.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p> <hr/>
Minimum Page Count	<p>The Minimum Page Count field contains the minimum number of pages that can compose a document identified by this Automated Indexing form.</p> <p>This setting must be set to at least a value of 1.</p>
Maximum Page Count	<p>The Maximum Page Count field contains the maximum number of pages that can compose a document identified by this Automated Indexing form.</p> <p>Enter 0 if the Automated Indexing process should not attempt to break documents (i.e., Automated Indexing respects the pre-existing document breaks).</p>

Option	Description
VB Script	Use the VB Script drop-down select list to select a VB script to associate with the identification and processing of this Automated Indexing form. Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.
Copy Zones From	If you would like to copy all of the configured zones from an existing Automated Indexing form, select the existing form from this drop-down select list.
Indexing Note	Enter an indexing note of up to 250 character to be displayed for documents matched to this Automated Indexing form. This note is displayed to users when a document matched to this Automated Indexing form is displayed in the Working window during index verification and review.

- Set the configuration options for the Automated Indexing form being created.
- Click **Save** to save the form and close the **Form Definition** dialog box.

Tip: At any time, click **Save Configuration** in the Tools window to save the Automated Indexing form. It is considered a best practice to save your Automated Indexing form after the form is created and after any configuration change is made (e.g., any Form Identification, Registration, or Data Field Zones are added, deleted, modified, moved, etc.).

To discard all changes since the last time the form was saved, click **Discard Configuration Changes** in the Tools window.

MANUALLY MATCHING EXISTING AUTOMATED INDEXING FORMS TO DOCUMENTS IN THE BATCH

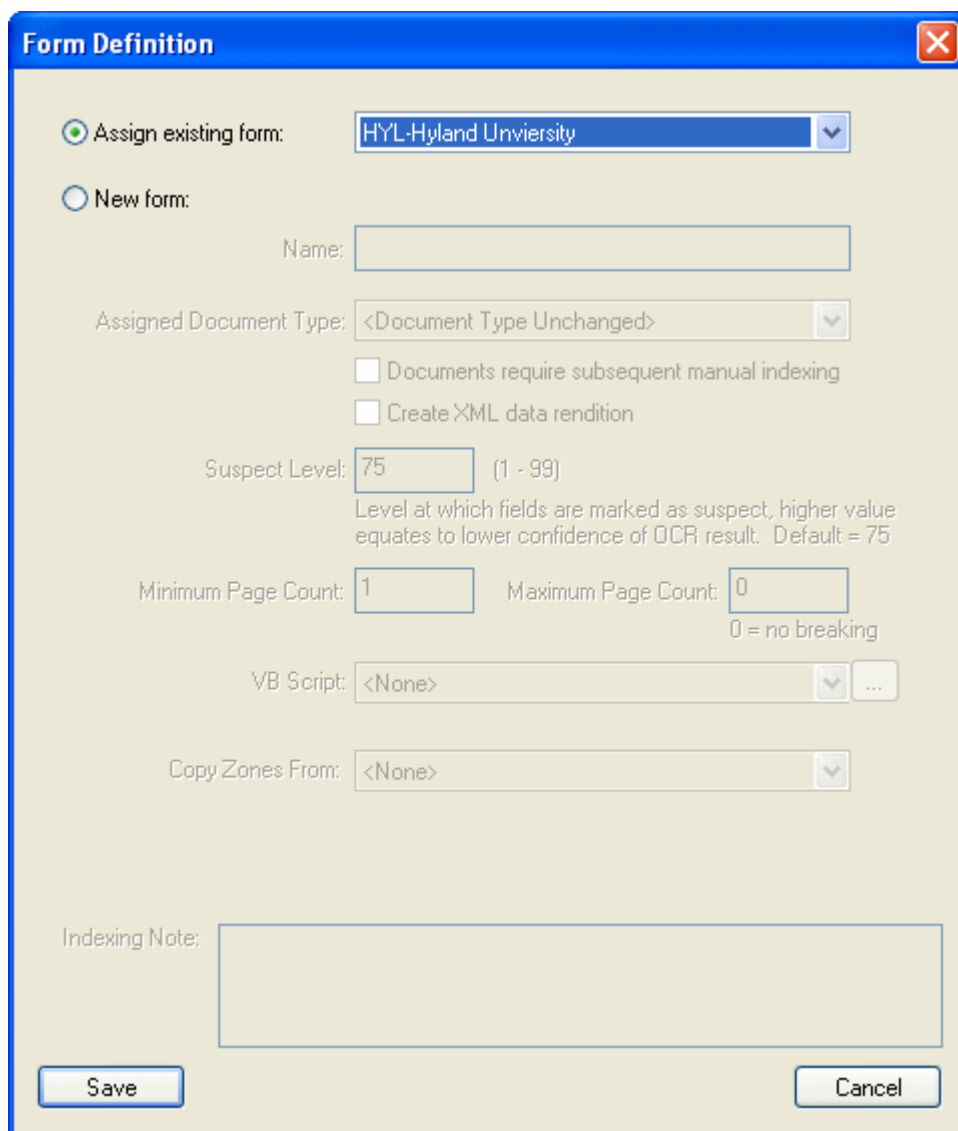
When a batch is processed in the **Awaiting Automated Indexing** batch status queue, the OCR engine automatically evaluates the documents in the batch against all available Automated Indexing forms.

Sometimes, however, it is necessary to manually associate a form to a document in the **Automated Index Configuration** window (such as when the text or image used as a form identifier is of poor quality for a scanned document or if a document's layout has been changed and its Automated Indexing form needs to be updated to match it).

To manually associate a form to a document:

1. From the **Automated Index Configuration** window, ensure that the document the Automated Indexing form is being matched to is displayed in the Document Viewer. If needed, select the correct document from the **Batch Documents** window.
2. The **Form Definition** dialog box is automatically displayed. If it is not displayed, click **Create New Form Definition** in the Tools window.

3



The image shows a 'Form Definition' dialog box with a blue title bar and a close button. It contains two radio buttons: 'Assign existing form' (selected) and 'New form'. The 'Assign existing form' option has a dropdown menu showing 'HYL-Hyland Unviersity'. The 'New form' option has a 'Name' text field. Below these are 'Assigned Document Type' (dropdown, '<Document Type Unchanged>'), two checkboxes for 'Documents require subsequent manual indexing' and 'Create XML data rendition', and a 'Suspect Level' (text field '75' with a range '(1 - 99)' and a description: 'Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75'). There are also 'Minimum Page Count' (text field '1') and 'Maximum Page Count' (text field '0' with a note '0 = no breaking'). Below these are 'VB Script' (dropdown '<None>' with an ellipsis button) and 'Copy Zones From' (dropdown '<None>'). At the bottom is an 'Indexing Note' text area. 'Save' and 'Cancel' buttons are at the bottom.

Form Definition

☒ Assign existing form: HYL-Hyland Unviersity

☐ New form:

Name:

Assigned Document Type: <Document Type Unchanged>

☐ Documents require subsequent manual indexing

☐ Create XML data rendition

Suspect Level: 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: 1 Maximum Page Count: 0
0 = no breaking

VB Script: <None> ...

Copy Zones From: <None>

Indexing Note:

Save Cancel

3. Select the **Assign existing form** radio button. Once selected, use the drop-down to select the Automated Indexing form to be associated with the document.

Note: The **Assign existing form** radio button and the associated drop-down menu are only enabled when at least one Automated Indexing form is configured for your Automated Indexing solution.

If no Automated Indexing form exists that matches the currently-displayed document, you will need to configure a new Automated Indexing form to match it. For more information on creating a new Automated Indexing form, see *Creating a New Automated Indexing Form* on page 73.

4. Click **Save**. The Automated Indexing form is applied to the document in the Document Viewer. Form Identification, Page Registration and Data Field Zones are highlighted on the displayed document.

Tip: At any time, click **Save Configuration** in the Tools window to save the Automated Indexing form. It is considered a best practice to save your Automated Indexing form after the form is created and after any configuration change is made (e.g., any Form Identification, Registration, or Data Field Zones are added, deleted, modified, moved, etc.).

To discard all changes since the last time the form was saved, click **Discard Configuration Changes** in the Tools window.

Now that the form is associated with the currently displayed document, it can be used as-is, modified or deleted.

CONFIGURING DOCUMENT BREAKS

If the documents matched to the Automated Indexing form consist of multiple pages, you can configure the form to automatically identify the beginning and end of a document (i.e., document breaks) in the batch based on the minimum and maximum number of pages allowed for a document matched to that form.

Note: If document breaks are specified by the scan format you are using to scan documents, then this step is not required.

To specify the minimum and maximum number of pages allowed for a document matched to a form, use the **Minimum Page Count** and **Maximum Page Count** fields on the **Form Definition** dialog box when creating a new form or modifying an existing form.

- **Minimum Page Count.** Set the **Minimum Page Count** field to the minimum number of pages that can compose a document identified by this Automated Indexing form. This setting must be set to at least a value of **1**.
- **Maximum Page Count.** Set the **Maximum Page Count** field to the maximum number of pages that can compose a document identified by this Automated Indexing form. Enter **0** if the OCR engine should not attempt to break documents (i.e., Automated Indexing should respect the pre-existing document breaks).

The following process describes how Automated Indexing attempts to determine how to break documents:

1. When the OCR engine encounters a new document (i.e., the first page of the first document in the batch or the first unclassified page after the end of the previously-identified document), it processes the page and attempts to match it to a form.
2. Once the page is matched to a form, the OCR engine determines the minimum number of pages required for documents matched to that form and appends a number of pages to the current page to create a document containing the minimum number of pages.
3. Once that document is created, the OCR engine moves to the next page in the batch to attempt to match it to a form.
4. If this page is not matched to a form to create a new document, the page is appended to the previous document.
5. Steps 3 and 4 are repeated until a page can be matched to a form to start a new document or until the page count of the document reaches the **Maximum Page Count** configured for the Automated Indexing form.
6. Once the **Maximum Page Count** is reached, the next page in the batch is automatically treated as a new document. If it is not matched to a form, this document is considered unclassified and the OCR engine moves to the next page in the batch and attempts to match it to a form.

For example:

A 20-page batch is scanned and the first page in the batch (Page 1) is matched to a form with a **Minimum Page Count** of **5** and a **Maximum Page Count** of **8**.

The OCR engine automatically appends the next four pages in the batch (Pages 2-5) to Page 1 to create a 5-page document and processes the next page in the batch (Page 6).

- If Page 6 can be matched to a form, it is treated as the first page in a new document.
- If Page 6 cannot be matched to a form, then it is appended to the previous document because it would now only consist of six pages, not the maximum of eight.

If Pages 6, 7, and 8 all could not be matched to a Automated Indexing form and are appended to the first document, then Page 9 is automatically treated as the first page of a new document because the maximum number of pages was reached for the first document.

If Page 9 cannot be matched to a form, it is treated as a unclassified document, and the OCR engine processes Page 10 to attempt to match it to a Automated Indexing form.

3

SPECIFYING AUTOMATED INDEXING FORM AVAILABILITY BY DOCUMENT TYPE

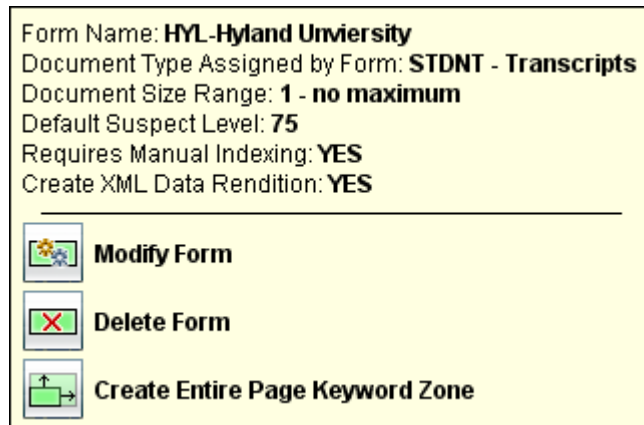
By default, a form is available only to the Document Type of the document it was associated with when it was created (i.e., only documents belonging to the same Document Type as the document used to create the Automated Indexing form can be matched to the Automated Indexing form).

However, you can configure a form to be available to additional Document Types (including unindexed documents) during the Automated Indexing process.


Tip: In addition to using this option to restrict Automated Indexing form availability by Document Type, you may also restrict Automated Indexing form availability by scan queue or the presence of a Keyword Value. These restrictions can be used separately or in conjunction with one another. Because a limitless number of Automated Indexing forms can exist in your Automated Indexing solution, it is considered a best practice to limit the number of Automated Indexing forms available as much as possible (by Document Type, Keyword Value, and/or scan queue) in order to save time and processing resources.


To specify a form's availability by Document Type:


1. Create the Automated Indexing form as usual.
2. Once the form has been created, from the **Automated Index Configuration** window, right-click on the document associated with the form in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.



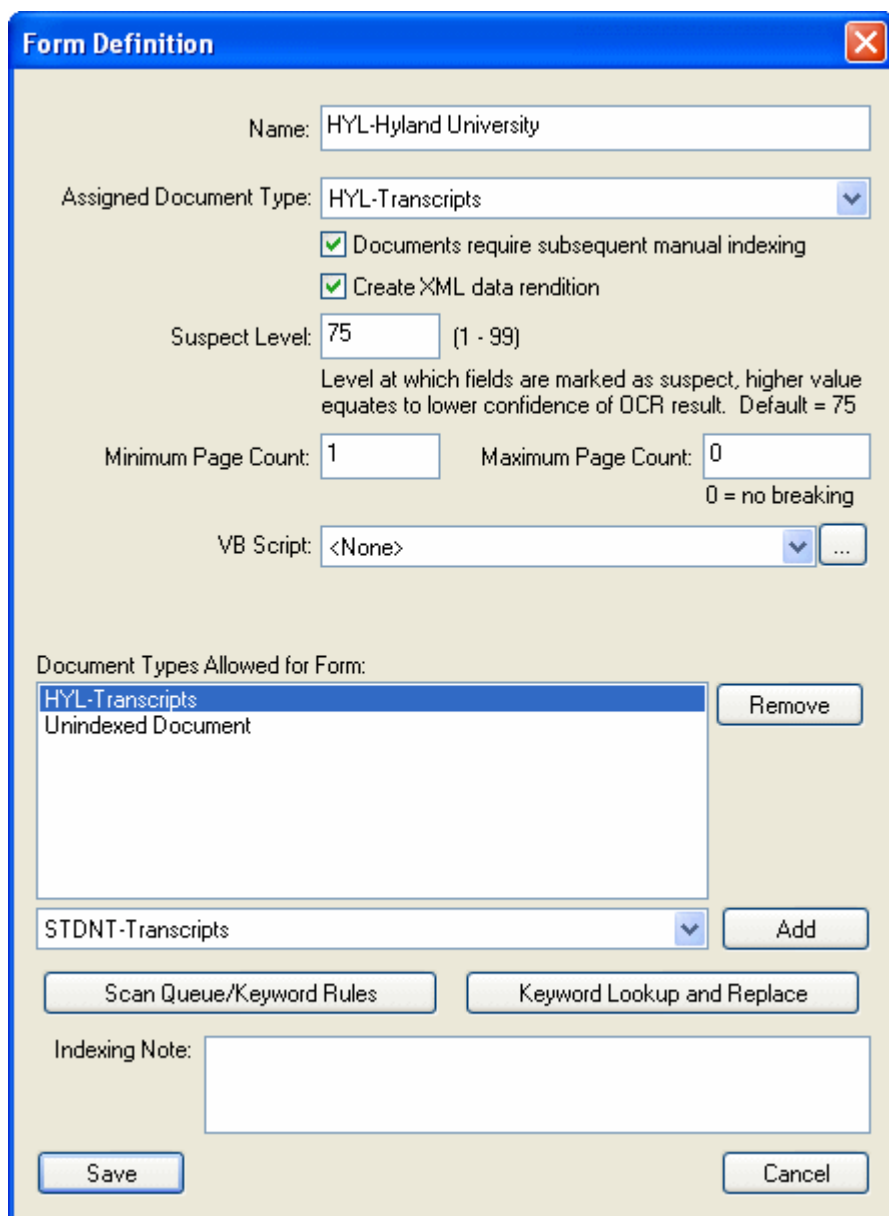
Form Name: **HYL-Hyland University**
Document Type Assigned by Form: **STDNT - Transcripts**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Requires Manual Indexing: **YES**
Create XML Data Rendition: **YES**

 **Modify Form**

 **Delete Form**

 **Create Entire Page Keyword Zone**

3. Click **Modify Form**. The **Form Definition** dialog box is displayed.

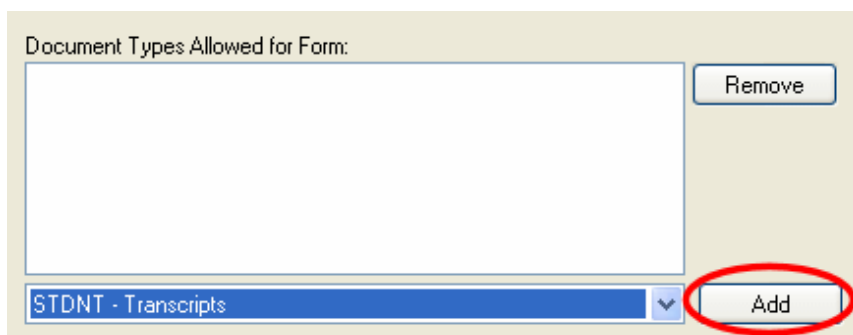


The **Form Definition** dialog box is shown with the following fields and controls:

- Name:** HYL-Hyland University
- Assigned Document Type:** HYL-Transcripts (dropdown menu)
- ☒ Documents require subsequent manual indexing
- ☒ Create XML data rendition
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1
- Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None> (dropdown menu with an ellipsis button)
- Document Types Allowed for Form:**
 - HYL-Transcripts (highlighted)
 - Unindexed Document
- Remove** button
- STDNT-Transcripts** (dropdown menu)
- Add** button
- Scan Queue/Keyword Rules** button
- Keyword Lookup and Replace** button
- Indexing Note:** (text area)
- Save** button
- Cancel** button

Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Automated Indexing form.

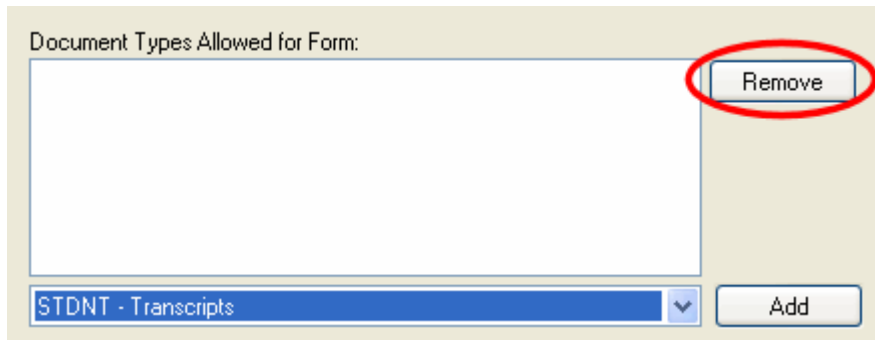
4. To add a Document Type allowed for the form (i.e., to allow documents from a Document Type to be matched to the Automated Indexing form), select the Document Type you would like to add from the drop-down below the Document Types Allowed for Form list and click **Add**.



Note: Only Document Types assigned to the scan queue are available in the drop-down.

Documents of this Document Type can now be matched to this Automated Indexing form. Depending on the Automated Indexing form's configuration, after the document is matched to the Automated Indexing form, its Document Type may be changed.

5. To remove a Document Type allowed for the form (i.e., to prevent documents from a Document Type from being matched to the Automated Indexing form), select the Document Type you would like to remove from the Document Types Allowed for Form list and click **Remove**.



Documents of this Document Type now cannot be matched to this Automated Indexing form. During processing, documents assigned to this Document Type are not compared to this Automated Indexing form.

6. Once you set the Document Types correctly for the Automated Indexing form, click **Save**.

CONFIGURING SCAN QUEUE OR KEYWORD VALUE RULES FOR A FORM


You may configure scan queue or Keyword Value requirements for associating documents with a form (i.e., in order to be associated with this Automated Indexing form, documents must have been scanned using a specified scan queue or must have one or more specified Keyword Values).


Additionally, you can also assign default Keyword Values for documents associated with a form (e.g., all documents associated with the **HYL-Hyland University** Automated Indexing form are assigned a Keyword Value of **RECEIVED** for the **Status** Keyword Type).


To configure scan queue or Keyword Value rules for a form:

1. Create the Automated Indexing form as usual.
2. Once the form has been created, from the **Automated Index Configuration** window, right-click on the document displayed in the Document Viewer. The **Modify/Delete Form** dialog box is displayed.

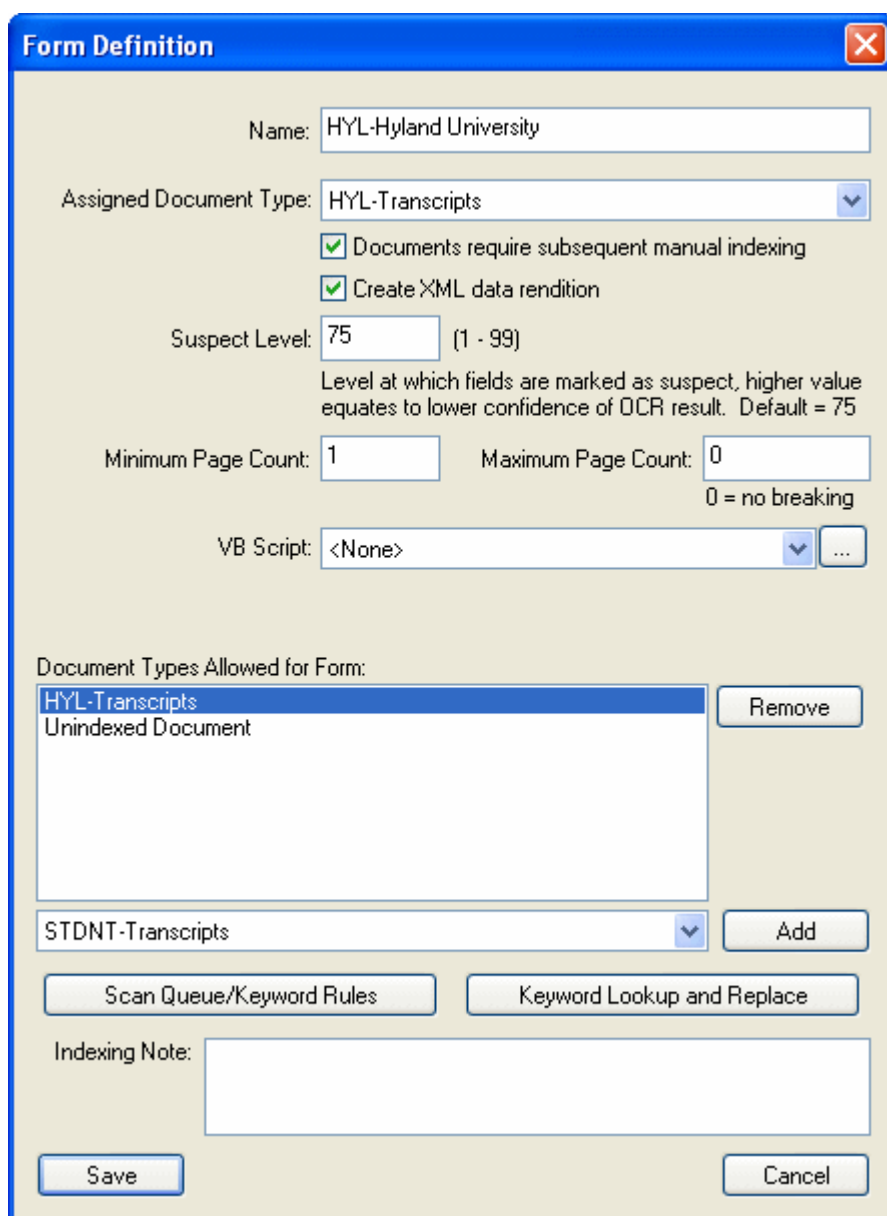
Form Name: HYL-Hyland Unviersity
Document Type Assigned by Form: STDNT - Transcripts
Document Size Range: 1 - no maximum
Default Suspect Level: 75
Requires Manual Indexing: YES
Create XML Data Rendition: YES

 **Modify Form**

 **Delete Form**

 **Create Entire Page Keyword Zone**

3. Click **Modify Form**. The **Form Definition** dialog box is displayed.

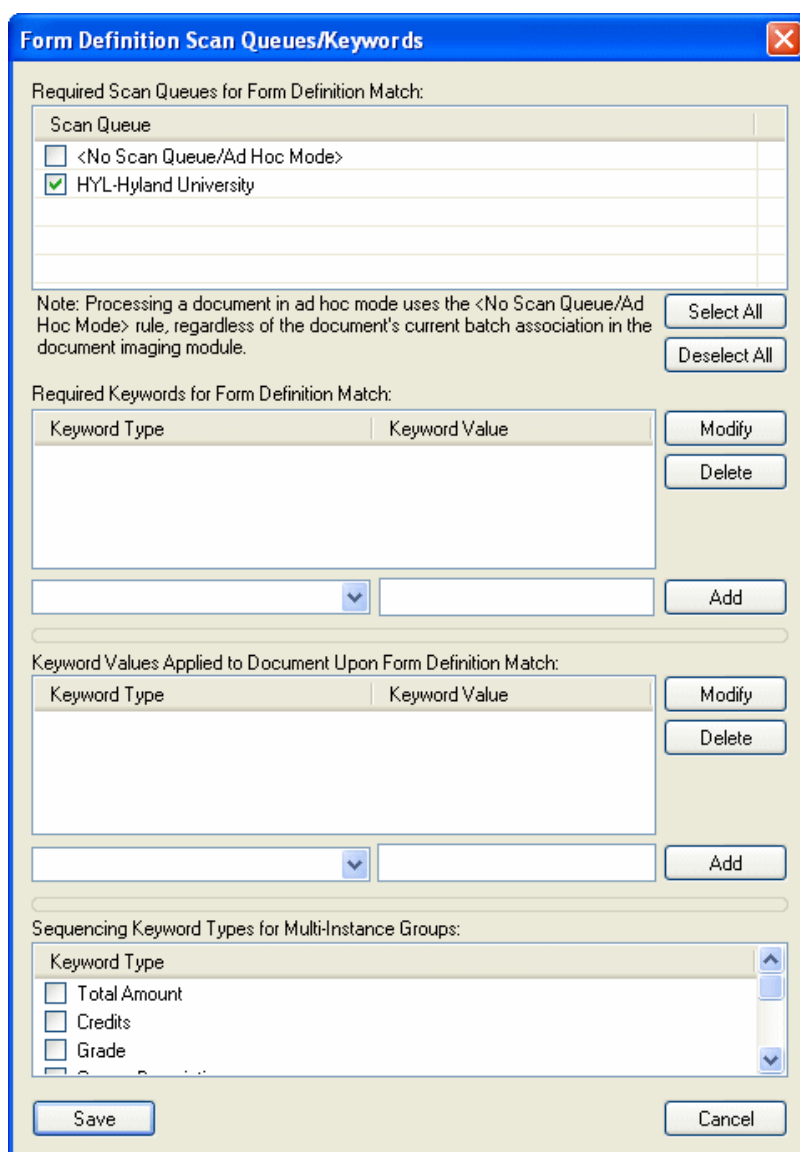


The **Form Definition** dialog box is shown with the following settings:

- Name:** HYL-Hyland University
- Assigned Document Type:** HYL-Transcripts
- ☒ Documents require subsequent manual indexing
- ☒ Create XML data rendition
- Suspect Level:** 75 (1 - 99)
Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75
- Minimum Page Count:** 1 **Maximum Page Count:** 0
0 = no breaking
- VB Script:** <None>
- Document Types Allowed for Form:**
 - HYL-Transcripts (selected)
 - Unindexed Document
- Buttons:** Remove, Add, Scan Queue/Keyword Rules, Keyword Lookup and Replace
- Indexing Note:** (empty text box)
- Buttons:** Save, Cancel

Note: The **Form Definition** dialog box displayed here is slightly different from the dialog box that is displayed when you are creating a new Automated Indexing form. For more information on these options and modifying an existing Automated Indexing form, see *Modifying a Form* on page 106.

- Click **Scan Queue/Keyword Rules**. The **Form Definition Scan Queues/Keywords** dialog box is displayed.



The dialog box is titled "Form Definition Scan Queues/Keywords" and contains three main sections for configuration.

Required Scan Queues for Form Definition Match:

Scan Queue
<input type="checkbox"/> <No Scan Queue/Ad Hoc Mode>
<input checked="" type="checkbox"/> HYL-Hyland University

Note: Processing a document in ad hoc mode uses the <No Scan Queue/Ad Hoc Mode> rule, regardless of the document's current batch association in the document imaging module.

Required Keywords for Form Definition Match:

Keyword Type	Keyword Value
--------------	---------------

Buttons: Select All, Deselect All, Modify, Delete, Add.

Keyword Values Applied to Document Upon Form Definition Match:

Keyword Type	Keyword Value
--------------	---------------

Buttons: Modify, Delete, Add.

Sequencing Keyword Types for Multi-Instance Groups:

Keyword Type
<input type="checkbox"/> Total Amount
<input type="checkbox"/> Credits
<input type="checkbox"/> Grade

Buttons: Save, Cancel.

5. Configure the scan queue and/or Keyword Value rules as needed:
 - **To specify the scan queues that the Automated Indexing form is available to (i.e., only documents scanned from the specified scan queues can be matched to the current Automated Indexing form):**

Tip: In addition to using this option to restrict Automated Indexing form availability by scan queue, you may also restrict Automated Indexing form availability by Document Type or the presence of a Keyword Value. These restrictions can be used separately or in conjunction with one another. Because a limitless number of Automated Indexing forms can exist in your Automated Indexing solution, it is considered a best practice to limit the number of Automated Indexing forms available as much as possible (by Document Type, Keyword Value, and/or scan queue) in order to save time and processing resources.

By default, a Automated Indexing form is available to only the scan queue that the document used to create for the form was scanned with and to documents that are not associated with a scan queue.

Note: Depending on your configuration, a form may be available to all scan queues once it is created. This default behavior is controlled by the **Assign Current Scan Queue** option in the Tools window of the **Automated Index Configuration** window.

To make the Automated Indexing form available to additional scan queues (or restrict it from additional scan queues):

In the Required Scan Queues for Form Definition Match section, select the check boxes next to the scan queues that the Automated Indexing form is available to. Click **Select All** to select all scan queues. Click **Deselect All** to deselect all scan queues.

To make the Automated Indexing form available to documents not associated with a scan queue or those documents that undergo Ad-Hoc Automated Indexing (where a scan queue association would not be known), select the **<No Scan Queue/Ad Hoc Mode>** check box.

Required Scan Queues for Form Definition Match:

Scan Queue	
<input type="checkbox"/> <No Scan Queue/Ad Hoc Mode>	
<input checked="" type="checkbox"/> HYL- Hyland University	

Note: Processing a document in ad hoc mode uses the <No Scan Queue/Ad Hoc Mode> rule, regardless of the document's current batch association in the document imaging module.

Select All

Deselect All

3

In the example illustrated above, only documents imported via the **HYL-Hyland University** scan queue can be matched to the Automated Indexing form.

- **To only allow documents with a specified Keyword Value to be matched to the Automated Indexing form:**

Tip: In addition to using this option to restrict Automated Indexing form availability by the presence of a Keyword Value, you may also restrict Automated Indexing form availability by Document Type or scan queue. These restrictions can be used separately or in conjunction with one another. Because a limitless number of Automated Indexing forms can exist in your Automated Indexing solution, it is considered a best practice to limit the number of Automated Indexing forms available as much as possible (by Document Type, Keyword Value, and/or scan queue) in order to save time and processing resources.

In the Required Keywords for Form Definition Match section, use the **Keyword Type** drop-down to select the Keyword Type, enter the Keyword Value in the **Keyword Value** field that documents must have in order to be associated with this Automated Indexing form, and then click **Add**.

Required Keywords for Form Definition Match:

Keyword Type	Keyword Value
Course ID	PSY111

Buttons: Modify, Delete

Below table: Keyword Type: Course ID, Keyword Value: , Add

Note: Only Keyword Types common to all Document Types specified in the Document Types Allowed for Form list for the Automated Indexing form can be selected from the **Keyword Type** drop-down. If only **Unindexed Document** is specified in the Document Types Allowed for Form list, no Keyword Types are available to be selected (i.e., no Keyword Value restrictions can be configured).

In the example illustrated above, only documents with a value of **PSY111** for the **Course ID** Keyword Type can be matched to the Automated Indexing form.

Repeat this process for as many Keyword Value restriction rules as are needed for the Automated Indexing form. To modify a default Keyword Value rule, double-click it in the Keyword Type/Keyword Value list or select it from the list and click **Modify**. To delete a default Keyword Value rule, select it from the Keyword Type/Keyword Value list and click **Delete**.

- **To add a default Keyword Value to all documents matched to the Automated Indexing form:**

In the Keyword Values Applied to Document Upon Form Definition Match section, use the **Keyword Type** drop-down to select the Keyword Type and enter the Keyword Value in the **Keyword Value** field that is to be automatically assigned as a default Keyword Value to documents matched to this Automated Indexing form, and then click **Add**.

Keyword Type	Keyword Value
Status	RECEIVED

Keyword Type: ▼ Status Keyword Value: Add

3

Note: Only Keyword Types assigned to the Document Type specified as the **Assigned Document Type** for the Automated Indexing form can be selected from the **Keyword Type** drop-down. If only **Unindexed Document** is specified in the Document Types Allowed for Form list and the Assigned Document Type is set to **<Document Type Unchanged>**, then no Keyword Types are available to be selected (i.e., no default Keyword Values can be configured)

In the example illustrated above, a value of **RECEIVED** is assigned to the **Status** Keyword Type for all documents that are matched to this Automated Indexing form. Repeat for as many default Keyword Values you would like to add to documents matched to this Automated Indexing form. To modify a default Keyword Value rule, double-click it in the Keyword Type/Keyword Value list or select it from the list and click **Modify**. To delete a default Keyword Value rule, select it from the Keyword Type/Keyword Value list and click **Delete**.

- When all scan queue and/or Keyword Value rules have been configured, click **Save**.

CONFIGURING KEYWORD LOOKUP/REPLACE DICTIONARIES

Your Automated Indexing solution can be configured to use a Keyword Lookup/Replace dictionary to replace Keyword Values extracted from documents via the Automated Indexing process with other values.

For example:

Your institution stores **Term** Keyword Values as a season and a four-digit year (e.g., **Fall 2008**, **Spring 2009**, **Summer 2009**, etc.). However, some institutions that you receive

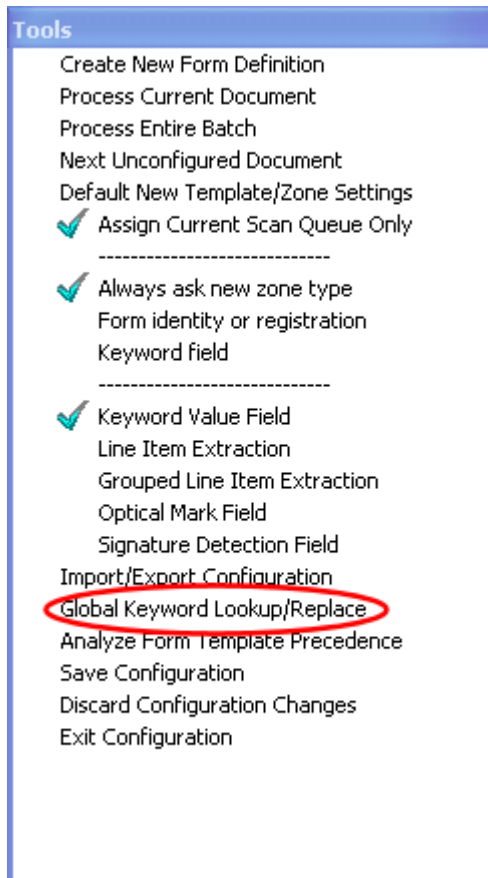
transcripts from identify terms using a season and a two-digit year (e.g., **Fall 08, Spring 09, Summer 09**).

You can configure a Keyword Lookup/Replace dictionary to automatically replace the undesirably-formatted Keyword Values extracted from the transcripts with your preferred values.

Keyword Lookup/Replace dictionaries can be configured in two ways:

- **Globally.** When a global Keyword Lookup/Replace dictionary is configured, a Keyword Value extracted using any Automated Indexing form in your solution can be replaced by an entry in the dictionary.

Entries in the Global Keyword Lookup/Replace dictionary are configured by clicking **Global Keyword Lookup/Replace** in the Tools Window of the **Automated Index Configuration** window.

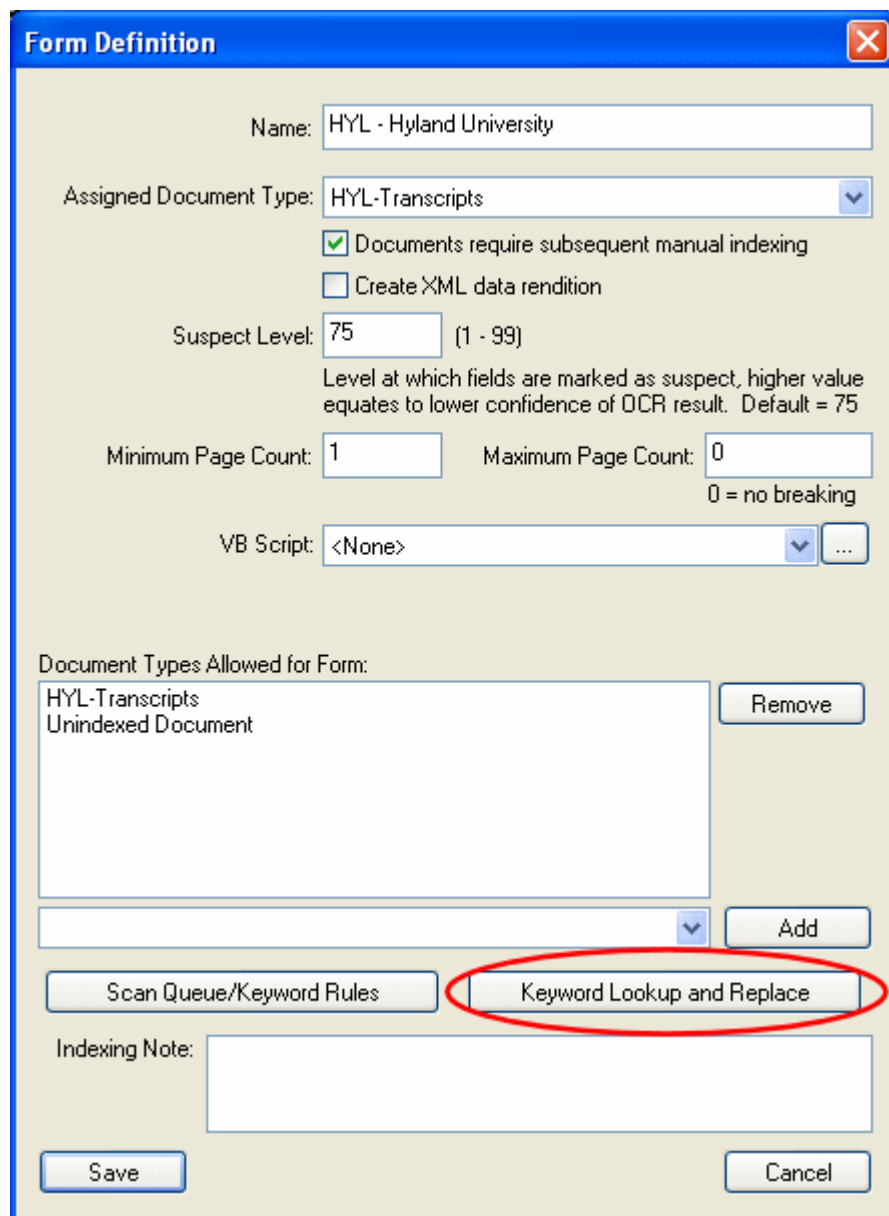


For example: You have several dozen Automated Indexing forms configured for your solution, and your global Keyword Lookup/Replace dictionary specifies that **Fall 08** should be replaced by **Fall 2008** for the **Term** Keyword Value. As a result, all instances where the **Term** Keyword Value is identified as **Fall 08** by the Automated Indexing process are replaced by **Fall 2008**, regardless of which Automated Indexing form was matched to the document the Keyword Value was extracted from.

- **At the Automated Indexing Form Level.** When a form-level Keyword Lookup/Replace dictionary is configured, a Keyword Value is only be replaced if it is specified in the Keyword Lookup/Replace dictionary for the form used to identify it.

Form-level Keyword Lookup/Replace dictionaries are configured by clicking **Keyword Lookup and Replace** in the **Form Definition** dialog box when modifying an existing Automated Indexing form.

3



The image shows a screenshot of the 'Form Definition' dialog box. The 'Name' field is 'HYL - Hyland University'. The 'Assigned Document Type' is 'HYL-Transcripts'. There are two checkboxes: 'Documents require subsequent manual indexing' (checked) and 'Create XML data rendition' (unchecked). The 'Suspect Level' is set to '75' with a range of '(1 - 99)'. Below this is a text explanation: 'Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75'. The 'Minimum Page Count' is '1' and the 'Maximum Page Count' is '0' with a note '0 = no breaking'. The 'VB Script' is set to '<None>'. There is a list of 'Document Types Allowed for Form' containing 'HYL-Transcripts' and 'Unindexed Document', with 'Remove' and 'Add' buttons. At the bottom, there are two buttons: 'Scan Queue/Keyword Rules' and 'Keyword Lookup and Replace', which is circled in red. There is also an 'Indexing Note' field and 'Save' and 'Cancel' buttons at the very bottom.

Form Definition

Name: HYL - Hyland University

Assigned Document Type: HYL-Transcripts

☒ Documents require subsequent manual indexing

☐ Create XML data rendition

Suspect Level: 75 (1 - 99)

Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: 1 Maximum Page Count: 0

0 = no breaking

VB Script: <None>

Document Types Allowed for Form:

HYL-Transcripts
Unindexed Document

Remove

Add

Scan Queue/Keyword Rules **Keyword Lookup and Replace**

Indexing Note:

Save Cancel

For example: You have two Automated Indexing forms in your solution (**HYL-Hyland University** and **WSU-Westlake State University**), and a Keyword Lookup/Replace dictionary is configured for the Hyland University form (but not the Westlake State University form) to replace **Term** Keyword Value with **Fall 2008** when it is identified by the Automated Indexing process as **Fall 08**. As a result, all instances of the **Term** Keyword set to **Fall 08** are replaced with **Fall 2008** for documents matched to Hyland University form, but **Fall 08** remains set as the **Term** Keyword Value for all documents matched to the Westlake State University form.

If both a global and a form-level Keyword Lookup/Replace dictionary entry are configured for the same Keyword Value, then the form-level entry will override the global entry.

Note: A Keyword Value can only be replaced once. If a Keyword Value is replaced by another value, and the replacement Keyword Value also has an entry in the Keyword Lookup/Replace dictionary, the replacement value is not replaced.

3

Creating a Keyword Lookup/Replace Dictionary Entry

1. Determine whether you would like to create an entry in the global Keyword Lookup/Replace dictionary or a form-level Keyword Lookup/Replace dictionary:
 - *To create an entry in the global Keyword Lookup/Replace dictionary:* click **Global Keyword Lookup/Replace** in the Tools Window of the **Automated Index Configuration** window.

- To create an entry in a form-level Keyword Lookup/Replace dictionary: click **Keyword Lookup and Replace** in the **Form Definition** dialog box when modifying an existing Automated Indexing form.

The **Keyword Lookup and Replace** dialog box is displayed.

[illegible]

Note: The **Context Level** label near the bottom of the **Keyword Lookup and Replace** dialog box indicates if you are creating an entry in the global dictionary (i.e., **Context Level: Global**) or if you are creating an entry in a form-level dictionary (**Context Level: <Automated Indexing Form Name>**).

- Using the **Keyword Type** drop-down select list, select the Keyword Type associated with Keyword Values that are to be replaced.

Note: If you are creating an entry in a global dictionary, all Keyword Types configured for your Automated Indexing solution are available in the **Keyword Type** drop-down select list. If you are creating an entry in a form-level dictionary, only Keyword Types assigned to Document Types associated with the Automated Indexing form are available in the **Keyword Type** drop-down select list.

- Click **Add**. The **Edit Keyword Lookup and Replace Entry** dialog box is displayed.

Edit Keyword Lookup and Replace Entry

Keyword Type: Course Description

Compare Value:

(Literal text value enclosed in quote marks, or regular expression not in quotes)

Replacement Value:

☒ Replace with static text ☐ Replace using capture groups
☐ Replace with <blank> ☐ Leave value unchanged

☒ Suspect Level Remains Unchanged
☐ Mark Value as Suspect
☐ Mark Value as Not Suspect

- In the **Compare Value** field, enter the Keyword Value that is to be replaced if it is found.

This value can be a literal text string (i.e., specific text, such as **Fall 08**) or a regular expression (i.e., text that matches a certain pattern, such as a value containing four letters, a space, and two digits).

Literal values must be enclosed in quotation marks to differentiate them from regular expressions. Multiple values (literal values, regular expression, and/or both) can be configured by separating them by a space in the **Compare Value** field. Literal compare values are not case-sensitive.

5. Select the radio button that describes how you would like the compare value to be treated once it is identified:

- **Replace with static text.** Select this option if you would like to replace the compare value with a static text string. Enter the static text you would like to replace the compare value with in the **Replacement Value** field.

For example: You would like to replace all **Term** Keywords set to **Fall 08** with **Fall 2008**. Set the compare value to **"Fall 08"** (a literal text value), select the **Replace with static text** radio button, and specify the replacement value as **Fall 2008**.

- **Replace with <blank>.** Select this option if you would like to replace the compare value with a blank value. Once this option is selected, the **Replace Value** field is disabled.

For example: You would like to discard all **Term** Keywords set to **Fall 08** and have a user manually enter a new value during Index Verification and Review. Set the compare value to **"Fall 08"** (a literal text value) and select the **Replace with <blank>** radio button.

Replace using captured groups. Select this option if you would like to replace the compare value with a new value composed of sections of the compare value.

Note: This feature is only available when specifying a regular expression for the compare value.

By enclosing sections of your regular expression compare value in parenthesis, you can specify the portions of the compare value (i.e., the building blocks) that are used to create the replacement value. Then, specify the sections of the compare value (e.g., **[1] [2] [4]**) you would like to use to compose the new Keyword Value in the **Replacement Value** field.

For example: Your institution stores Keyword Values for the **Term** Keyword using a **<Season> <Four-Digit Year>** format (e.g., **Fall 2008, Spring 2009, Summer 2009**). However, some transcripts received from other institutions use a **<Season> Semester <Four-Digit Year>** format for **Term** Keyword Values (e.g., **Fall Semester 2008, Spring Semester 2009, Summer Semester 2009**).

Create the following regular expression to identify the compare value:

(\S{4,6}\s)(SEMESTER)\s(\d{4}). As you can see, parenthesis are used to divide the value into three sections: the season **(\S{4,6}\s)**, Semester **(SEMESTER)**, and the year **(\d{4})**.

To drop the Semester “building block” (section #2) from the compare value and use only the season and the year “building blocks” (sections #1 and #3) to create the new value, enter **[1] [3]** in the Replacement Value field. **Term** Keyword Values matching the **<Season> Semester <Four-Digit Year>** format are replaced with Keyword Values matching the **<Season> <Four-Digit Year>** format.

- **Leave value unchanged.** Select this option if you would like to leave the compare value as-is. Once this option is selected, the **Replace Value** field is disabled.

Tip: This option is typically used in a form-level Keyword Lookup/Replace dictionary entry to override an entry in the global dictionary.

For example: You have an entry in the global Keyword Lookup/Replace dictionary that specifies that **Term** Keywords set to **Fall 08** should be changed to **Fall 2008**. However, you have one non-transcript document for which you would like to leave the **Term** Keyword Value as-is.

Leave the global dictionary entry in place and configure a form-level dictionary entry for the Automated Indexing form that is matched to the non-transcript document where the compare value is set to **"Fall 08"** (a literal text value) and select the **Leave Value Unchanged** radio button. Because the form-level dictionary overrides the global dictionary, Fall 08 is changed to Fall 2008 for **Term** Keywords on all documents indexed via Automated Indexing except for the one non-transcript document you configured the form-level dictionary entry for.

6. Select the radio button that determines how the replacement value should be treated in terms of Suspect Level.
 - **Suspect Level Remains Unchanged.** The replacement value is given the Suspect Level score given to the compare value when it was identified by the OCR engine. The replacement value's score is compared to the Suspect Level threshold configured for the zone (or the form itself) to determine if it should be considered suspect.
 - **Mark Value as Suspect.** The replacement value is automatically marked as suspect, regardless of the Suspect Level score given to the compare value.
 - **Mark Value as Not Suspect.** The replacement value is automatically marked as not suspect, regardless of the Suspect Level score given to the compare value.

- Click **Save**. The **Edit Keyword Lookup and Replace Entry** dialog box is closed and you are returned to the **Keyword Lookup and Replace** dialog box. The entry you created in Steps 2-6 is added to the dictionary.

[illegible]

8. Repeat Steps 2-6 to add additional entries to the Keyword Lookup/Replace dictionary.
 - To modify the order of the entries in the dictionary, select an entry and click **Move Up** or **Move Down**.

Tip: If multiple entries in the Keyword Lookup/Replace dictionary affect the same Keyword Value, only the first entry in the dictionary will modify its value (remember -- a Keyword Value can only be replaced once). Use the **Move Up** and **Move Down** buttons to ensure the order of the entries will correctly identify and replace Keyword Values in your solution.

- To delete an entry, select it and click **Delete**.

9. Once the Keyword Lookup/Replace dictionary is complete, click **Save**. The **Keyword Lookup and Replace** dialog box is closed.

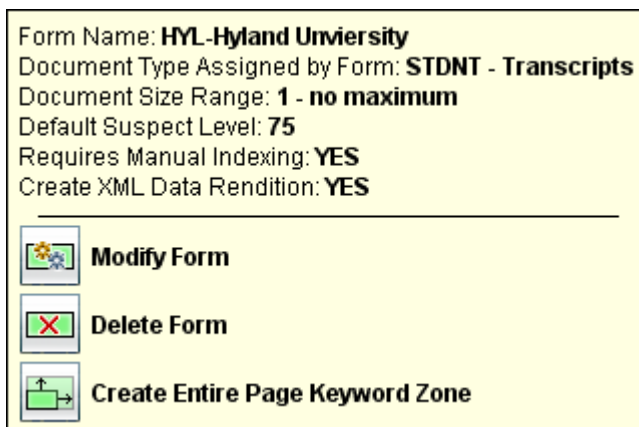
Adding an Additional or Modifying/Deleting a Keyword Lookup/Replace Dictionary Entry

New entries can be created and existing entries can be modified or deleted at any time for either the global or a form-level Keyword Lookup/Replace dictionary by accessing the **Keyword Lookup and Replace** dialog box.

- *To access the global Keyword Lookup/Replace dictionary:* click **Global Keyword Lookup/Replace** in the Tools Window of the **Automated Index Configuration** window.
- *To access a form-level Keyword Lookup/Replace dictionary:* click **Keyword Lookup and Replace** in the **Form Definition** dialog box when modifying an existing Automated Indexing form.




MODIFYING OR DELETING A FORM

An existing Automated Indexing form can be modified or deleted by right-clicking on the document it has been applied to in the Document Viewer.



A context menu is displayed with a yellow background. It contains the following text and icons:

- Form Name: **HYL-Hyland Unviersity**
- Document Type Assigned by Form: **STDNT - Transcripts**
- Document Size Range: **1 - no maximum**
- Default Suspect Level: **75**
- Requires Manual Indexing: **YES**
- Create XML Data Rendition: **YES**

-  **Modify Form**
-  **Delete Form**
-  **Create Entire Page Keyword Zone**

Information about the current Automated Indexing form is displayed above the following options:

- **Modify Form.** Click to display the **Form Definition** dialog box to modify the Automated Indexing form's configuration settings. For more information, see *Modifying a Form* on page 106.

From **Form Definition** dialog box, you can also specify the Document Types that the Automated Indexing form is available to be applied to. For more information, see *Specifying Automated Indexing Form Availability by Document Type* on page 81.

- **Delete Form.** Click to delete the selected Automated Indexing form. All associated Form Identification, Registration and Data Field Zones are deleted.
- **Create Entire Page Keyword Zone.** Click to create a Keyword Data Zone that covers the entire page.

Modifying a Form

The **Form Definition** dialog box is displayed:

Form Definition

Name:

Assigned Document Type: ▼

☒ Documents require subsequent manual indexing

☒ Create XML data rendition

Suspect Level: (1 - 99)

Level at which fields are marked as suspect, higher value equates to lower confidence of OCR result. Default = 75

Minimum Page Count: Maximum Page Count:

0 = no breaking

VB Script: ▼ ...

Document Types Allowed for Form:

HYL-Transcripts	Remove
Unindexed Document	

▼ Add

Indexing Note:

Use the following options (as needed) to modify the current Automated Indexing form.

Option	Description
Name	<p>Enter the name of the Automated Indexing form.</p> <hr/> <p>Tip: To create a new Automated Indexing form based on an existing form, modify the existing form and rename it.</p> <hr/>
Assigned Document Type	<p>Select the Document Type that documents are assigned to once they are matched to this Automated Indexing form.</p> <p>To leave the Document Type of documents undergoing Automated Indexing processing unchanged (e.g., to ensure that the document remains unindexed or that the document still belongs to the Document Type assigned when the document was scanned/imported, select <Document Type Unchanged>.</p>
Documents require subsequent manual indexing	<p>This check box determines if documents associated with this Automated Indexing form are routed to the Index in Progress batch status queue after undergoing Automated Indexing for additional manual indexing.</p> <ul style="list-style-type: none"> • When this check box is selected, all documents matched to this Automated Indexing form are automatically routed for index verification and review. • When this check box is not selected, documents are not automatically routed for index verification and review. Documents are routed for index verification and review only if one or more captured values associated with the document have been marked as suspect by the OCR engine.
Create XML data rendition	<p>Select this check box to create XML renditions of the documents matched to this Automated Indexing form.</p> <p>The XML renditions contain indexing values (e.g., the Document Date and Keyword Values) identified during the Automated Indexing process. The structure of the XML file (i.e., the element that the Keyword Value is contained in) is defined when configuring the Data Field Zone for that Keyword Value.</p>

Option	Description
Suspect Level	<p>Enter the Suspect Level threshold, 1-99, in this field.</p> <p>The Suspect Level is the level of confidence placed in the data values captured using this form.</p> <p>After a zone is processed, the OCR engine gives the result a score between 1 and 99, depending on how confident it is in the results it returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set cause the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no score could be returned that was higher and no result could be marked as suspect.</p> <p>Setting the Suspect Level at 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Suspect Levels can also be set at the zone level for individual zones if more or less stringent standards for their values.</p> <hr/> <p>Tip:By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p> <hr/>
Minimum Page Count	<p>The Minimum Page Count field contains the minimum number of pages that can compose a document identified by this Automated Indexing form.</p> <p>This setting must be set to at least a value of 1.</p>
Maximum Page Count	<p>The Maximum Page Count field contains the maximum number of pages that can compose a document identified by this Automated Indexing form.</p> <p>If set to 0, the Automated Indexing process will not attempt to break documents (i.e., Automated Indexing respects the pre-existing document breaks).</p>

Option	Description
VB Script	<p>Use the VB Script drop-down select list to select a VB script to associate with the identification and processing of this Automated Indexing form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>
Document Types Allowed for Form	This option allows you to specify the Document Types that this Automated Indexing form is available to. (i.e., the allowable “incoming” Document Types for this form).
Scan Queue/Keyword Rules	Click this button to configure scan queue or Keyword Value rules configured for this Automated Indexing form.
Keyword Lookup and Replace	<p>Click this button to configure a form-level Keyword Lookup/Replace dictionary for this Automated Indexing form. A Keyword Lookup/Replace dictionary is used to replace Keyword Values identified by the Automated Indexing process with different values.</p> <p>For example, if you require a Term Keyword Value to contain a Season and a Four Digit Year (e.g., Fall 2010) and some documents display a Season and a Two Digit Year (e.g., Fall 10), you could configure a lookup/replace action to automatically replace Term Keywords Values identified as Fall 10 with Fall 2010.</p>
Indexing Note	<p>Enter an indexing note of up to 250 character to be displayed for documents matched to this Automated Indexing form.</p> <p>This note is displayed to users when a document matched to this form is displayed in the Working window during index verification and review.</p>

Once the form has been modified, click **Save** to save your changes and close the **Form Definition** dialog box.

Configuring A Form Identification Zone

Once the general Automated Indexing form options have been set in the **Form Definition** dialog box, a Form Identification Zone must be configured for the form.

A Form Identification Zone is used to detect the an image(s) or text value(s) on a document that is unique to that document layout. Automated Indexing uses these images or text values to determine if a Automated Indexing form is a match to a document (i.e., if the image or text value is detecting in the Form Identification Zone, then the document is determined to be a match to that Automated Indexing form).

Tip: Multiple Form Identification zones can be configured for a Automated Indexing form.

Once a document is matched to a form, data values can be extracted from the document via the Automated Indexing process.

To create a Form Identification Zone:

Note: Prior to creating the Form Identification Zone, ensure that the **Default Zone Type** option in the Tools window is set to either **Always Ask** or **Form identity or registration**.

1. With the document that the Automated Indexing form is being created from displayed in the Document Viewer, click and hold the left mouse button and use the pointer to draw a box around the text or image that is to be used as the Form Identification value.

Automated Index Configuration

Configuration << Collapse

- Current Form Definition
 - HYL-Hyland University
 - Document Types Allowed for Form
 - STDNT - Transcripts
 - Document Type Assigned by Form
 - STDNT - Transcripts
 - Form Identification Rules
 - Required Scan Queues
 - HYL- Hyland University
 - <No Scan Queue/Ad Hoc Mode>
 - Form Keyword Fields List
 - <None>
 - All Form Definitions
 - All Document Types

Result Verification

Hyland University

Tools

- Default New
- Assign
- Always
- Form id
- Keywor
- Keywor

Official Transcript

This is not an official transcript without Registrar's signature and college seal.

Reagan Ronald W. 012345 12/25
(Last Name) (First Name, Middle Initial) (ID#) (Birthdate)

Course	Course Title	CRD	GRD	GRDPT
FALL 2006				
PSY101	Intro to College Statistics	3.0	A-	3.5
SPC201	Speech Communications I	2.0	B	3.0
FIN200	Reading Financial Information	1.0	A	4.0
ENG111	Composition I	3.0		
	Total GPA: 3.5			4.0
	Cum GPA: 3.5			4.0
	Cum Honor Points: 3.0			
	Cum Credits Attempted: 9.0			

End of official record

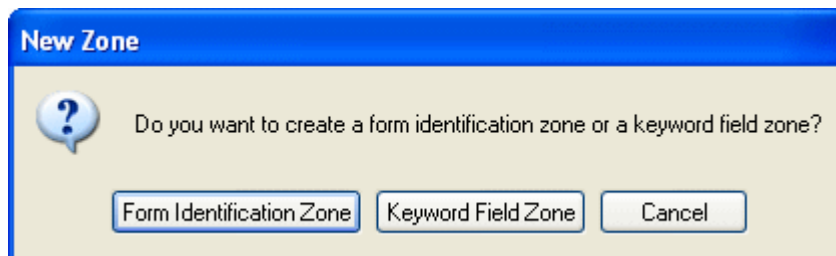
RECEIVED
AUG 11 2007

John Doe, Registrar

Tip: Note that the area of the document that you selected is displayed in the Result Verification window.

2. Release the left mouse button to display the **Form Identification Zone** dialog box.

Depending on the setting of the **Default Zone Type** option in the Tools window, the **New Zone** dialog box may be displayed.



If the **New Zone** dialog box is displayed, click **Form Identification Zone** to continue to the **Form Identification Zone** dialog box.

Form Identification Zone

Text Match Image Match

Match value: **HYLAND UNIVERSITY**

☒ Exact match ☐ Contains text ☐ Fuzzy match Max errors:

☒ Remove all whitespace

Filtering

Allowed character types:

<input checked="" type="checkbox"/> Numerals	<input checked="" type="checkbox"/> Punctuation
<input checked="" type="checkbox"/> Uppercase	<input checked="" type="checkbox"/> Miscellaneous
<input checked="" type="checkbox"/> Lowercase	<input type="checkbox"/> Exclude Asian

Suspect level: (0 - 99)
Form Default = 75, Actual = 70

Combined rule expressions:
<Required Keyword Value Rules>
Identification Group #1
<This Identification Rule>

VB Script:

☒ Use registration point:

Hyland University

3. Select the type of identifier you are using to match the Automated Indexing form to incoming documents.
 - If you are using a text value as an identifier to match the Automated Indexing form to a document, select the Text Match tab.
 - If you are using an image, such as a logo, as an identifier to match the Automated Indexing form to a document, select the Image Match tab.
4. Configure the Form Identification Zone using the **Form Identification Zone** dialog box.
5. Once you have set all necessary options on the **Form Identification Zone** dialog box, click **Save**. The **Form Identification Zone** dialog box is closed and the Form Identification Zone is highlighted in magenta on the document in the Document Viewer.

The image shows a sample document titled "Official Transcript" from Hyland University. A magenta rectangular box highlights the text "Hyland University" at the top left of the document, indicating it as the Form Identification Zone. The document also contains a logo, a date field "August 30, 2007 (Issue Date)", a disclaimer "This is not an official transcript without Registrar's signature and college seal.", and several fields for personal information: "Reagan (Last Name)", "Ronald W (First Name, Middle Initial)", "012345 (ID #)", and "12/25 (Semester)".

6. Repeat Steps 1-5 for each Form Identification Zone that you wish to create for the currently-displayed Automated Indexing form.

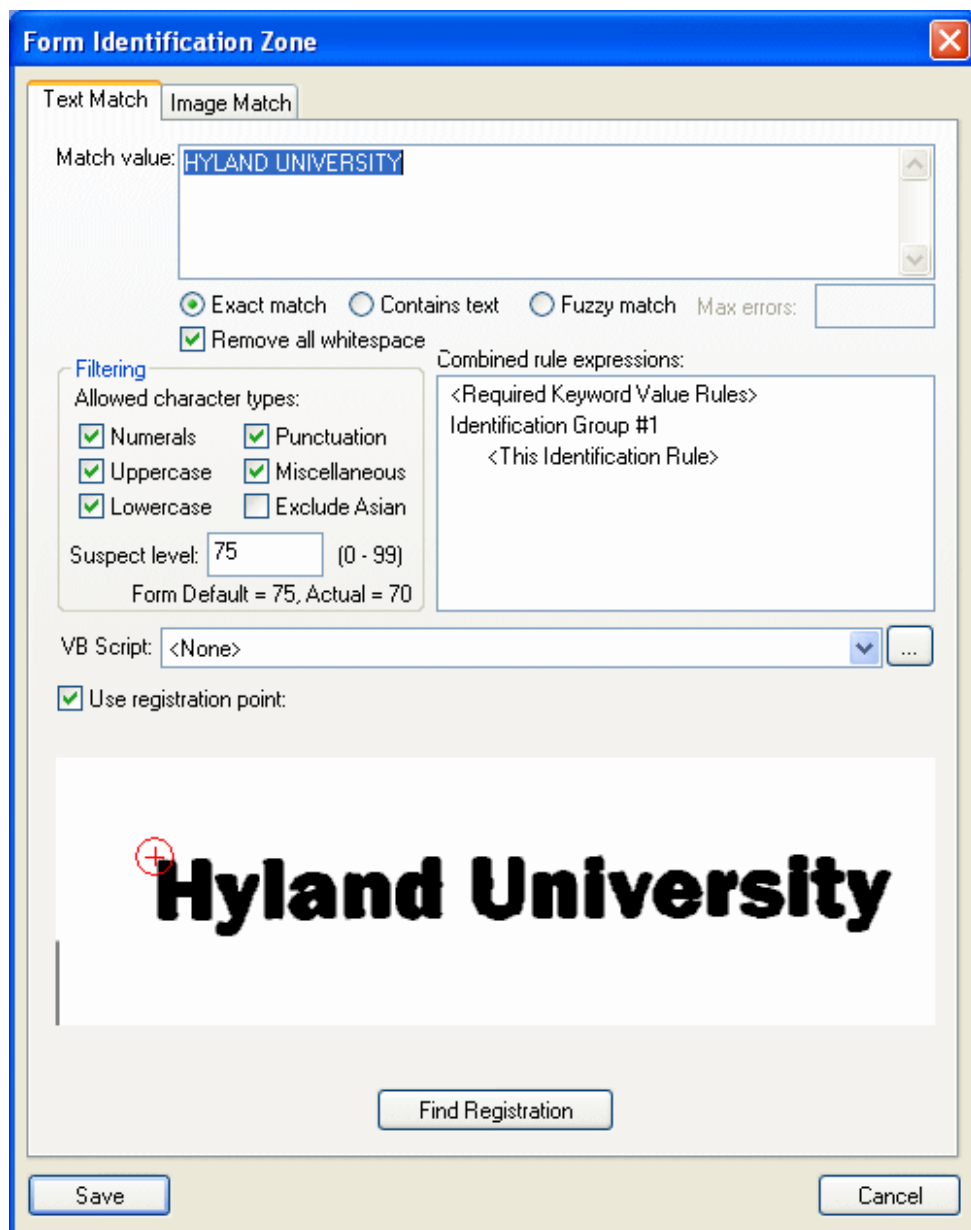
Tip: Depending on your scan queue configuration, documents that have the same Keyword Values as existing OnBase documents can be appended to existing documents. See the Document Imaging documentation for more information on the **Append to Existing Documents Matched on Keyword(s)** scan queue setting.

Tip: At any time, click **Save Configuration** in the Tools window to save the form to the database. It is considered a best practice to save your Automated Indexing form configuration after the creation and configuration of every Automated Indexing form and each form's Form Identification, Registration and Data Field Zones.

To discard all configuration changes (including any newly-created Form Identification Zones) made since the last time the form was saved, click **Discard Configuration Changes** in the Tools window.

CONFIGURING A TEXT VALUE FORM IDENTIFIER

A text value form identifier is configured using the options on the Text Match tab of the **Form Identification Zone** dialog box.



The **Form Identification Zone** dialog box is shown with the **Text Match** tab selected. The **Match value:** field contains **HYLAND UNIVERSITY**. The **Exact match** radio button is selected, and the **Remove all whitespace** checkbox is checked. The **Filtering** section shows **Allowed character types:** with **Numerals**, **Uppercase**, **Lowercase**, **Punctuation**, and **Miscellaneous** checked, and **Exclude Asian** unchecked. The **Suspect level:** is set to **75** (range 0 - 99), with **Form Default = 75, Actual = 70** noted below. The **Combined rule expressions:** field contains **<Required Keyword Value Rules>**, **Identification Group #1**, and **<This Identification Rule>**. The **VB Script:** dropdown is set to **<None>**. The **Use registration point:** checkbox is checked. A preview of the text **Hyland University** is shown with a red registration point icon. The **Find Registration**, **Save**, and **Cancel** buttons are at the bottom.

Form Identification Zone

Text Match Image Match

Match value: HYLAND UNIVERSITY

☒ Exact match ☐ Contains text ☐ Fuzzy match Max errors:

☒ Remove all whitespace

Filtering

Allowed character types:

☒ Numerals ☒ Punctuation

☒ Uppercase ☒ Miscellaneous

☒ Lowercase ☐ Exclude Asian

Suspect level: 75 (0 - 99)

Form Default = 75, Actual = 70

Combined rule expressions:

<Required Keyword Value Rules>

Identification Group #1

<This Identification Rule>

VB Script: <None>

☒ Use registration point:

Hyland University

Find Registration

Save Cancel

Form Identification Zone Text Match Options	Description
Match Value	<p>This field displays the text value returned from the OCR engine's evaluation of the Field Identification Zone.</p> <hr/> <p>Note: The Match Value is also displayed in the Result Verification window below the enlarged image of the Field Identification Zone.</p> <hr/> <p>You may select the match rules for the Field Identification Zone by selecting one of the match rule radio buttons.</p> <ul style="list-style-type: none"> • Exact match. Select this radio button if the value identified by the OCR engine when a document is processed must exactly match the Match Value in order to match the document to the Automated Indexing form. • Contains text. Select this radio button if the Match Value can be part of a longer text string identified by the OCR engine for the document being processed in order for Automated Indexing to match the document to the Automated Indexing form. <hr/> <p>Note: When using this option, only the Match Value portion of the text string is taken into consideration to determine the Suspect Level of the value.</p> <hr/> <ul style="list-style-type: none"> • Fuzzy match. Select this radio button if the Match Value should match the value detected by the OCR engine for the document, but some margin of error is allowed between the two values. <p>Enter a value in the Max Errors field to set the number of discrepancies allowed between the Match Value and the value read from the document being processed.</p> <p>For example, if your Match Value is HYLAND UNIVERSITY and the value detected by the OCR engine from the processed document is HYL4ND UNIVERS1TY, the document will still be matched to the Automated Indexing form if the Max Errors value is set to 2 or more.</p>

Form Identification Zone Text Match Options	Description
Remove all whitespace	<p>Select this check box if you would like to remove all whitespace (e.g., spaces, paragraph returns, etc.) from the Match Value after it is read by the OCR engine.</p> <p>For example, if the Match Value read by the OCR engine is ENG 313 and the Remove all whitespace check box is selected, the Match Value would be changed to ENG313.</p>
Allowed character types	<p>Select the check box next to each character type that may be included in the Match Value. The available options are:</p> <ul style="list-style-type: none"> • Numerals. Numeric characters, 0-9. • Uppercase. Uppercase alphabetic characters. • Lowercase. Lowercase alphabetic characters. • Punctuation. Punctuation marks (i.e., . ! ?). • Miscellaneous. Other ASCII characters that do not fall into one of the above categories (i.e., # \$ * @). <p>When a character is recognized by the OCR engine that is not part of a selected character set, the character is replaced by a tilde (~), and the value is automatically marked as suspect.</p> <hr/> <p>Tip:Using the Allowed character type options can sometimes help the OCR engine more easily read the correct value by eliminating characters that are obviously not correct (e.g., an I is correctly identified instead of a 1 because numeric characters are filtered and prevented from being recognized as part of the value).</p> <hr/> <p>Select the Exclude Asian check box to instruct the OCR engine to skip Asian (i.e., double-byte) characters.</p> <hr/> <p>Tip:Selecting the Exclude Asian check box allows you to identify numeric data (i.e., Date Keyword Values, Currency Keyword Values and Document Dates) when performing Automated Indexing on documents configured to contain Asian (i.e., double-byte) characters.</p> <hr/>

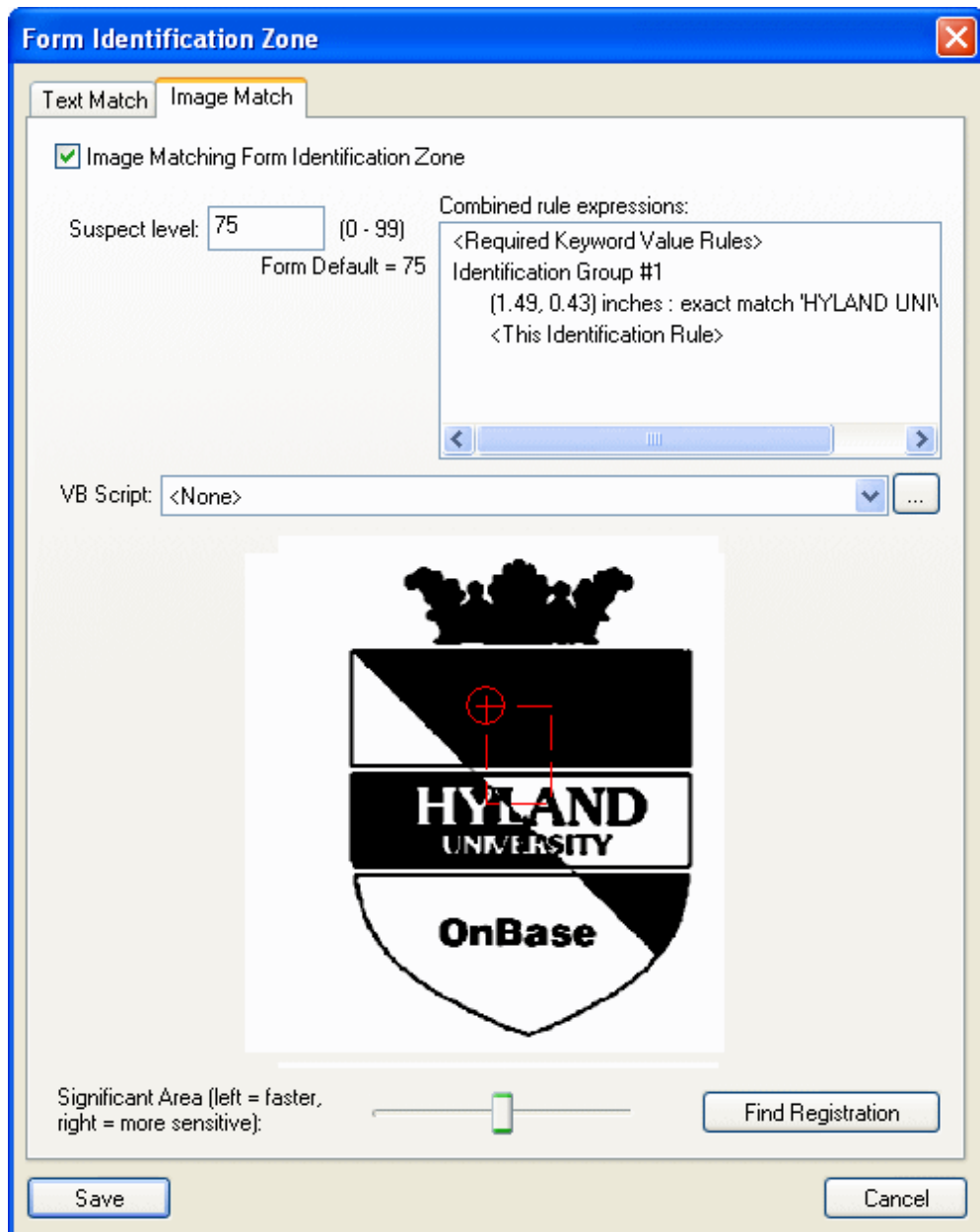
Form Identification Zone Text Match Options	Description
Suspect Level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured using this field. The default Suspect Level for the Automated Indexing form and the actual Suspect Level detected for the selected Form Identification Zone are displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and not result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p>

Form Identification Zone Text Match Options	Description
Combined rule expressions	<p>In the case where multiple Form Identification Zones are needed to match a document to a form, you can configure the how these zones are processed using the Combined rule expressions option.</p> <p>Rules can be dragged and dropped into the proper order in this field and can be combined using the AND or OR boolean operators.</p> <ul style="list-style-type: none"> • All rules placed before the current (<This Identification Rule>) rule are combined using an OR operator. • All rules placed after the current (<This Identification Rule>) rule are combined using an AND operators. <p>Within the Combined rule expressions box, Form Identification Zones can be copied so that they can be used multiple times within the same rule or in other rules. To copy a Form Identification Zone, right-click it in the Combined rule expressions box and drag it to its new location in the current rule or another rule.</p> <hr/> <p>Note:Keyword rules or Identification Groups cannot be copied.</p> <hr/>
VB Script	<p>Use the VB Script drop-down select list to select a VB script to associate with the identification of this Automated Indexing form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>

Form Identification Zone Text Match Options	Description
Use registration point	<p>Select this check box to enable the Registration Point feature.</p> <hr/> <p>Note: The Registration Point is the starting point (top, left corner) of the value detected in the Form Identification Zone. Click Find Registration to automatically set the Registration Point at the starting point of the value in the Form Identification Zone, or double-click a location in the field below the check box to manually set the registration point at that position.</p> <hr/> <p>The position of the Registration Point on the Automated Indexing form is compared to the same position on the document being processed to determine the offset (i.e., the skew or rotation) of any imported documents. The position of the configured Form Identification Zone is adjusted on the document being processed to account for the detected offset to ensure that the Automated Indexing process is able to process the document properly.</p>

CONFIGURING AN IMAGE MATCH FORM IDENTIFIER

An image identifier is configured using the options on the Image Match tab of the **Form Identification Zone** dialog box.



3

Form Identification Zone Image Match Options	Description
Image Matching Form Identification Zone	<p>Select this check box to configure this Form Identification Zone to use an image identifier. By default, Form Identification Zones are configured to use text values as identifiers.</p> <p>In order to match a document to this Automated Indexing form, the OCR engine must find the image displayed in this Form Identification Zone in the same location on the document being processed.</p> <p>This check box must be selected in order to enable the other options on the Image Match tab of the Form Identification Zone dialog box.</p>

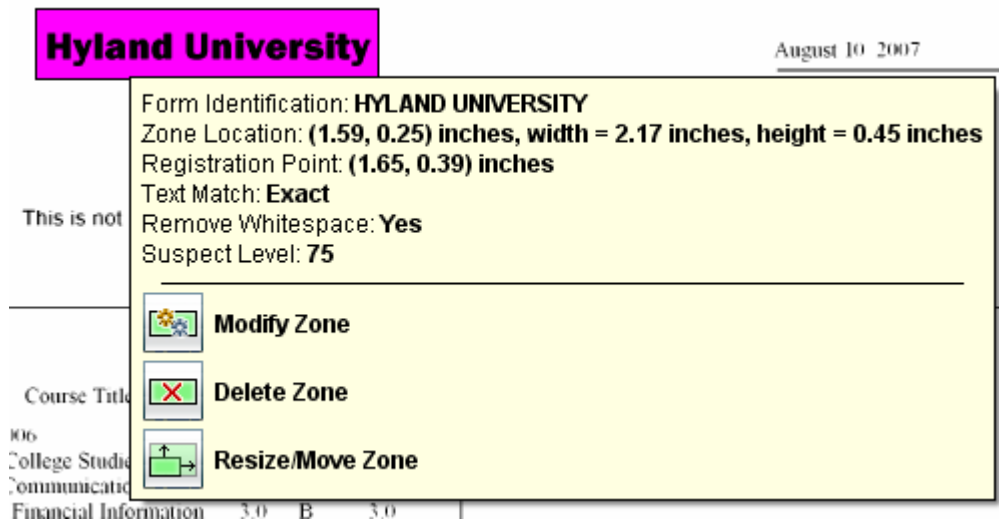
Form Identification Zone Image Match Options	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured using this field. The default Suspect Level for the Automated Indexing form is displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from that zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents.</p> <hr/>

Form Identification Zone Image Match Options	Description
Combined rule expressions	<p>In the case where multiple Form Identification Zones are needed to match a document to a form, you can configure the how these zones are processed using the Combined rule expressions option.</p> <p>Rules can be dragged and dropped into the proper order in this field and can be combined using the AND or OR boolean operators.</p> <ul style="list-style-type: none"> • All rules placed before the current (<This Identification Rule>) rule are combined using an OR operator. • All rules placed after the current (<This Identification Rule>) rule are combined using an AND operators. <p>Within the Combined rule expressions box, Form Identification Zones can be copied so that they can be used multiple times within the same rule or in other rules. To copy a Form Identification Zone, right-click it in the Combined rule expressions box and drag it to its new location in the current rule or another rule.</p> <hr/> <p>Note: Keyword rules or Identification Groups cannot be copied.</p> <hr/>
VB Script	<p>Use the VB Script drop-down select list to select a VB script to associate with the identification of this Automated Indexing form.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited. For more information on these options, contact your System Administrator.</p>
Significant Area	<p>Use the slider to control the size of the significant processing area. The significant processing area is the red box displayed on the image snippet.</p> <p>Increasing the size of the significant processing area (i.e., moving the slider to the right) will increase the accuracy of the image match process, but will decrease the speed of the process.</p> <p>Decreasing the size of the significant processing area (i.e., moving the slider to the left) will increase the speed of the image match process, but will decrease the accuracy of the process.</p>

Form Identification Zone Image Match Options	Description
Find Registration	<p>The Registration Point feature is required when using an image identifier.</p> <p>The Registration Point is the starting point (top, left corner) of the image's significant processing area (the red box displayed on the image snippet). It is used to determine the page's offset (i.e., any skew or rotation) that might have occurred during scanning.</p> <p>The position of the Registration Point on the Automated Indexing form is compared to the same position on the document being processed. The position of the configured Form Identification Zone is adjusted on the document being processed to account for the detected offset to ensure that the Automated Indexing process is able to process the document properly.</p> <p>Click Find Registration to automatically set the Registration Point, or double-click a location on the image to manually set the registration point at that position.</p>

MODIFYING, DELETING OR RESIZING/MOVING A FORM IDENTIFICATION ZONE

An existing Form Identification Zone can be modified, deleted or resized/moved by right-clicking on it in the Document Viewer. If multiple zones overlap, the **Previous** and **Next** links can be used to select the desired zone.

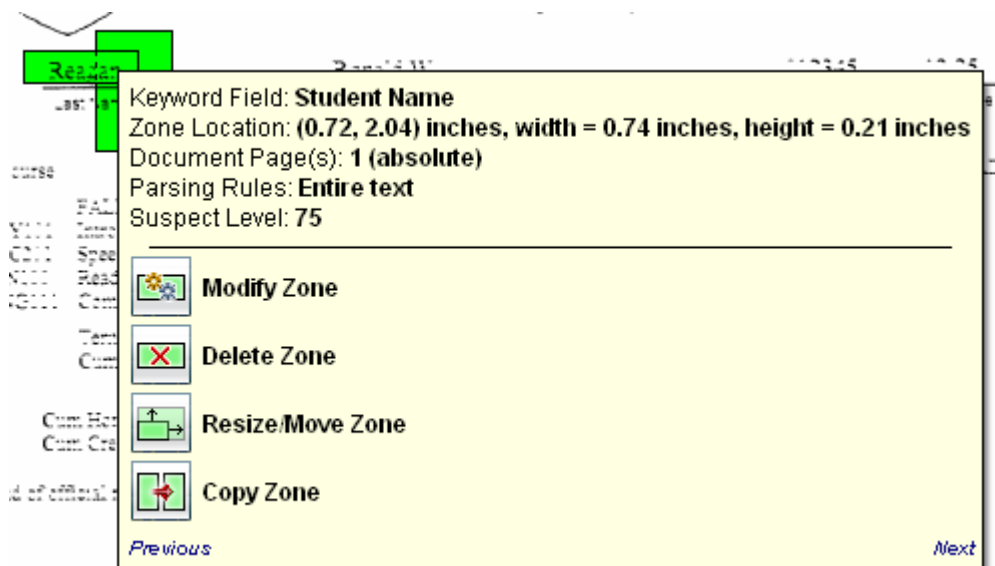


Information about the selected zone is displayed above the following options:

- **Modify Zone.** Click to display the **Form Identification Zone** dialog box to modify any of the Form Identification Zone's configuration settings.
- **Delete Zone.** Click to delete the selected Form Identification Zone.
- **Resize/Move Zone.** Click to make the Form Identification Zone editable in the Document Viewer. It can be resized and/or moved without changing any of its existing configuration settings using the pointer.

Modifying, Deleting, Resizing/Moving, or Copying Overlapping Zones

Depending on your document layout, zones may overlap. Overlapping zone borders are displayed with a dotted line.



When right-clicking on an area of the document with multiple zones, the name of the currently-selected zone is displayed at the top of the dialog box. To switch between zones, click the **Previous** and **Next** links; once the desired zone is selected, it can be modified, deleted, resize/moved, or copied.

Configuring A Data Field Zone

Data Field Zones are used to identify, extract, and assign Keyword Values to the document.

Data Field Zones can identify Keyword Values in a number of ways:

- By extracting text from the document and assigning that text as a Keyword Value.
- By performing OMR or signature detection and assigning pre-configured Keyword Values based on if a mark or a signature is present.

To create a Data Field Zone:

Note: Prior to creating a Data Field Zone, ensure that the **Default Zone Type** option in the Tools window is set to either **Always Ask** or **Keyword field**.

1. With the document that the Automated Indexing form is being created from displayed in the Document Viewer, click and hold the left mouse button and use the pointer to draw a box around the value that is to be extracted/evaluated.

3

Hyland University

August 10, 2007
(Issue Date)

Official Transcript

This is not an official transcript without Registrar's signature and college seal.

Reagan Ronald W.
(Last Name) (First Name, Middle Initial)

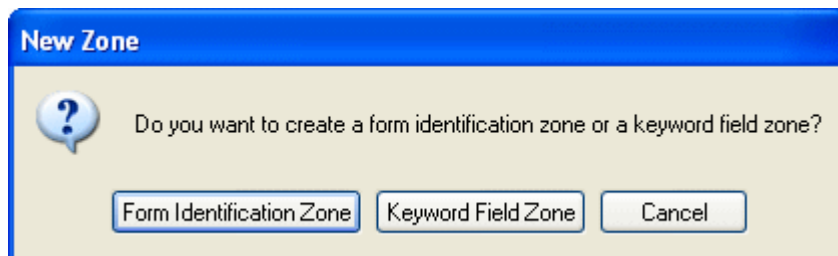
012345 12/25
(ID #) (Birthdate)

Course	Course Title	CRD	GRD	GRDPT
PSY101	FALL 2006 Intro to College Studies	0.0	W	0.000
SPC201	Speech Communication I	3.0	A-	3.5
FIN100	Reading Financial Information	3.0	B	3.0
ENG111	Composition I	3.0	A	4.0
Term GPA 2.5		Credit		0.0

Tip: Note that the area of the document that you selected is displayed in the Result Verification window.

2. Release the left mouse button to display the **Data Field Zone** dialog box.

Depending on the setting of the **Default Zone Type** option in the Tools window, the **New Zone** dialog box may be displayed.



If the **New Zone** dialog box is displayed, click **Keyword Field Zone** to continue to the **Data Field Zone** dialog box.

Data Field Zone

Grouped Line Item Extraction | **Text Field** | Optical Mark Field | Signature Detection Field

Line Item Extraction

Keyword type: **<None>**

Data Application

- ☒ Entire zone value
 - ☐ Remove line breaks
- ☐ First line only
- ☐ Table style
- ☐ Find by tag/expression
- ☐ Person name extraction

☐ Remove all whitespace from result

Page Location(s)

- ☒ Absolute page:
 - 1
 - Field only appears on a single page; always the same absolute page number (one-based)
- ☐ Relative page(s):
 - ☐ First page
 - ☐ First page (not only page)
 - ☐ Interior pages
 - ☐ Last page
 - ☐ Last page (not only page)

Filtering

Character search hints:

- ☒ Numerals
- ☒ Punctuation
- ☒ Uppercase
- ☒ Miscellaneous
- ☒ Lowercase
- ☐ MICR data
- ☐ Exclude Asian

Advanced filtering:

VB script: **<None>**

Suspect level: **75** (0-99) Form Default = 75

Specific allowed characters:


Advanced recognition: **Machine print text/OCR**

Save Cancel

3. Select the type of data the OCR engine needs to process in order to identify, extract and assign Keyword Values to the document being processed.
 - If the OCR engine is reading a text value from the image document and assigning that value as a Keyword Value, select the Text Field tab.
 - If the OCR engine is going to be reading multiple Keyword Values in a table format, select the Line Item Extraction tab.
 - If the OCR engine is going to be reading multiple Keyword Values from a single zone when line item data is separated into multiple groups on the document, select the Grouped Line Item Extraction tab.
 - If the OCR engine is going to be performing optical mark recognition and assigning a pre-configured Keyword Value based on the presence or lack of an optical mark, select the Optical Mark tab.

- If the OCR engine is going to be performing signature detection and assigning a pre-configured Keyword Value based on the presence or lack of a signature, select the Signature Detection tab.
4. Configure the Data Field Zone using the **Data Field Zone** dialog box.
 5. Once you have set all necessary options on the **Data Field Zone** dialog box, click **Save**. The **Data Field Zone** dialog box is closed, and the Data Field Zone is temporarily saved to the Automated Indexing form.

The Data Field Zone is highlighted in green on the document in the Document Viewer.



Hyland University

August 10, 2007
(Issue Date)

Official Transcript

This is not an official transcript without Registrar's signature and college seal.

Reagan

(Last Name)

Ronald W.

(First Name, Middle Initial)

012345

(ID #)

12/25

(Birthdate)

Course	Course Title	CRD	GRD	GRDPT	Course	Course Title	CRD	GRD	GRDPT
FALL 2006									
PSY101	Intro to College Studies	0.0	W	00000					
SPC201	Speech Communication I	3.0	A-	3.5					
FIN100	Reading Financial Information	3.0	B	3.0					
ENG111	Composition I	3.0	A	4.0					
Totals		GPA	3.5	Credits	12.0				

6. Repeat steps 1-5 for each Data Field Zone that you wish to create for the currently-displayed Automated Indexing form.

Tip: At any time, click **Save Configuration** in the Tools window to save the form to the database. It is considered a best practice to save your Automated Indexing form configuration after the creation and configuration of every Automated Indexing form and each form's Form Identification, Registration and Data Field Zones.

To discard all configuration changes (including all Data Field Zones created since the last time the form was saved), click **Discard Configuration Changes** in the Tools window.

CONFIGURING A TEXT DATA FIELD ZONE

A Text Data Field Zone is configured using the options on the Text Field tab of the **Data Field Zone** dialog box.

Data Field Zone

Grouped Line Item Extraction | Optical Mark Field | Signature Detection Field

Text Field | Line Item Extraction

Keyword type: <None>

Data Application

- ☒ Entire zone value
- ☐ Remove line breaks
- ☐ First line only
- ☐ Table style
- ☐ Find by tag/expression
- ☐ Person name extraction

☐ Remove all whitespace from result

Page Location(s)

☒ Absolute page: 1

Field only appears on a single page; always the same absolute page number (one-based)

☐ Relative page(s):

- ☐ First page
- ☐ First page (not only page)
- ☐ Interior pages
- ☐ Last page
- ☐ Last page (not only page)

Filtering

Character search hints:

- ☒ Numerals
- ☒ Punctuation
- ☒ Uppercase
- ☒ Miscellaneous
- ☒ Lowercase
- ☐ MICR data
- ☐ Exclude Asian

Advanced filtering:

VB script: <None>

Suspect level: 75 (0-99) Form Default = 75

Specific allowed characters:

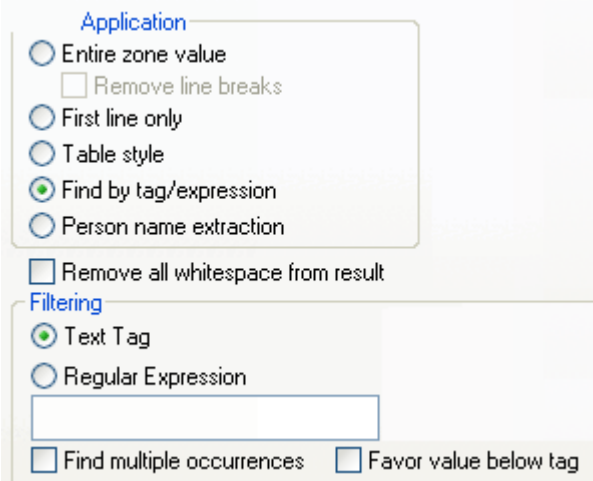
Advanced recognition: Machine print text/OCR

Save Cancel

3

Data Field Zone Keyword Field Options	Description
Keyword type	<p>Using this drop-down select list, select the Keyword Type that values extracted from the Data Field Zone are to be assigned to.</p> <p>If the currently-displayed Automated Indexing form has an assigned Document Type, only the Keyword Types associated with the assigned Document Type are available in the drop-down select list. If no Document Type is assigned to the currently-displayed Automated Indexing form, all Keyword Types are available in the drop-down select list.</p> <p>Some notes about selecting a Keyword Type:</p> <ul style="list-style-type: none"> • If the Keyword Value identified by the OCR engine exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is rejected and the document is marked as suspect. • If you are configuring a Data Field Zone to capture a Date Keyword Value, the date format of the value you are attempting to capture must match the workstation's regional format. <p>If the value's date format that does not match the workstation's regional format, then you must configure a VBscript for the Data Field Zone to parse the value into the correct format.</p> <ul style="list-style-type: none"> • If you are configuring a Data Field Zone to capture a Currency Keyword Value, the Add implied decimal point check box is displayed below the Suspect Level field. Select this check box to automatically insert a decimal point into a Currency Keyword Value if one is not found when the zone is processed. <hr/> <p>Note:When the Person name extraction Data Application option is selected, the Keyword type drop-down is replaced by the Last Name Keyword drop-down.</p> <hr/>
XML Node	<hr/> <p>Note:This field is only displayed if the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Automated Indexing form).</p> <hr/> <p>Enter the name of the XML node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p>

Data Field Zone Keyword Field Options	Description
Data Application	<p>Determine how the Data Field Zone is processed by selecting one of the Data Application radio buttons.</p> <p>Entire zone value: Select this option if all text detected in the Data Field Zone should be processed and assigned to the Keyword Value.</p> <p>If multiple lines of text are displayed in the Data Field Zone, select the Remove line breaks check box to remove the line breaks in the Keyword Value.</p> <p>First line only: Select this option if multiple lines of text may be displayed in the Data Field Zone, but only the text in the first detected line should be assigned to the Keyword Value.</p> <p>Table style: Select this option if multiple lines of text should be read as a table. Each line will be assigned as a separate instance of the Keyword.</p> <p>Multi-Instance Keyword Type Groups can be used with values read as a table to group related values together accordingly. Each Multi-Instance Keyword Type Group is constructed with Keyword Values in the order they were detected.</p> <p>For example: You have an invoice for multiple line items. You would like to extract indexing information using three Keyword Types (Course ID, Course Name and Credits) for each individual line item, and you would like to keep the Keyword Values for each line item grouped together using Multi-Instance Keyword Type Groups.</p> <p>In order to correctly identify and group the Keyword Values, you must create a separate Data Field Zone for each of the Keyword Types, and each zone must be configured to read each zone as a table. Each recognized Keyword Value will be associated with the others that compose the Multi-Instance Keyword Type Group in the order they were read (i.e., the first Class ID Keyword Value is grouped with the first Class Name and Credits Keyword Values, the second Class ID Keyword Value is grouped with the second Class Name and Credits Keyword Values, etc.)</p> <hr/> <p>Caution: In order to read Data Field Zones as a table to create Multi-Instance Keyword Type Groups, a value must exist for every Keyword Value in the Multi-Instance Keyword Type Group. If a value is missing, the subsequent values for the Keyword Type may be unintentionally assigned to the wrong Multi-Instance Keyword Type Group.</p> <hr/>

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>Find by tag/expression: Select this option if you want to identify information to be stored as Keyword Values based on a specific tag or regular expression.</p> <hr/> <p>Note:When extracting Keyword Values using text tags or regular expressions, you may configure the entire page as a Data Field Zone instead of configuring individual Data Field Zones. For more information, see Creating an Entire Page Zone for Find by Tag/Expression on page 142.</p> <hr/> <p>When this option is selected, the Text Tag and Regular Expression options are displayed in the Filtering section below, allowing you to define a Text Tag or Regular Expression to identify the Keyword Value to be extracted.</p>  <p>Select either the Text Tag or the Regular Expression radio button and enter the tag or the regular expression rule in the associated field.</p> <ul style="list-style-type: none"> • Text Tag: The OCR engine searches the zone for a tag and extracts the next logical value (either to the right of or underneath the tag) as a Keyword Value. Tags can be literal (i.e., the text specified in the text box field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and text matching that pattern is the tag). Literal text tags must be enclosed in quotation marks.

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>The tag is case insensitive. However, if the case of the tag and document value match, the suspect level of the Keyword Value is lower than it would be if they did not match.</p> <p>Select the Favor Value Below Tag check box to suggest to the OCR engine that the desired value is likely to be found below the tag rather than next to it. This lends weight to the value below the tag, but does not force the OCR engine to select it.</p> <ul style="list-style-type: none"> • Regular Expression: The OCR engine searches the zone for text that matches the regular expression rule. If text matching the rule is found, it is extracted as a Keyword Value. <p>For example: the regular expression rule for a Social Security Number would be <code>\d{3}-\d{2}-\d{4}</code> (i.e., 3 digits, dash, 2 digits, dash, 4 digits). If the OCR engine identifies text matching this pattern in the zone, that text is stored as a Keyword Value.</p> <hr/> <p>Note:All regular expressions must be ECMA compliant.</p> <hr/> <p>Select the Extract capture group only check box if you would like to use a regular expression to identify the Keyword Value, but you would only like to extract part of the identified value as a Keyword Value. The portion of the value you wish to capture must be enclosed in parenthesis in the regular expression.</p> <p>For example: A document displays a student's Social Security Number (i.e., SSN: 123-45-6789), of which you only wish to capture the last four digits (6789) as a Keyword Value. To capture this value, you would select the Extract capture group only check box and enter the following regular expression in the Regular Expression field: <code>\d{3}-\d{2}-(\d{4})</code>.</p> <p>Select the Find multiple occurrences check box if you would like to extract multiple Keyword Values from the zone (e.g., multiple values in the zone meet the regular expression rule or the tag appears multiple times in the zone and you would like to extract each instance as a Keyword Value).</p>

3

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>Person name extraction: Select this option if you want the OCR engine to identify and extract any text in the zone that is formatted like a name.</p> <p>When this option is selected, the Person Name options are displayed in the Filtering section below, allowing you to define how the name should be identified and extracted as a Keyword Value.</p> <div><div><input type="checkbox"/> Match against any letter case</div><div><div>Filtering</div><div>First Name Keyword:<div></div></div><div>Middle Name Keyword:<div></div></div><div><input type="checkbox"/> Format is Last First [Middle] with or without comma</div><div>Find after tag/expression:<div></div></div></div></div> <p>The following name formats are supported by default.</p> <ul style="list-style-type: none">• Last, First [Middle]• LAST, FIRST [MIDDLE]• First [Middle] Last• FIRST [MIDDLE] LAST <hr/> <p>Note:A middle name value is optional.</p> <hr/> <p>If the name being extracted is displayed in the Last First [Middle] format, select the Format is Last First [Middle] with or without comma check box to distinguish it from other supported formats.</p> <p>While the entire name is captured in the zone, the text is parsed into separate values (first name, middle name, and last name) and each value can be stored as an individual Keyword Value.</p> <ul style="list-style-type: none">• Use the Last Name Keyword drop-down select list to select the Keyword Type that the last name is assigned to.• Use the First Name Keyword drop-down select list to select the Keyword Type that the first name is assigned to.• Use the Middle Name Keyword drop-down select list to select the Keyword Type that the middle name is assigned to.

Data Field Zone Keyword Field Options	Description
Data Application (cont.)	<p>You may also identify a text tag or a regular expression to find the name values within the zone. When a text tag is configured, the OCR engine searches the zone for the specified text and extracts the next logical value (either to the right of or underneath the tag) as a Keyword Value. When a regular expression is configured, the text matching the pattern of the regular expression is identified as a name value.</p> <p>To configure a tag or regular expression, enter the text tag or regular expression in the Find after tag/expression: field. Text tags must be enclosed in quotation marks so not to be treated as regular expressions. Multiple tags and/or regular expressions can be configured; they must be separated by a space in the Find after tag/expression field.</p> <p>To ignore the case of the text value being extracted as a Keyword Value, select the Match against any letter case check box.</p> <hr/> <p>Tip: The Match against any letter case option should be used when the zone contains only the name and is not surrounded by other text.</p> <hr/>
Remove all whitespace	<hr/> <p>Note: This option is not displayed if the Person name extraction radio button is selected in the Data Application section.</p> <hr/> <p>Select this check box if you would like to remove all whitespace (e.g., spaces, paragraph returns, etc.) from the Keyword Value after it is read by the OCR engine.</p> <p>For example, if the Keyword Value read by the OCR engine is PSY 103 and the Remove all whitespace check box is selected, the Keyword Value is modified to PSY103.</p>

Data Field Zone Keyword Field Options	Description
Character Search Hints	<p>Note: These options are only available when Entire zone value, First line only, or Table style is selected in the Data Application section.</p> <p>These options are available to assist the OCR engine determine if the value it reads is correct. Select the check box next to each character type that is allowed to be included in the Keyword Value. The available options are:</p> <ul style="list-style-type: none"> • Numerals. Numeric characters, 0-9. • Uppercase. Uppercase alphabetic characters. • Lowercase. Lowercase alphabetic characters. • Punctuation. Punctuation marks (e.g., . ! ?). • Miscellaneous. Other ASCII characters that do not fall into one of the above categories (e.g., # \$ * @). • MICR data. Select this option only when extracting Keyword Values from a MICR line when performing Automated Indexing on check documents. <p>When a character is recognized by the OCR engine that is not part of an allowable character set, the character is replaced by a tilde (~) and the value is automatically marked as suspect.</p> <p>Tip: Using the allowed character type options can sometimes help the OCR engine more easily determine the correct value by eliminating characters that are obviously not correct (e.g., an I is correctly identified instead of a 1 because numeric characters are filtered and prevented from being recognized as part of the value).</p> <p>Select the Exclude Asian check box to instruct the OCR engine to skip Asian (i.e., double-byte) characters.</p> <p>Tip: Selecting the Exclude Asian check box allows you to identify numeric data (e.g., Date Keyword Values, Currency Keyword Values and Document Dates) when performing Automated Indexing on documents configured to contain Asian (i.e., double-byte) characters.</p>

Data Field Zone Keyword Field Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the Keyword Value being read by the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the Keyword Value is always displayed on page 1). Enter the page number that the Keyword Value is located on in the associated field. • Select the Relative page radio button if the Keyword Value may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the Data Field Zone may be located on. <ul style="list-style-type: none"> Select the First page check box if the Data Field Zone is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes. Select the First page (not only page) check box if the Data Field Zone is located on the first page and other pages in the document. Select the Interior pages check box if the Data Field Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes. Select the Last page check box if the Data Field Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes. Select the Last page (not only page) check box if the Data Field Zone is located on the last page and other pages in the document.
Specific allowed characters	<p>The Specific allowed characters field allows you to configure the OCR engine to identify only the characters that you specify in the field.</p> <p>For example, if you entered ABC in the Specific allowed characters field and the OCR engine identified the value as ABCDEFGH, the DEFG characters would be stripped from the value and the Keyword Value associated with the document would be ABC.</p>

Data Field Zone Keyword Field Options	Description
Advanced Recognition	<p>Note: This option is only enabled if your solution is licensed for Intelligent Character Recognition (ICR).</p> <p>Using this drop-down select list, select the type of processing you would like to perform on this Data Field Zone.</p> <ul style="list-style-type: none"> • Machine print text/OCR. The OCR engine will perform optical character recognition in order to read machine-printed text. This option is selected by default. • Handwriting/ICR - numerals/grouping punctuation (North American style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the North American style (e.g., the 7 character is not crossed). • Handwriting/ICR - numerals/grouping punctuation (European style). The OCR engine will perform intelligent character recognition to read hand-written text. This option will only attempt to recognize numeric characters written in the European style (e.g., the 7 character is crossed). • Handwriting/ICR - alphanumeric/punctuation. The OCR engine will perform intelligent character recognition in order to read hand-written text. This option will attempt to recognize all alphanumeric characters. <p>If one of the Handwriting/ICR options is selected, the Remove vertical/horizontal lines from zone before processing check box is displayed. Select this option to automatically strip vertical or horizontal lines (i.e., constraint boxes) from the zone before it is processed.</p>
VB script	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>

Data Field Zone Keyword Field Options	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Automated Indexing form and the actual Suspect Level detected for the selected value are displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <hr/> <p>Tip:By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that suspect Keyword Values are being consistently identified.</p> <hr/>

Data Field Zone Keyword Field Options	Description
Find in keyword dataset and correct value	<p>Note: This option is not enabled if the Person name extraction option is selected in the Data Application section. Additionally, it is only enabled if the Keyword Type selected in the Keyword type drop-down select list is configured to use a Keyword Data Set (but not a Cascading Data Set).</p> <p>Select this option if you to validate the value identified by the OCR engine using values specified for the Keyword Type's Data Set.</p> <p>If the extracted value is an exact or fuzzy match to a value found in the Data Set (i.e., the number of mismatched characters in the extracted value is less than 10% of the number of characters in the extracted value), then the value found in the Data Set is assigned as the Keyword Value.</p> <p>If the Find in keyword dataset and correct value check box is not selected, the OCR engine does not attempt to validate the extracted value against any values in the Keyword Data Set and the extracted value is assigned as the Keyword Value.</p>
Add implied decimal point	<p>Note: This check box is only enabled if a Currency Keyword Type is selected in the Keyword Type drop-down select list.</p> <p>Select this check box if you would like the OCR engine to automatically insert a decimal point in the extracted value if one is not detected.</p> <p>For example, if this option is selected and the extracted value for the Total Amount Due Keyword is \$12345, the OCR engine would automatically modify the value to \$123.45.</p> <p>If this option is selected and a decimal point is detected in the value, or if this option is not selected, then the extracted value is left as-is.</p>

Creating an Entire Page Zone for Find by Tag/Expression


Note: The **Create Entire Page Keyword Zone** should only be used in using text tags or regular expressions to identify and extract Keyword Values.


When using a tag or regular expression to identify Keyword Values, you may elect to search the entire page instead of a specific Data Field Zone.


To configure an entire page zone:

1. Right-click on the document that the Automated Indexing form is being created from in the Document Viewer to display the **Modify/Delete Form** options.

Form Name: **HYL-Hyland Unviersity**
Document Type Assigned by Form: **STDNT - Transcripts**
Document Size Range: **1 - no maximum**
Default Suspect Level: **75**
Requires Manual Indexing: **YES**
Create XML Data Rendition: **YES**

 **Modify Form**

 **Delete Form**

 **Create Entire Page Keyword Zone**

- Click **Create Entire Page Keyword Zone**. The **Data Field Zone** dialog box is displayed.

Data Field Zone

Grouped Line Item Extraction Optical Mark Field Signature Detection Field

Text Field Line Item Extraction

Keyword type: <None>

Data Application

- ☒ Entire zone value
 - ☐ Remove line breaks
- ☐ First line only
- ☐ Table style
- ☐ Find by tag/expression
- ☐ Person name extraction

☐ Remove all whitespace from result

Page Location(s)

☒ Absolute page: 1
Field only appears on a single page; always the same absolute page number (one-based)

☐ Relative page(s):

- ☐ First page
- ☐ First page (not only page)
- ☐ Interior pages
- ☐ Last page
- ☐ Last page (not only page)

Filtering

Character search hints:

- ☒ Numerals
- ☒ Punctuation
- ☒ Uppercase
- ☒ Miscellaneous
- ☒ Lowercase
- ☐ MICR data
- ☐ Exclude Asian

Advanced filtering:

VB script: <None>

Suspect level: 75 (0-99) Form Default = 75

Specific allowed characters:

Advanced recognition: Machine print text/OCR

Save Cancel

- Ensure that the Text Field tab is displayed.
- Create a Text Field Data Zone, ensuring that the you are using the **Find by tag/expression** option in the Data Application section.
- When finished, click **Save**.

CONFIGURING A LINE ITEM EXTRACTION FIELD ZONE

A Line Item Extraction Field Zone allows you to extract multiple Keyword Values from a single zone when the values to be extracted are arranged in line item format using rows and columns. Values in each column of the table are assigned to the same Keyword Type.

Tip: Line Item Extraction Field Zones are intended to be used with Multi-Instance Keyword Type Groups in order to capture line item data from while maintaining the Keyword Values' relationship to one another.

Note: Keyword Values can also be extracted from line-item-formatted data into a Multiple Instance Keyword Type Group using multiple Text Data Field Zones (one for each column of data) set to use the **Table style** Data Application option.

Line Item Extraction Zones have two modes:

- **Automatic Table Decomposition.** The OCR engine attempts to automatically detect the location of individual table columns within the Data Field Zone.
- **Manual Table Decomposition.** Individual table columns are manually defined when configuring the Data Field Zone.

Both Line Item Extraction modes are configured using the options on the Line Item Extraction tab of the **Data Field Zone** dialog box. Once the table decomposition mode-specific options are set, you are able to set other, non-mode specific configuration options.

Data Field Zone

Grouped Line Item Extraction | Optical Mark Field | Signature Detection Field

Text Field | **Line Item Extraction**

☒ Use line item style data extraction XML Node (Parent):

Column	Keyword Type	Req	Expression	XML Node

Begin table: End table: Invalid row:

☐ Inclusive ☐ Inclusive

Page Location(s)

☒ Absolute page:
Field only appears on a single page; always the same absolute page number (one-based)

☐ Relative page(s):

- ☐ First page
- ☐ First page (not only page)
- ☐ Interior pages
- ☐ Last page
- ☐ Last page (not only page)

Options

☐ Manual table decomposition mode

☐ Automatic continued-row merging

VB script:

Suspect level:
(0-99) Form Default = 75

Save Cancel

Configuring Line Item Extraction in Automatic Table Decomposition Mode

1. Select the **Use line item style data extraction** check box. The remaining options on the tab are enabled.
2. Use the **Keyword type** drop-down select list (located below the Column and Keyword Type columns) to select the Keyword Type that values in the first column of the table are assigned to as Keyword Values.

To ignore a column or capture it as XML data only, select **<None>** from the drop-down select list.

3. If this Keyword Value is required, select the **Required** check box (located to the right of the Keyword type drop-down select list). If a Keyword Value that is marked as required is missing, the data from the entire row is discarded.
4. *Optional.* To specify a regular expression rule for the extracted text, enter the rule in the **Expression** field (located next to the **Required** check box). The OCR engine will compare the extracted text to the defined regular expression rule; if the text is a match, the value is stored as a Keyword Value. If the text does not match, it is discarded.

For example, if you specify the following regular expression rule:

[[[:upper:]][:lower:]][:digit:]][:space:]]+

Any value containing a character that is not a letter, number, or space is discarded.

Note: All regular expressions must be ECMA compliant.

3

Tip: Regular expressions can be used to discard column data that consists of entirely unwanted characters (e.g., a **123***456**, where you want to capture **123** and **456** as separate Keyword Values and discard the asterisk “separator” column). However, if the unwanted characters are located in the middle of valid data (e.g., **123***456**, where you want to capture **123456** as one Keyword Value), you can configure a Keyword Lookup/Replace dictionary entry to replace the data.

5. Click **Add**.
6. Repeat Steps 2-5 for all columns in the table. Columns must be added to the list as they appear in the document from left-to-right in order for Keyword Values to be extracted correctly.

Select a configured column and use the **Move Up** and **Move Down** buttons to re-arrange it as needed. Select a configured column and click **Delete** to delete the column configuration.
7. Click **Save**.

Configuring Line Item Extraction in Manual Table Decomposition Mode

1. Select the **Use line item style data extraction** check box. The remaining options on the tab are enabled.
2. In the Options section select the **Manual table decomposition mode** check box. A preview of the Data Field Zone is displayed to the right of the **Data Field Zone** dialog box.

3. *Optional.* Select the **Automatic continued-row merging** check box.

When this option is selected, the OCR engine compares the average spacing between rows. If a row is closer than average to the row that precedes it and the row does not have data in at least half of its processed columns, the OCR engine recognizes it as a continuation of the previous row and merges the extracted values.

4. In the preview of the Data Field Zone displayed to the right of the **Data Field Zone** dialog box, click and hold the left mouse button and use the pointer to draw a box around the first column of data to be extracted.

The **Modify Keyword Result for Column** dialog box is displayed.

5. Use the **Keyword type** drop-down select list to select the Keyword Type that values in the selected column are assigned to as Keyword Values.

To ignore the data in the selected column or capture it as XML data only, select **<None>** from the **Keyword type** drop-down select list.

6. *Optional.* To specify a regular expression rule for the extracted text, enter the rule in the **Expression to Match** field. The OCR engine will compare the extracted text to the defined regular expression rule; if the text is a match, the value is stored as a Keyword Value. If the text does not match, it is discarded.

For example, if you specify the following regular expression rule:

[[[:upper:]][:lower:]][:digit:]][:space:]]+

Any value containing a character that is not a letter, number, or space is discarded.

Note: All regular expressions must be ECMA compliant.

Tip: Regular expressions can be used to discard column data that consists of entirely unwanted characters (e.g., a **123***456**, where you want to capture **123** and **456** as separate Keyword Values and discard the asterisk “separator” column). However, if the unwanted characters are located in the middle of valid data (e.g., **123***456**, where you want to capture **123456** as one Keyword Value), you can configure a Keyword Lookup/Replace dictionary entry to replace the data.

7. If the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the Automated Indexing form), the **XML Node** drop-down select list is displayed.

Use the **XML Node** drop-down select list to select the XML node (i.e., the element) that Keyword Values extracted from this column are contained in when the XML rendition of the document is created.

8. If this Keyword Value is required, select the **Column Data Required** check box. If a Keyword Value that is marked as required is missing, the data from the entire row is discarded.
9. Click **Save**. The Column Extraction Zone is highlighted in the Data Field Zone preview.
10. Repeat Steps 4-9 for any remaining columns in the Data Field Zone preview whose data you wish to capture.

Additional Line Item Extraction Configuration Options

Once you have configured the Line Item Extraction Field Zone’s table decomposition mode, you are able to set other, non-mode specific configuration options.

Data Field Zone Line Item Extraction Option	Description
XML Node (Parent)	<p>Note: This field is only displayed if the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Automated Indexing form).</p> <p>Enter the name of the XML parent node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p>

Data Field Zone Line Item Extraction Option	Description
Begin table marker	<p>Use the Begin table marker field to specify the text that indicates the beginning of the information being extracted.</p> <p>The markers can be configured as literal text (i.e., the text specified in the Begin Table Marker field is the marker) or regular expressions (i.e., the text in the Begin Table Marker field defines a regular expression, and text matching that pattern is the marker).</p> <p>Literal markers must be enclosed in quotation marks to differentiate them from the regular expression markers.</p> <p>Multiple begin table markers (either literal, regular expression, or both) can be configured by separating the text by a space in the Begin table marker field.</p> <p>By default, text used as a Begin table marker is not included in the Keyword Value(s) being extracted. To include this text in the Keyword Vale, select the Inclusive check box beneath the Begin table marker field.</p>
End table marker	<p>Use the End table marker field to specify the text that indicates the end of the information being extracted.</p> <p>The markers can be configured as literal text (i.e., the text specified in the End Table Marker field is the marker) or regular expressions (i.e., the text in the End Table Marker field defines a regular expression, and text matching that pattern is the marker).</p> <p>Literal markers must be enclosed in quotation marks to differentiate them from the regular expression markers.</p> <p>Multiple end table markers (either literal, regular expression, or both) can be configured by separating the text by a space in the End table marker field.</p> <p>By default, text used as an End table marker is not included in the Keyword Value(s) being extracted. To include this text in the Keyword Vale, select the Inclusive check box beneath the End table marker field.</p>

Data Field Zone Line Item Extraction Option	Description
Invalid Row	<p>Use the Invalid Row field to specify text that indicates the beginning of the row that should be discarded</p> <p>The markers can be configured as literal text (i.e., the text specified in the Invalid Row field is the marker) or regular expressions (i.e., the text in the Invalid Row field defines a regular expression, and text matching that pattern is the marker).</p> <p>Literal markers must be enclosed in quotation marks to differentiate them from the regular expression markers.</p> <p>Multiple invalid row markers (either literal, regular expression, or both) can be configured by separating the text by a space in the Invalid Row field.</p>

Data Field Zone Line Item Extraction Option	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the table being read in the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the table is always displayed on page 1). Enter the page number that the table is located on in the associated field. • Select the Relative page radio button if the table may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the table may be located on. <p>Select the First page check box if the table is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the table is located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the table is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the table is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the table is located on the last page and other pages in the document.</p>
VB script	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>

Data Field Zone Line Item Extraction Option	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Automated Indexing form and the actual Suspect Level detected for the selected table are displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that suspect Keyword Values are being consistently identified.</p>

CONFIGURING A GROUPED LINE ITEM EXTRACTION DATA FIELD ZONE

A Grouped Line Item Extraction Data Field Zone allows you to extract Keyword Values from multiple groups, or tables, within a single zone.

Keyword Values in each group can be identified by a tag (either a literal tag or a regular expression), regular expression, or by columns in line item data.

Tip: Grouped Line Item Extraction Data Field Zones are intended to be used with Multi-Instance Keyword Type Groups in order to capture data from multiple tables while maintaining the Keyword Values' relationship to one another.

For example:

Fall 2009				
BIO111	Biology 1	B	3.0	
ENG212	Composition 1	A-	3.0	
ECO211	Micro Econ 1	C	3.0	
PHS111	Ice Skating	W	0.0	
TOTAL			9.0	

Spring 2010				
PSY 101	Psychology 1	B	3.0	
ENG 312	Lit Analysis	A	3.0	
MTH 151	Calculus 1	C	3.0	
TOTAL			9.0	

Here you have two groups of course data organized by term. For each course, you want to identify five pieces of information: the term the course was taken, the Course ID, the Course Description, Grade and Credits). This course information is to be stored in a Multi-Instance Keyword Type Group; each instance represents information about one course the student has completed.

In this example, the **Term** Keyword Value is identified by a regular expression, and the remaining course information is identified by the column that it is displayed in.

A Grouped Line Item Data Field Zone is configured using the options on the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box.

Data Field Zone

Text Field | Line Item Extraction

Grouped Line Item Extraction | Optical Mark Field | Signature Detection Field

☒ Use grouped line item style data extraction XML Node (Parent):

Column	Keyword Type	Req	Expression	XML Node

Keyword Type	Tag/Expression ...	Req	Position	XML Node

New Tag
Modify
Delete

Page Location(s)

☒ Absolute page: 1
Field only appears on a single page; always the same absolute page number (one-based)

☐ Relative page(s):
☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)

Options

☐ Automatic continued-row merging
☐ Extract only between tags
 VB script: <None> ...
 Suspect level: 75
 (0-99) Form Default = 75, Actual = 70

Save Cancel

Unlike Line Item Extraction Zones, Grouped Line Item Extraction Zones use only manual table decomposition.

To configure a Grouped Line Item Extraction Zone, select the **Use line item style data extraction** check box. The remaining options on the tab are enabled and a preview of the Data Field Zone is displayed to the right of the **Data Field Zone** dialog box.

There are several steps required to configure a Grouped Line Item Extraction Zone:

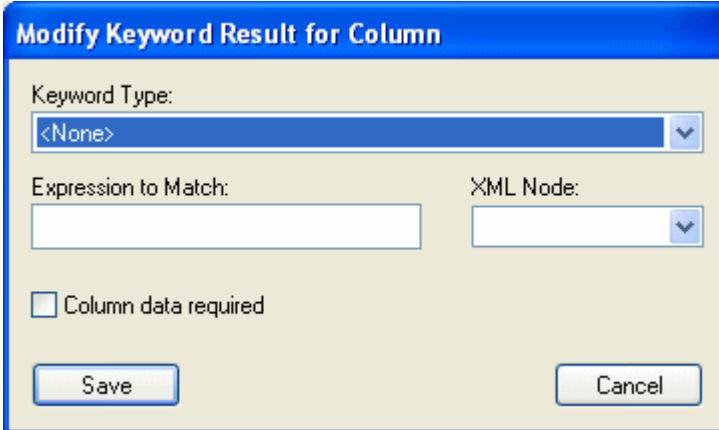
1. **Specify How Line Item Data is to be Extracted as Keyword Values** Each column in the group is mapped to a Keyword Type.
2. **Specify How Data is to be Extracted as Keyword Values Using Tags or Regular Expressions.** Data identified by a tag (tags can be literal text or regular expressions) or a regular expression can be assigned as Keyword Values to specified Keyword Types.
3. **Configure the Additional Zone Configuration Options.** Once you have specified how the OCR engine should address the data in the group, you can specify other configuration options for the zone (e.g., page locations, VB scripts, Suspect Level, etc.).

Identifying Keyword Values from Line Item Data

To identify Keyword Values from line item data:

1. In the preview of the Data Field Zone displayed to the right of the **Data Field Zone** dialog box, click and hold the left mouse button and use the pointer to draw a box around a column of data to be extracted as Keyword Values.

The **Modify Keyword Result for Column** dialog box is displayed.



2. Use the **Keyword Type** drop-down select list to select the Keyword Type that values in the selected column are assigned to as Keyword Values.

To ignore the data in the selected column or capture it as XML data only, select **<None>** from the **Keyword type** drop-down select list.

3. *Optional.* To specify a regular expression rule for the extracted text, enter the rule in the **Expression to Match** field. The OCR engine will compare the extracted text to the defined regular expression rule; if the text is a match, the value is stored as a Keyword Value. If the text does not match, it is discarded.

For example, if you specify the following regular expression rule:

[[[:upper:]][:lower:]][:digit:]][:space:]]+

Any value containing a character that is not a letter, number, or space is discarded.

Note: All regular expressions must be ECMA compliant.

Tip: Regular expressions can be used to discard column data that consists of entirely unwanted characters (e.g., a **123***456**, where you want to capture **123** and **456** as separate Keyword Values and discard the asterisk “separator” column). However, if the unwanted characters are located in the middle of valid data (e.g., **123***456**, where you want to capture **123456** as one Keyword Value), you can configure a Keyword Lookup/Replace dictionary entry to replace the data.

4. If the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the Automated Indexing form), the **XML Node** drop-down select list is displayed.

Use the **XML Node** drop-down select list to select the XML node (i.e., the element) that Keyword Values extracted from this column are contained in when the XML rendition of the document is created.

5. If this Keyword Value is required, select the **Column Data Required** check box. If a Keyword Value that is marked as required is missing, the data from the entire row is discarded.
6. Click **Save**. The Column Extraction Zone is highlighted in the Data Field Zone preview.
7. Repeat Steps 1-6 for any remaining columns in the Data Field Zone preview whose data you wish to capture.

Identifying Keyword Values Using Tags or Regular Expressions

To identify Keyword Values using tags (literal or regular expression) or regular expressions:

1. From the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box, click **New Tag** to create the text tag or regular expression to be used to identify the Keyword Value. The **Modify Line Item Tagged Value** dialog box is displayed.

Modify Line Item Tagged Value

Keyword Type:
<None>

Select the desired radio button on the left to activate the associated entry fields on the right:

Tag Text:
☒ [Text Field]

Regular Expression Match:
☐ [Text Field]
☐ Extract capture group only

Person Name Extraction:
☐ [Text Field]
First Name: [Dropdown] Middle Name: [Dropdown]

Meta Data Value:
☐ [Dropdown]

Line Item Totals Marker:
☐ [Text Field]
☐ No tag value/always last row in group

☒ This item PRECEDES the associated line item group
☐ This item FOLLOWS the associated line item group
☐ Treat this row as a line item after reading this tagged value
☐ The position of this item is not significant

☐ Data value required XML Node: [Dropdown]

Save Cancel

2. Using the **Keyword Type** drop-down select list, select the Keyword Type that the value you are identifying is to be assigned to.

If you only want to capture this value as XML data (not as a Keyword Value), select **<None>**.
3. Select the radio button that describes how the OCR engine will identify the Keyword Value:
 - **Tag Text.** The OCR engine searches the Data Field Zone for a tag (e.g., **Name:**, **Date:**, etc.) and identifies the value to the right of the tag as the Keyword Value.

Note: When identifying a Keyword Value, a tag can only be used to identify the value to the right of the tag. A tag cannot be used to identify the value below the tag.

3

Enter the tag text in the **Tag Text** field. Tags can be literal (i.e., the text specified in the field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and the text matching that pattern is the tag).

Literal tags must be enclosed in quotation marks in order to differentiate them from regular expression tags. Regular expressions must be ECMA compliant.

- **Regular Expression Match.** The OCR engine searches the zone for text that matches the regular expression rule. If text matching the pattern set by the regular expression rule is found, it is identified as the Keyword Value.

Enter the regular expression in the **Regular Expression Match** field. All regular expressions must be ECMA compliant.

- Select the **Extract Capture Group Only** check box if you would to use a regular expression to identify the Keyword Value, but you would only like to extract part of the value identified by the regular expression as a Keyword Value. The portion of the Keyword Value you wish to capture must be enclosed in parenthesis in the regular expression.

For example: A document displays a student's Social Security Number (i.e., **SSN: 123-45-6789**), of which you only wish to capture the last four digits (**6789**) as a Keyword Value. To capture this value, you would select the **Extract Capture Group Only** check box and enter the following regular expression in the **Regular Expression Match** field: **\d{3}-\d{2}-(\d{4})**.

Additionally, you can use the **Extract Capture Group Only** option to capture only one value from a row of line item data. For example: a document contains a **Totals** line that contains tallied data for five columns, of which you only want to capture the Total value from the third column (i.e., **\$310.71**) as a Keyword Value.

TOTAL \$101.35 \$177.01 \$310.71 \$285.00 \$526.22

To capture this value, you would select the **Extract Capture Group Only** check box and enter the following regular expression in the **Regular Expression Match** field: **TOTAL\s+\\$\d{3}.\d{2}\s+\\$\d{3}.\d{2}\s+(\\$\d{3}.\d{2})\s+\\$\d{3}.\d{2}\s+\\$\d{3}.\d{2}**

- **Person Name Extraction.** Similar to **Tag Text** option, the OCR engine searches the Data Field Zone for a literal tag or regular expression tag; however, once the tag is identified, the value following the tag is stored as a name instead of a single Keyword Value.

When a name is stored, the extracted value is parsed into separate values (first name, middle name, and last name) and each value is stored separately as an individual Keyword Value.

The following are valid name formats (middle name values are not required):

- **Last, First [Middle]**
- **LAST, FIRST [MIDDLE]**
- **First [Middle] Last**
- **FIRST [MIDDLE] LAST**

To select the Keyword Types that the name values are assigned to:

- Use the **Last Name Keyword** drop-down select list (renamed from **Keyword Type**) to select the Keyword Type that the last name is assigned to.
- Use the **First Name** drop-down select list to select the Keyword Type that the first name is assigned to.
- Use the **Middle Name** drop-down select list to select the Keyword Type that the middle name is assigned to.

Enter the tag text in the **Person Name Extraction** field. Tags can be literal (i.e., the text specified in the field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and the text matching that pattern is the tag).

Literal tags must be enclosed in quotation marks in order to differentiate them from regular expression tags. Regular expressions must be ECMA compliant.

- **Meta Data Value.** The OCR engine identifies the Keyword Value using data obtained when processing the document. Use the **Meta Data Value** drop-down select list to select the data to be used as the Keyword Value:
 - **<Group Number>.** This option refers to the order of the groups in the zone. For example, if a zone contains three groups (e.g., **Fall 2009, Spring 2010, Fall 2010**), then Group Number for the first group (**Fall 2009**) is **1**, the Group Number for the second group (**Spring 2010**) is **2**, and the Group Number for the third group (**Fall 2010**) is **3**.
 - **<Line Number>.** This option refers to the line on which the data is displayed once the zone is processed. The first line at the top of the zone is 1, the second line is 2, etc.
 - **<None>.** This option is reserved for future use.
 - **<Page Number>.** This option refers to the page on which the data is displayed.
- **Line Item Totals Marker.** The OCR engine identifies the Keyword Value as the last line in the group (i.e., a “totals” line where column data is tallied).

Enter the tag text in the **Line Item Totals** field. Tags can be literal (i.e., the text specified in the field is the tag) or regular expressions (i.e., the text in the text box field defines a regular expression, and the text matching that pattern is the tag).

Select **No tag value/always last line in group** check box if the data is not denoted by a tag, but should be captured because it is the last line in the group.

4. Select the radio button that describes the location of tag or regular expression in the group.
 - **This item PRECEDES the associated line item group.** Select this radio button if the Keyword Value is located before the line item data.
 - **This item FOLLOWS the associated line item group.** Select this radio button if the Keyword Value is located after the line item data.
 - **Treat this row as a line item after reading this tagged value.** Select this radio button if the line item containing the Keyword Value should be also be treated as a row of line item data.
 - **The position of this item is not significant.** Select this radio button if the location of the Keyword Value is not significant in the zone.
5. Select the **Data Value required** check box if the Keyword Value is a required value. If this check box is selected and no value is found, then the entire group is discarded.
6. If the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the **Create XML data rendition** option is selected for the Automated Indexing form), the **XML Node** drop-down select list is displayed.

Use the **XML Node** drop-down select list to select the XML node (i.e., the element) that Keyword Value is contained in when the XML rendition of the document is created.

7. Click **Save**. The **Modify Line Item Tagged Value** dialog box is closed and you are returned to the Grouped Line Item Extraction tab of the **Data Field Zone** dialog box.

Configuring Additional Group Line Item Extraction Options

The following configuration options for the zone can be configured in the Data Field Zone dialog box.

Data Field Zone Grouped Line Item Extraction Option	Description
XML Node (Parent)	<div>Note: This field is only displayed if the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Automated Indexing form).</div> <div>Enter the name of the XML parent node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</div>

Data Field Zone Grouped Line Item Extraction Option	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the groups are only displayed on one page and is always displayed on the same page (e.g., the groups are always displayed on page 1). Enter the page number that the groups are located on in the associated field. • Select the Relative page radio button if the groups may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the groups may be located on. <p>Select the First page check box if the groups are located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the groups are located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the groups are located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the groups are located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the groups are located on the last page and other pages in the document.</p>
Automatic continued-row merging	<p>Select the Automatic continued-row merging check box to compare the average spacing between rows.</p> <p>If a row is closer than average to the row that precedes it and the row does not have data in at least half of its processed columns, the OCR engine recognizes it as a continuation of the previous row and merges the extracted values.</p>

Data Field Zone Grouped Line Item Extraction Option	Description
Extract only between tags	<p>Select the Extract only between tags check box if all Keyword Values to be extracted can be found between two tags (i.e., a start marker and an end marker).</p> <hr/> <p>Tip: This option should be selected when your groups contain “extra” data that can be ignored by the OCR engine because it does not contain Keyword Values.</p> <hr/>
VB script	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>

Data Field Zone Grouped Line Item Extraction Option	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Automated Indexing form is displayed below the Suspect level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that suspect Keyword Values are being consistently identified.</p> <hr/>

CONFIGURING AN OPTICAL MARK DATA FIELD ZONE

An OMR Data Field Zone is configured using the options on the Optical Mark Field tab of the **Data Field Zone** dialog box.

The screenshot shows the 'Data Field Zone' dialog box with the 'Optical Mark Field' tab selected. The dialog is divided into several sections:

- Text Field** and **Line Item Extraction** tabs are at the top.
- Grouped Line Item Extraction**, **Optical Mark Field**, and **Signature Detection Field** sub-tabs are below the main tabs.
- Keyword type:** A dropdown menu set to 'Grade'.
- XML Node:** A dropdown menu.
- OMR item group mode:** An unchecked checkbox.
- Keyword Values:** Two text input fields for 'Positive result value' and 'Negative result value'.
- Page Location(s):** A section with two radio buttons: 'Absolute page:' (selected) and 'Relative page(s):'. The 'Absolute page:' option has a text input field containing '1' and a descriptive text: 'Field only appears on a single page; always the same absolute page number (one-based)'. The 'Relative page(s):' option has five checkboxes: 'First page', 'First page (not only page)', 'Interior pages', 'Last page', and 'Last page (not only page)'. All are unchecked.
- Options:** A section with two groups of radio buttons: 'Mark frame detection:' with 'Framed', 'Unframed', and 'Auto-detect' (selected); and 'Sensitivity:' with 'Highest' (selected), 'Low', 'Lower', and 'Lowest'. There is also an unchecked checkbox for 'Support 'filled-in-error' detection'.
- Advanced filtering:** A section with a 'VB script:' dropdown menu set to '<None>' and a 'Test' button.
- Buttons:** 'Save' and 'Cancel' buttons at the bottom.

Data Field Zone Optical Mark Field Options	Description
Keyword type	<p>Using this drop-down select list, select the Keyword Type that this Data Field Zone will assign a Keyword Value to.</p> <p>If the currently-displayed Automated Indexing form has an assigned Document Type, only the Keyword Types associated with the assigned Document Type are available in the drop-down select list. If no Document Type is assigned to the currently-displayed Automated Indexing form, all Keyword Types are available in the drop-down select list.</p> <hr/> <p>Note: If the Keyword Value identified by the OCR engine exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is rejected and the document is marked as suspect.</p> <hr/>
XML Node	<hr/> <p>Note: This field is only displayed if the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Automated Indexing form).</p> <hr/> <p>Enter the name of the XML node (i.e., the element) that a Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p>

Data Field Zone Optical Mark Field Options	Description
OMR item group mode	<p>Note: To use this feature, multiple optical marks must be reside in the Data Field Zone.</p> <p>Select this check box to allow the OCR engine to process multiple optical marks inside the Data Field Zone. Each optical mark must share all of the same configuration options (e.g., Keyword Type, Framed/Unframed, Sensitivity level, etc.) except for their associated positive and negative Keyword Values.</p> <p>When this check box is selected, the Keyword Values section of the Data Field Zone dialog box is displayed as a list containing Item #, Positive Value, Negative Value and an image snippet of the area of the document selected for the Data Field Zone is displayed next to the Data Field Zone dialog box.</p> <p>From the image snippet, use the pointer to draw a box around the first optical mark in the Data Field Zone. The Modify Keyword Result for OMR Group Item dialog box is displayed, allowing you to specify the positive and negative Keyword Value for that optical mark.</p> <p>If the Automated Indexing form is configured to create XML renditions of the documents it is matched to, the XML Node field is displayed, giving you the opportunity to name the XML node that the Keyword Values are contained in when the XML rendition is created.</p> <p>When finished, click Save. The optical mark configuration zone is highlighted on the image snippet. Right-click on the optical mark configuration within the image snippet to move, resize or delete the configuration information for that optical mark configuration zone.</p> <p>Repeat this process for each optical mark displayed in the Data Field Zone.</p>

Data Field Zone Optical Mark Field Options	Description
Keyword Values	<p>In the Positive Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that an optical mark is present.</p> <p>In the Negative Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that an optical mark is not present.</p> <p>For example: you are processing an application and one of the questions asks the applicant to select a check box if he/she has previously applied to the university. Depending on the answer to this question, a Keyword Value is assigned to the Previously Applied Keyword Type.</p> <ul style="list-style-type: none">• If the check box is selected, Yes is assigned as the Previously Applied Keyword Value.• If the check box is not selected, No is assigned as the Previously Applied Keyword Value.

Data Field Zone Optical Mark Field Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the data being read in the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the optical mark is always displayed on page 1). Enter the page number that the optical mark is located on in the associated field. • Select the Relative page radio button if the data may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the optical mark may be located on. <ul style="list-style-type: none"> Select the First page check box if the optical mark is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes. Select the First page (not only page) check box if the optical mark is located on the first page and other pages in the document. Select the Interior pages check box if the optical mark is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes. Select the Last page check box if the optical mark is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes. Select the Last page (not only page) check box if the optical mark is located on the last page and other pages in the document.

Data Field Zone Optical Mark Field Options	Description
Mark Frame Detection	<p>Select the radio button that describes the optical mark being read by the OCR engine.</p> <ul style="list-style-type: none">• Framed. Select the Framed radio button if the optical mark is located in a pre-defined space, such as a check box or bubble.• Unframed. Select the Unframed radio button if the optical mark is located in an undefined or blank space.• Auto-detect. Select the Auto-detect radio button if the OCR engine should determine if the optical mark is framed or unframed. <hr/> <p>Tip: Selecting the Framed or Unframed selection, where appropriate, will help to increase the accuracy of the Automated Indexing process.</p> <hr/>

Data Field Zone Optical Mark Field Options	Description
Sensitivity	<p>Select a radio button to determine the Suspect Level for this Data Field Zone.</p> <ul style="list-style-type: none"> • Highest. Sets the Suspect Level for this Data Field Zone to 70. • Low. Sets the Suspect Level for this Data Field Zone to 75. • Lower. Sets the Suspect Level for this Data Field Zone to 80. • Lowest. Sets the Suspect Level for this Data Field Zone to 85. <p>The Suspect Level is the level of confidence placed in the processing results for this field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The Sensitivity level selected for this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Sensitivity to Lowest would indicate you have a fair amount of confidence in the result returned by the OCR engine because few higher scores could be returned and fewer results would be determined to be suspect.</p> <p>Setting the Suspect Level to Highest would indicate you have less confidence in the result because a great number of lower scores could be returned and more results would be determined as suspect.</p>

Data Field Zone Optical Mark Field Options	Description
Support 'filled-in error' detection	<p>Note: This option is enabled only when the Framed or Auto-detect radio button is selected in the Options section.</p> <p>Select this check box to enable optical mark "filled in error" detection.</p> <p>When this option is selected, the OCR engine will attempt to determine if a user has "crossed out" an optical mark on the document in order to indicate it is not present.</p> <p>For example, users occasionally select a check box and later realize that they made an error (i.e., they meant to leave the check box unselected), so they attempt to cross it out. This option allows the OCR engine to attempt to distinguish between a selected check box and a crossed-out check box.</p> <p>If the OCR engine determines that a field has been marked in error (i.e., crossed out), the field is marked as suspect.</p> <p>If this option is not selected, the OCR engine will not attempt to discern the difference between a selected check box and a crossed-out check box.</p>
Advanced filtering	<p>Use the VB script drop-down to select a VBscript to associate with the processing of this Data Field Zone.</p> <p>Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.</p>
Test	<p>Click the Test button to have the OCR engine perform a test process on the Data Field Zone and attempt to detect an optical mark using the options you have specified.</p> <p>Once the test process is complete, a dialog box is displayed indicating if an optical mark was detected and the Keyword Value that would be assigned based on this result.</p>

CONFIGURING A SIGNATURE DETECTION DATA FIELD ZONE

A Signature Detection Data Field Zone is configured using the options on the Signature Detection Field tab of the **Data Field Zone** dialog box.

The screenshot shows the 'Data Field Zone' dialog box with the 'Signature Detection Field' tab selected. The 'Keyword type' is set to 'Status'. The 'XML Node' is empty. Under 'Keyword Values', there are empty text boxes for 'Positive result value' and 'Negative result value'. Under 'Page Location(s)', 'Absolute page' is selected with a value of '1'. The text 'Field only appears on a single page; always the same absolute page number (one-based)' is displayed. Under 'Options', 'Signature is framed/underlined' is selected. Sensitivity is set to 'Highest'. The 'Exclude machine print text' checkbox is unchecked. The 'Advanced filtering' section shows 'VB script' set to '<None>'. There are 'Test', 'Save', and 'Cancel' buttons.

Data Field Zone

Text Field | Line Item Extraction

Grouped Line Item Extraction | Optical Mark Field | **Signature Detection Field**

Keyword type: **Status** XML Node:

Keyword Values

Positive result value:

Negative result value:

Page Location(s)

☒ Absolute page: ☐ Relative page(s):

Field only appears on a single page; always the same absolute page number (one-based)

☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)

Options

☒ Signature is framed/underlined
☐ Signature area is unframed

Sensitivity: ☒ Highest ☐ Low ☐ Lower ☐ Lowest

☐ Exclude machine print text

Test

Advanced filtering:
VB script: <None>

Save Cancel

3

Data Field Zone Signature Detection Field Options	Description
Keyword type	<p>Using this drop-down select list, select the Keyword Type that this Data Field Zone will assign a Keyword Value to.</p> <p>If the currently-displayed Automated Indexing form has an assigned Document Type, only the Keyword Types associated with the assigned Document Type are available in the drop-down select list. If no Document Type is assigned to the currently-displayed Automated Indexing form, all Keyword Types are available in the drop-down select list.</p> <hr/> <p>Note: If the Keyword Value identified by the OCR engine exceeds the maximum length of the Keyword Type it is assigned to, then the Keyword Value is rejected and the document is marked as suspect.</p> <hr/>
XML Node	<hr/> <p>Note: This field is only displayed if the Automated Indexing form is configured to create an XML rendition of documents matched to it (i.e., the Create XML data rendition option is selected for the Automated Indexing form).</p> <hr/> <p>Enter the name of the XML node (i.e., the element) that Keyword Values extracted from this zone are contained in when the XML rendition of the document is created.</p>
Keyword Values	<p>In the Positive Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is present.</p> <p>In the Negative Result Value field, enter the Keyword Value that is assigned if the OCR engine determines that a signature is not present.</p> <p>For example: you are processing an application that requires a signature to indicate that the data has been verified. Depending on if a signature is present or not, a Keyword Value is assigned to the Verified Keyword Type.</p> <ul style="list-style-type: none"> • If a signature is detected, Yes is assigned as the Verified Keyword Value. • If a signature is not detected, No is assigned as the Verified Keyword Value.

Data Field Zone Signature Detection Field Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Data Field Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the data being read in the Data Field Zone is only displayed on one page and is always displayed on the same page (e.g., the signature is always displayed on page 1). Enter the page number that the optical mark is located on in the associated field. • Select the Relative page radio button if the data may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the signature may be located on. <p>Select the First page check box if the signature is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the signature is located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the signature is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the signature is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the signature is located on the last page and other pages in the document.</p>

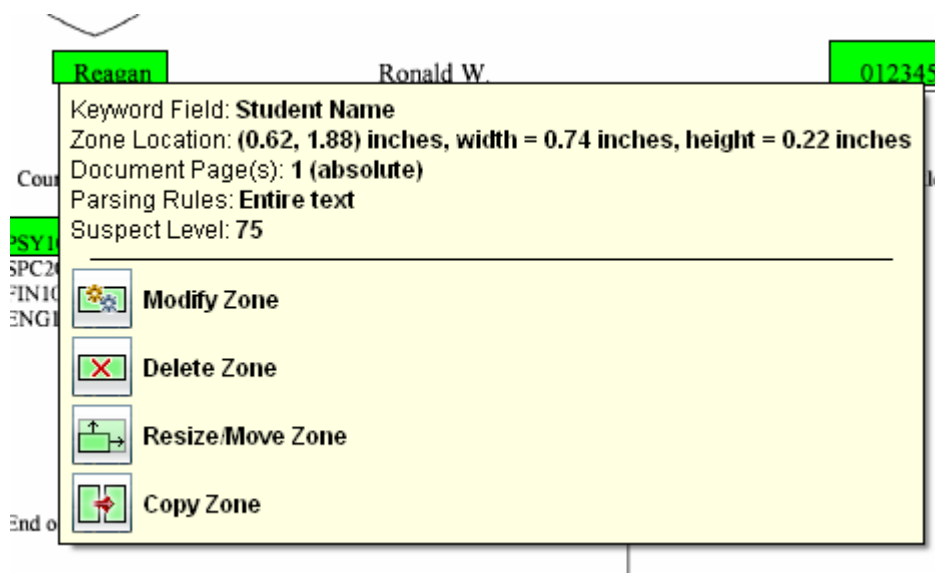
Data Field Zone Signature Detection Field Options	Description
Options	<p>Select the radio button that describes the signature being read by the OCR engine.</p> <ul style="list-style-type: none"> • Signature is framed/underlined. Select this radio button if the signature is located in a pre-defined space, such as a box or on a signature line. • Signature area is unframed. Select this radio button if the signature is located in an undefined or blank space. <p>To ignore machine-printed text in the Data Field Zone (e.g., Signature, Applicant Signature, etc.), select the Exclude machine print text check box.</p> <p>This prevents the signature detection results from being impacted by the presence of machine-printed text.</p>
Sensitivity	<p>Select a radio button to determine the Suspect Level for this Data Field Zone.</p> <ul style="list-style-type: none"> • Highest. Sets the Suspect Level for this Data Field Zone to 70. • Low. Sets the Suspect Level for this Data Field Zone to 75. • Lower. Sets the Suspect Level for this Data Field Zone to 80. • Lowest. Sets the Suspect Level for this Data Field Zone to 85. <p>The Suspect Level is the level of confidence placed in the processing results for this field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that was returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The Sensitivity level selected for this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured from the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Sensitivity to Lowest would indicate you have a fair amount of confidence in the result returned by the OCR engine because few higher scores could be returned and fewer results would be determined to be suspect.</p> <p>Setting the Suspect Level to Highest would indicate you have less confidence in the result because a great number of lower scores could be returned and more results would be determined as suspect.</p>

Data Field Zone Signature Detection Field Options	Description
Advanced filtering	Use the VB script drop-down to select a VB script to associate with the processing of this Data Field Zone. Click the ... button to open the VB Scripts dialog box. Here, the selected script can be re-configured or edited.
Test	Click the Test button to have the OCR engine perform a test process on the Data Field Zone and attempt to detect a signature using the options you have specified. Once the test process is complete, a dialog box is displayed indicating if a signature was detected and the Keyword Value that would be assigned based on this result.

3

MODIFYING, DELETING, RESIZING/MOVING, OR COPYING A DATA FIELD ZONE

An existing Data Field Zone can be modified, deleted, resized/moved, or copied by right-clicking on it in the Document Viewer. If multiple zones overlap, the **Previous** and **Next** links can be used to select the desired zone.



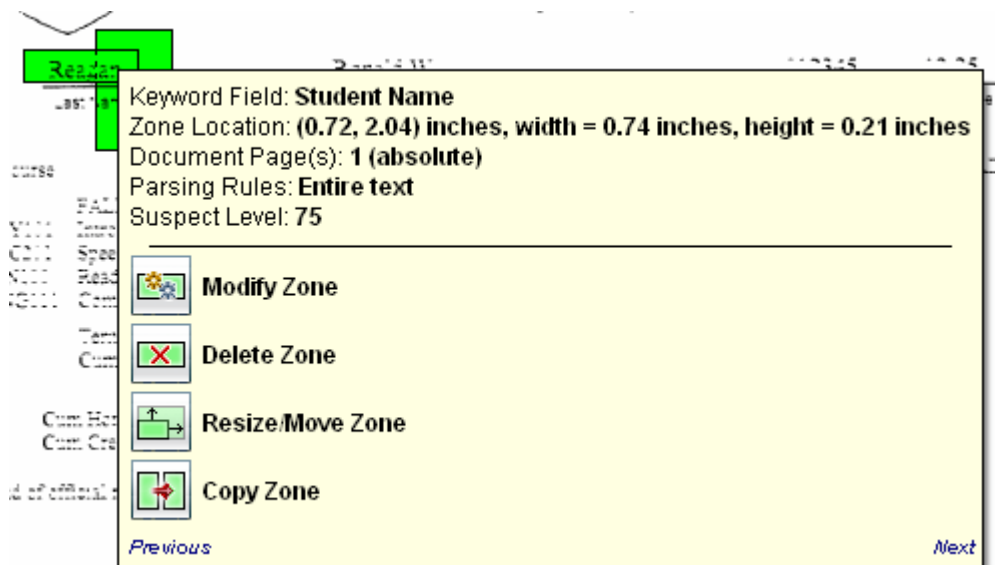
Information about the selected zone is displayed above the following options:

- **Modify Zone.** Click to display the **Data Field Zone** dialog box to modify any of the Data Field Zone's configuration settings.

- **Delete Zone.** Click to delete the selected Data Field Zone.
- **Resize/Move Zone.** Click to make the Data Field Zone editable in the Document Viewer. It can be resized and/or moved without changing any of its existing configuration settings using the pointer.
- **Copy Zone.** Click to create a duplicate of the selected Data Field Zone in the Document Viewer. The new zone is created directly on top of the zone it was created from; it can be resized as needed and then dragged to a new location, while still inheriting all of the configuration settings of the zone it was created from.

Modifying, Deleting, Resizing/Moving, or Copying Overlapping Zones

Depending on your document layout, zones may overlap. Overlapping zone borders are displayed with a dotted line.



When right-clicking on an area of the document with multiple zones, the name of the currently-selected zone is displayed at the top of the dialog box. To switch between zones, click the **Previous** and **Next** links; once the desired zone is selected, it can be modified, deleted, resize/moved, or copied.

Configuring a Page Registration Zone

If you are creating or modifying a form for a multi-page document, you can create a Page Registration Zone on pages of the document (other than the first page) to determine the offset (i.e., the skew or rotation) of the page so that it can be accounted for when performing the Automated Indexing process.

Note: A Page Registration Zone cannot be configured for the first page of a multi-page document. You can configure a Registration Point within the Form Identification Zone to determine and account for offset on the first page.

To create a Page Registration Zone:

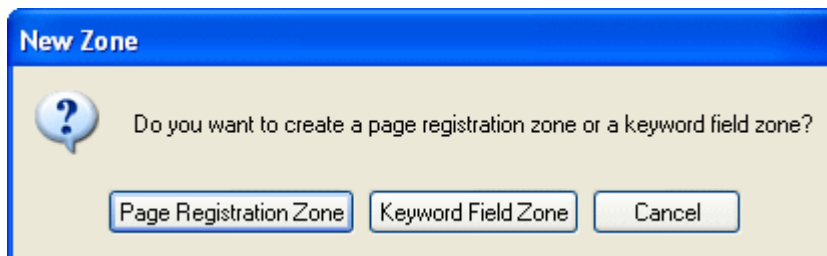
Note: Prior to creating the Page Registration Zone, ensure that the **Default Zone Type** option in the Tools window is set to either **Always Ask** or **Form identity or registration**.

1. With the page of the document that the Page Registration Zone is being created for displayed in the Document Viewer, click and hold the left mouse button and use the pointer to draw a box around the value that is to be used as the Page Registration value. Note that the area of the document that you selected is displayed in the Result Verification window.

Course	Course Title	CRD	GRD	GRDPT
FALL 2006				
PSY101	Intro to College Studies	0.0	W	0.000
SPC201	Speech Communication I	3.0	A-	3.5
FIN100	Reading Financial Information	3.0	B	3.0
ENG111	Composition I	3.0	A	4.0
Total		9.0		3.5

2. Release the left mouse button to display the **Page Registration Zone** dialog box.

Depending on the setting of the **Default Zone Type** option in the Tools window, the **New Zone** dialog box may be displayed.



If the **New Zone** dialog box is displayed, click **Page Registration Zone** to continue to the **Page Registration Zone** dialog box.

Page Registration Zone

Text Match | Image Match

Match value: OFFICIAL

☒ Exact match ☐ Contains text ☐ Fuzzy match Max errors:

☒ Remove all whitespace

Filtering

Allowed character types:

<input checked="" type="checkbox"/> Numerals	<input checked="" type="checkbox"/> Punctuation
<input checked="" type="checkbox"/> Uppercase	<input checked="" type="checkbox"/> Miscellaneous
<input checked="" type="checkbox"/> Lowercase	<input type="checkbox"/> Exclude Asian

Suspect level: 75 (0 - 99)
Form Default = 75, Actual = 70


Page Location(s)

☒ Absolute page: ☐ Relative page(s):

2

Field only appears on a single page; always the same absolute page number (one-based)

☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)


 Official

Find Registration

Save Cancel

3. Select if you would like to use a text value or an image to determine the page's registration.
 - If you are using a text value to determine the page's registration, select the Text Match tab of the **Page Registration Zone** dialog box.
 - If you are using an image to determine the page's registration, select the Image Match tab of the **Page Registration Zone** dialog box.
4. Configure the Page Registration Zone using the options on the **Page Registration Zone** dialog box.
5. Once you have set all necessary options on the **Page Registration Zone** dialog box, click **Save**. The **Page Registration Zone** dialog box is closed, and the Page Registration Zone is temporarily saved to the Automated Indexing form.

The Page Registration Zone is highlighted in yellow on the document in the Document Viewer.



Hyland University

Official Transcript

This is not an official transcript without Registrar's signature and

Clinton

(Last Name)

William J

(First Name, Middle Initial)

Course	Course Title	CRD	GRD	GRDPT	Course
Course					

FALL 2005

6. Repeat Steps 1-6 for each Page Registration field you would like to create for the document.

Tip: At any time, click **Save Configuration** in the Tools window to save the form to the database. It is considered a best practice to save your Automated Indexing form configuration after the creation and configuration of every Automated Indexing form and each form's Form Identification, Page Registration and Data Field Zones.

To discard all configuration changes (including Page Registration Zones created since the last time the form was saved), click **Discard Configuration Changes** in the Tools window.

CONFIGURING A PAGE REGISTRATION ZONE TO USE A TEXT VALUE

A text value can be configured to determine a page's registration by using the options on the Text Match tab of the **Page Registration Zone** dialog box.

Page Registration Zone

Text Match Image Match

Match value: OFFICIAL

☒ Exact match ☐ Contains text ☐ Fuzzy match Max errors:

☒ Remove all whitespace

Filtering

Allowed character types:

☒ Numerals ☒ Punctuation
☒ Uppercase ☒ Miscellaneous
☒ Lowercase ☐ Exclude Asian

Suspect level: 75 (0 - 99)
Form Default = 75, Actual = 70

Page Location(s)

☒ Absolute page: 2 ☐ Relative page(s):

☐ First page
☐ First page (not only page)
☐ Interior pages
☐ Last page
☐ Last page (not only page)

Field only appears on a single page; always the same absolute page number (one-based)

Official

Find Registration

Save Cancel

Page Registration Zone Text Match Options	Description
Match Value	<p>This field displays the text result returned from the OCR engine's evaluation of the Page Registration Zone.</p> <hr/> <p>Note: The Match Value is also displayed in the Result Verification window below the enlarged image of the Page Registration Zone.</p> <hr/> <p>You may select the match rules for the Page Registration Zone by selecting one of the match rule radio buttons.</p> <ul style="list-style-type: none"> • Exact match. Select this radio button if the OCR engine must recognize the Match Value exactly on a processed document in order for Automated Indexing to determine and account for offset using the Page Registration Zone. • Contains text. Select this radio button if the OCR engine needs only to identify the Match Value as part of a longer string on the document being processed in order to determine and account for offset using the Page Registration Zone. <hr/> <p>Note: When using this option, only the Match Value portion of the text string is taken into consideration to determine the Suspect Level of the value.</p> <hr/> <ul style="list-style-type: none"> • Fuzzy match. Select this radio button if the OCR engine needs to identify the Match Value on a processed document, but can allow for some margin for error between the Match Value and the value read from the document being processed in order to determine and account for offset using the Page Registration Zone. <p>Enter a value in the Max Errors field to set the number of discrepancies allowed between the Match Value and the value read from the document being processed.</p> <p>For example, if your Match Value is HYLAND UNIVERSITY and the value detected by the OCR engine from the processed document is HYL4ND UN1VERSITY, the Page Registration Zone can still be used to determine and account for offset if the Max Errors value is set to 2 or more.</p>

Page Registration Zone Text Match Options	Description
Remove all whitespace	<p>Select this check box if you would like to remove all whitespace (e.g., spaces, paragraph returns, etc.) from the Match Value after it is read by the OCR engine.</p> <p>For example, if the Match Value read by the OCR engine is PSY 103, if the Remove all whitespace check box is selected, the Match Value would be changed to PSY103.</p>
Allowed character types	<p>These options are available to assist the OCR engine determine if the value it reads is correct. Select the check box next to each character type that is allowed in the Match Value. The available options are:</p> <ul style="list-style-type: none"> • Numerals. Numeric characters, 0-9. • Uppercase. Uppercase alphabetic characters. • Lowercase. Lowercase alphabetic characters. • Punctuation. Punctuation marks (i.e., . ! ?). • Miscellaneous. Other ASCII characters that do not fall into one of the above categories (i.e., # \$ * @). <p>When a character is recognized by the OCR engine that is not part of an allowable character set, the character is replaced by a tilde (~) and the value is automatically marked as suspect.</p> <p>Tip: Using the allowed character type options can sometimes help the OCR engine more easily determine the correct value by eliminating characters that are obviously not correct (e.g., an I is correctly identified instead of a 1 because numeric characters are filtered and prevented from being recognized as part of the value).</p> <p>Select the Exclude Asian check box to instruct the OCR engine to skip Asian (i.e., double-byte) characters.</p> <p>Tip: Selecting the Exclude Asian check box allows you to identify numeric data (i.e., Date Keyword Values, Currency Keyword Values and Document Dates) when performing Automated Indexing on documents configured to contain Asian (i.e., double-byte) characters.</p>

Page Registration Zone Text Match Options	Description
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Automated Indexing form and the actual Suspect Level detected for the selected value are displayed below the Suspect Level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that is returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured by the zone to be marked as suspect. All scores lower than the Suspect Level threshold indicate that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <hr/> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that the page registration is being correctly determined.</p> <hr/>

Page Registration Zone Text Match Options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Page Registration Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the Page Registration Zone is only displayed on one page and is always displayed on the same page (e.g., the Page Registration Zone is always on page 1). Enter the page number that the Page Registration Zone is located on in the associated field. • Select the Relative page radio button if the Page Registration Zone may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the Page Registration Zone may be located on. <p>Select the First page check box if the Page Registration Zone is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the Page Registration Zone is located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the Page Registration Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the Page Registration Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the Page Registration Zone is located on the last page and other pages in the document.</p>
Find Registration	<p>The Registration Point is the starting point (top, left corner) of the value detected in the Page Registration Zone.</p> <p>Click Find Registration to automatically set the Registration Point at the starting point of the value in the Page Registration Zone, or double-click a location to manually set the registration point at that position.</p>

CONFIGURING A PAGE REGISTRATION ZONE TO USE AN IMAGE MATCH

An image can be configured to determine a page's registration by using the options on the Image Match tab of the **Page Registration Zone** dialog box.



3

Page Registration Zone Image Match options	Description
Image Matching Form Identification Zone	<p>By default, Page Registration Zones are configured to use text values to determine the page's registration.</p> <p>Select this check box to configure this Page Registration Zone to use an image to determine the page's registration.</p> <p>This check box must be selected in order to enable the other options on the Image Match tab of the Page Registration Zone dialog box.</p>
Suspect level	<p>Enter the Suspect Level threshold, 1-99, in this field. By default, this value is set to the default Suspect Level set for the Automated Indexing form.</p> <p>The Suspect Level is the level of confidence placed in data values captured in this zone. The default Suspect Level set for the Automated Indexing form is displayed below the Suspect Level field.</p> <p>After a zone is processed, the OCR engine gives the resulting value a score between 1 and 99, depending on how confident it is in the result that is returned. The higher the score is, the lower the OCR engine's confidence is in the results.</p> <p>The value you enter in this field is the threshold at which the OCR engine determines if a returned value is acceptable or suspect. A score returned by the OCR engine higher than the Suspect Level threshold you set causes the value captured by the zone to be marked as suspect. A score lower than the Suspect Level threshold indicates that the captured value is considered by the OCR engine to be acceptable.</p> <p>For example, setting the Suspect Level to 99 would indicate you completely trust the result returned by the OCR engine because no higher score could be returned and no result could be marked as suspect.</p> <p>Setting the Suspect Level to 1 would indicate you have no trust in the result, since no lower score could be returned and no result could be determined acceptable.</p> <p>Tip: By default, the Suspect Level threshold is set to 75 and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that the page registration is being correctly determined.</p>

Page Registration Zone Image Match options	Description
Page Location(s)	<p>The Page Location(s) options control the pages that the OCR engine searches for a particular Page Registration Zone.</p> <ul style="list-style-type: none"> • Select the Absolute page radio button if the Page Registration Zone is only displayed on one page and is always displayed on the same page (e.g., the Page Registration Zone is always on page 1). Enter the page number that the Page Registration Zone is located on in the associated field. • Select the Relative page radio button if the Page Registration Zone may be located on one or more pages relative to the length of the document. Select one or more of the following check boxes to indicate which page(s) that the Page Registration Zone may be located on. <p>Select the First page check box if the Page Registration Zone is located only on the first page of the document. This option can be used in conjunction with the Interior pages and/or Last page check boxes.</p> <p>Select the First page (not only page) check box if the Page Registration Zone is located on the first page and other pages in the document.</p> <p>Select the Interior pages check box if the Page Registration Zone is located on every page of the document other than the first or last page. This option can be used in conjunction with the First page and/or Last page check boxes.</p> <p>Select the Last page check box if the Page Registration Zone is located only on the last page of the document. This option can be used in conjunction with the First page and/or the Interior pages check boxes.</p> <p>Select the Last page (not only page) check box if the Page Registration Zone is located on the last page and other pages in the document.</p>
Significant Area	<p>Use the slider to control the size of the significant processing area. The significant processing area is the red box displayed on the image snippet.</p> <p>Increasing the size of the significant processing area (i.e., moving the slider to the right) will increase the accuracy of the image match process, but will decrease the speed of the process.</p> <p>Decreasing the size of the significant processing area (i.e., moving the slider to the left) will increase the speed of the image match process, but will decrease the accuracy of the process.</p>

Page Registration Zone Image Match options	Description
Find Registration	<p>The Registration Point feature is required when using an image identifier.</p> <p>The Registration Point is the starting point (top, left corner) of the image's significant processing area (the red box displayed on the image snippet). It is used to determine the page's offset (i.e., any skew or rotation) that might have occurred during scanning.</p> <p>The position of the Registration Point on the Automated Indexing form is compared to the same position on the document being processed. The position of the configured Page Registration Zone is adjusted on the document being processed to account for the detected offset to ensure that the Automated Indexing process is able to process the document properly.</p> <p>Click Find Registration to automatically set the Registration Point, or double-click a location on the image to manually set the registration point at that position.</p>

MODIFYING, DELETING, OR RESIZING/MOVING A PAGE REGISTRATION ZONE

An existing Page Registration Zone can be modified, deleted or resized/moved by right-clicking on it in the Document Viewer. If multiple zones overlap, the **Previous** and **Next** links can be used to choose the desired zone.



Official Transcript

This is not an official transcript without Registrar's signature and college seal

Clinton
(Last Name)

William J
(First Name, Middle Initial)

1234
(ID #)

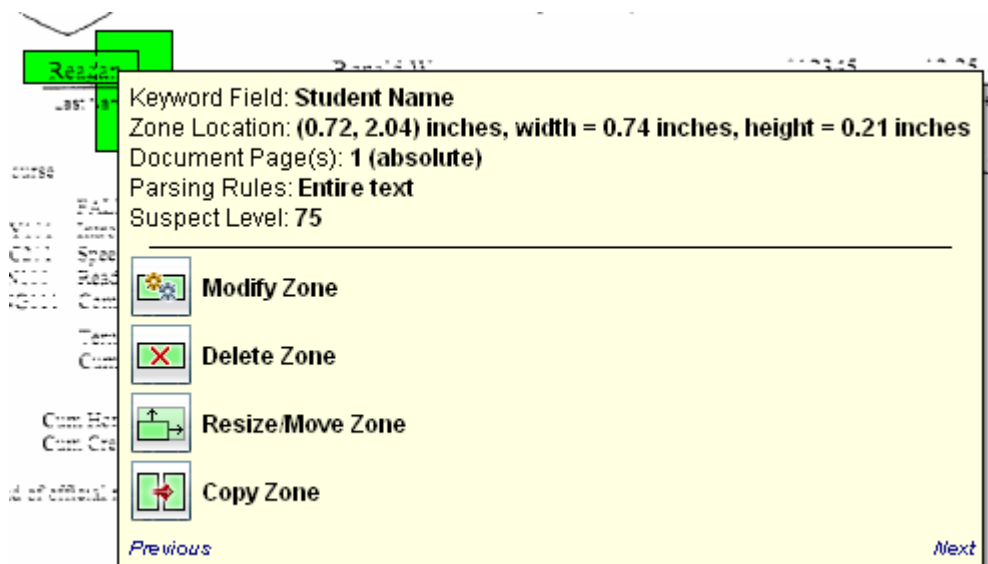
Course	Course Title	CRD	GRD	GRDPT	Course	Course I
<div> <div>Registration Zone: COURSE</div> <div>Zone Location: (0.32, 2.54) inches, width = 0.53 inches, height = 0.30 inches</div> <div>Registration Point: (0.38, 2.63) inches</div> <div>Text Match: Exact</div> <div>Remove Whitespace: Yes</div> <div>Suspect Level: 75</div> <div>Document Page(s): 2 (absolute)</div> <div> <div>Modify Zone</div> <div>Delete Zone</div> <div>Resize/Move Zone</div> </div> </div>						
ASN303	The Middle East	3.0	A	4.00		
SPN112	Beg Spanish I					

Information about the selected zone is displayed above the following options:

- **Modify Zone.** Click to display the **Page Registration Zone** dialog box to modify any of the Page Registration Zone's configuration settings.
- **Delete Zone.** Click to delete the selected Page Registration Zone.
- **Resize/Move Zone.** Click to make the Page Registration Zone editable on the Document Viewer. It can be resized and/or moved without changing any of its existing configuration settings using the pointer.

Modifying, Deleting, Resizing/Moving, or Copying Overlapping Zones

Depending on your document layout, zones may overlap. Overlapping zone borders are displayed with a dotted line.



When right-clicking on an area of the document with multiple zones, the name of the currently-selected zone is displayed at the top of the dialog box. To switch between zones, click the **Previous** and **Next** links; once the desired zone is selected, it can be modified, deleted, resize/moved, or copied.

Importing/Exporting Automated Indexing Forms

Automated Indexing forms can be imported into/exported from your Automated Indexing solution. This allows you to share forms that you have created with another Automated Indexing solution, or allows another Automated Indexing solution to share its forms with you.

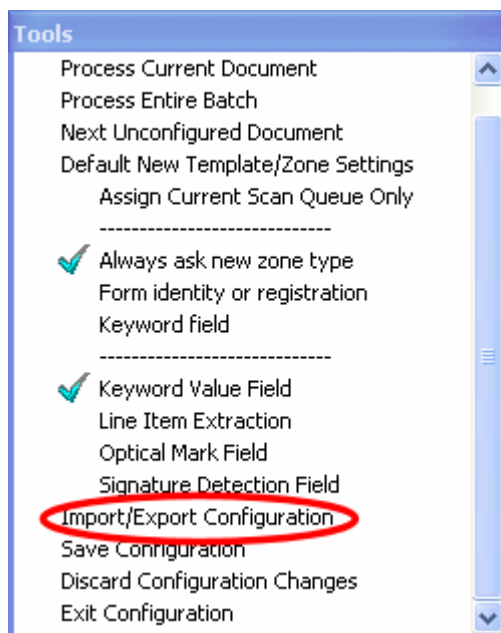
Once exported, Automated Indexing forms are saved as XML files for easy transmittal.

Tip: The ability to import/export Automated Indexing forms can reduce or eliminate the need for Automated Indexing solution users to repeat the task of creating the same Automated Indexing form.

Once created for your Automated Indexing solution, the Automated Indexing forms are available to any of your users with rights to access it (or the scan queue it is associated with). Automated Indexing forms do not need to be imported/exported between workstations that are part of the same Automated Indexing solution.

Importing Automated Indexing Forms

1. From the **Automated Index Configuration** window, click the **Import/Export Configuration** link in the Tools window.



2. A confirmation dialog is displayed. Click **Import**.

- The **Automated Index Configuration Import** dialog box is displayed.

3

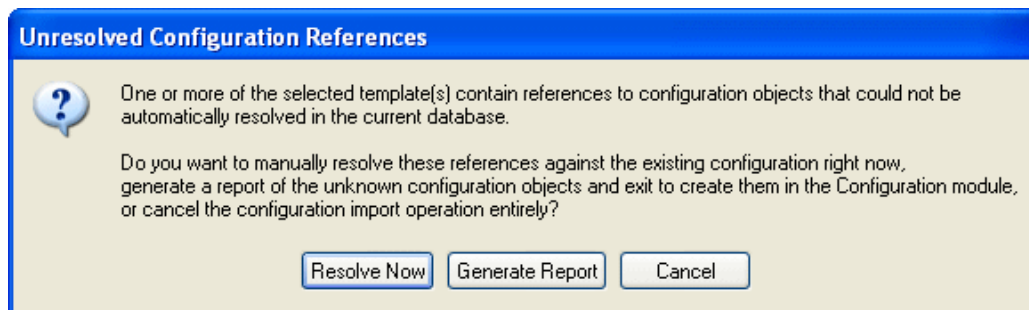
Information about each Automated Indexing form contained in the XML file is listed in the **Automated Index Configuration Import** dialog box. (e.g., the forms' names, their reconciliation status, and if they already exist in your solution)

- ©2011 Hyland Software, Inc.

7. A message is displayed informing you that the import process was successful, and you are prompted to perform Automated Indexing form precedence order analysis.

RECONCILING CONFIGURATION REFERENCES

When attempting to import Automated Indexing forms, you may be presented with a message warning you that the forms you are attempting to import contain unresolved configuration references.

**3**

This message indicates that the Automated Indexing forms you are trying to import have been configured using elements (e.g., Document Types, Keyword Types, scan queues) that do not exist in your solution.

For example, a form you are trying to import may be configured to be matched to documents belonging to the **WSU - Transcripts** Document Type, and your solution does not contain a Document Type named **WSU - Transcripts**.

You have two options for resolving these references:

- Click **Generate Report** to generate a list of all of the unresolved configuration references.

A text document is displayed containing a list of all of the unresolved configuration references for the form(s) you are trying to import. You can use this report as a reference to create these elements in the Configuration module.

Note: While this report lists all of the unresolved configuration elements (e.g., Document Types, Keyword Types, scan queues, etc.), it does not give any information about their configuration settings.

- Click **Resolve Now** to map the Automated Indexing form's unresolved configuration references to elements of your solution.

The **Configuration Object Mapping: <Configuration Element>** dialog box is displayed.

Configuration Object Mapping: Document Type

Select the appropriate Document Type for each unknown item listed
Select <None> for any items that do not have corresponding entries in this database

Item #	Source System Name	Corresponding Item
X 1	HYL-Transcripts	<unknown>

Set all references to:

Update All

OK

Cancel

Each of the unresolved reference for the form(s) you are importing are listed in the dialog box. Use the **Corresponding Item** drop-down select list to select an element in your solution to substitute for the unresolved element associated with the form. Or, select one element from your solution to substitute for all unresolved elements by using the **Set all references to** drop-down select list and clicking **Update All**.

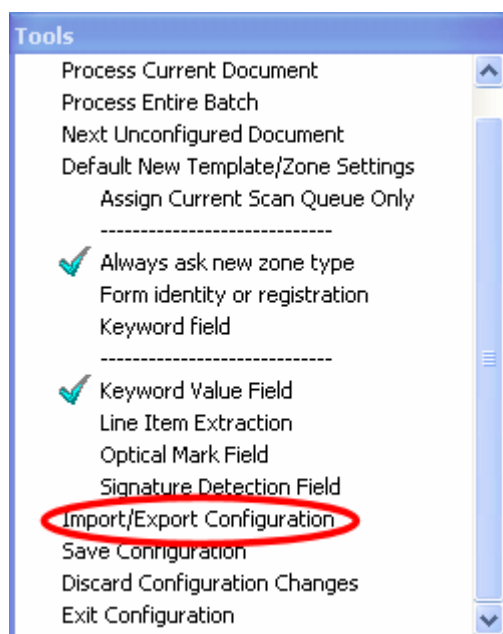
For example, a Automated Indexing form was associated with a Document Type named **HYL - Transcripts** in the solution it was exported from. However, this Document Type does not exist in the solution the form is being imported into, so you must use the **Corresponding Item** drop-down select list (or the **Set all references to** drop-down select list) to select a different Document Type to “substitute” for the **HYL - Transcripts** Document Type.

When you have selected a substitute for all of the unresolved elements listed in the dialog box, click **OK**. If you have unresolved configuration elements of a different type (e.g., Keyword Type, scan queue), another **Configuration Object Mapping: <Configuration Feature>** dialog box may be displayed.

Once all unresolved configuration features have been resolved, the Automated Indexing forms are imported into your solution. You are given the option to save the reconciliation mapping information you specified to automatically resolve any future Automated Indexing forms imported from this source.

Exporting Automated Indexing Forms

1. From the **Automated Index Configuration** window, click the **Import/Export Configuration** link in the Tools window.



2. A confirmation dialog is displayed. Click **Export**. The **Automated Index Configuration Export** dialog box is displayed.

[illegible]

Each Automated Indexing form configured for your solution is listed in the **Automated Index Configuration Export** dialog box.

The following information is displayed for each Automated Indexing form: the Form ID, the name of the Automated Indexing Form, the scan queue(s) that the Automated Indexing form is assigned to, and the Document Type that the documents matched to the Automated Indexing form are assigned to.

3. Select the check box next to each Automated Indexing form to be exported.
To select all Automated Indexing forms, click **Select All**. To deselect all Automated Indexing forms, click **Deselect All**.
4. Click **Browse**. Browse to the location where you would like to save the XML file containing the exported Automated Indexing form data and enter a name for the Automated Indexing form data file.
5. Click **Save** to close the **Save As** dialog box.

6. Click **Export**. Once the Automated Indexing form data is successfully exported, a confirmation message is displayed.

Automated Indexing Form Order Analysis

Your Automated Indexing solution keeps track of how frequently each form is matched to documents that undergo Automated Indexing.

As an administrator, you can analyze these statistics, and, based on this analysis, your Automated Indexing solution can update the order of its forms and attempt to match the more frequently-used forms to documents first.

For example:

You scan a batch of **AP - Invoice** documents daily, and within these batches, there are three potential Automated Indexing forms available to these documents (**AP - Computers R Us**, **AP - The Phone Shack**, and **AP - Office Supply Warehouse**).

Based on the volume of documents that are scanned and indexed:

- The **AP - Office Supply Warehouse** form is matched to 75% of the documents
- The **AP - Computers R Us** form is matched to 15% of the documents
- The **AP - The Phone Shack** form is matched to 10% of the documents.

Currently, Automated Indexing attempts to match the forms to incoming **AP - Invoice** documents in this order: **AP - The Phone Shack**, **AP - Computers R Us**, and then **AP - Office Supply Warehouse**.

In its current state, your **AP - Invoice** documents are not being indexed very efficiently because only 25% of all incoming documents are being matched to a form in less than three attempts.

After running the form order analysis, your Automated Indexing solution re-orders the forms so that it attempts to match all incoming **AP - Invoice** documents to the **AP - Office Supply Warehouse** form first; the **AP - Computers R Us** form second, and the **AP - The Phone Shack** form third.

Now, 85% of all incoming **AP - Invoice** documents are matched to a form in less than three attempts.

Note: If your solution is configured to use an Institutional Database, be aware that a super-user can view/re-order all forms across all institutions, while a non-super user can only view and re-order templates that are configured for his/her institution.

In order for the Automated Indexing form order analysis to provide accurate results, you must ensure that you have a sufficiently-large enough number of documents that have been successfully matched to a form currently in your system. By default, you can perform a form order analysis with a minimum of 100 documents and a maximum of 30,000 document

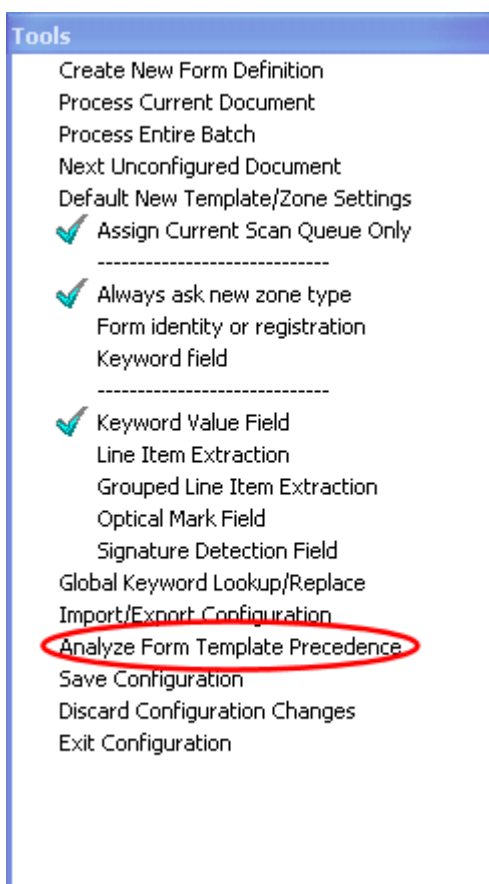
You can perform a form order analysis at any time from the **Automated Index Configuration** window. Additionally, you will always be prompted to perform a form order analysis after you have imported forms into your solution.

Tip: It is considered a best practice to periodically perform a form order analysis and re-order your forms. The frequency of this analysis should be based on the number of documents you are processing via Automated Indexing and the number of forms configured for your solution.

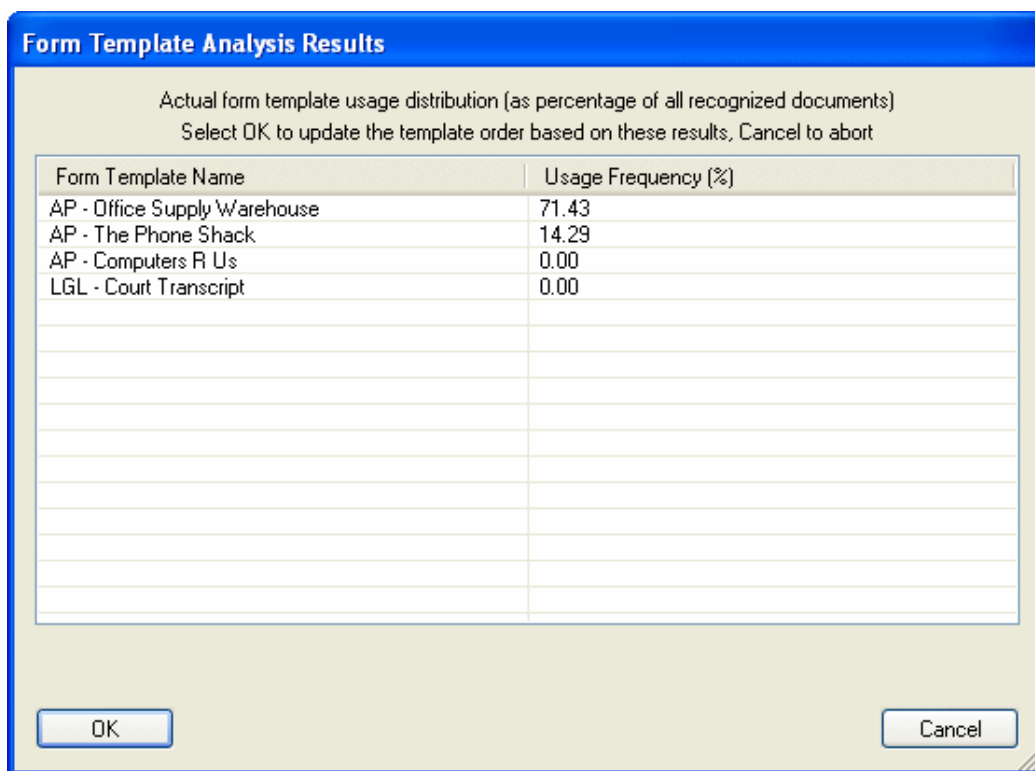
To perform a form order analysis:

1. From the **Automated Index Configuration** window, click the **Analyze Form Template Precedence** link in the Tools window.

3



2. Your Automated Indexing solution analyzes its form frequency statistics. Once completed, they are displed in the **Form Template Analysis Results** window.



The image shows a screenshot of the 'Form Template Analysis Results' window. The window has a blue title bar with the text 'Form Template Analysis Results'. Below the title bar, there is a light beige background with the following text: 'Actual form template usage distribution (as percentage of all recognized documents)' and 'Select OK to update the template order based on these results. Cancel to abort'. Below this text is a table with two columns: 'Form Template Name' and 'Usage Frequency (%)'. The table contains four rows of data: 'AP - Office Supply Warehouse' with a usage frequency of 71.43, 'AP - The Phone Shack' with 14.29, 'AP - Computers R Us' with 0.00, and 'LGL - Court Transcript' with 0.00. There are several empty rows below these. At the bottom of the window, there are two buttons: 'OK' on the left and 'Cancel' on the right.

Form Template Name	Usage Frequency (%)
AP - Office Supply Warehouse	71.43
AP - The Phone Shack	14.29
AP - Computers R Us	0.00
LGL - Court Transcript	0.00

3

The forms are displayed in descending order (i.e., the most frequently-used form is listed first and the least-frequently used form is listed last) along with their percentage of use.

3. Click **OK** to re-order the forms. Once the forms have been re-ordered, a confirmation message is displayed.

Now, your Automated Indexing solution will first attempt to match the most commonly used forms to incoming documents in order to more efficiently process documents.

Configuring XML Renditions

The Automated Indexing module can be configured to generate an XML rendition of extracted Keyword Values when indexing documents. In order to configure XML rendition generation, you must perform the following configuration steps:

1. **Configure the Document Types being indexed to allow renditions.** For more information on enabling renditions, see *Configuring Document Types for OCR* on page 56.
2. **Enable the creation of an XML data rendition on each form to be processed.** Automated Indexing Forms must have the **Create XML data rendition** option selected. To enable XML data rendition creation in new forms, see *Creating a New Automated Indexing Form* on page 73. To enable XML data rendition creation in existing forms, see *Modifying a Form* on page 106.
3. **Define XML nodes for all Data Field Zones.** Defining the XML node in a Data Field Zone generates the XML tags for each extracted Keyword Value. See *Configuring A Data Field Zone* on page 126 for more information on defining XML Nodes in Data Field Zones.

Testing Automated Indexing Forms

Once you have matched a form to one or more sample document(s), you can run a test Automated Indexing process to see if the Form Identification, Registration and Data Field Zones are processed correctly on the sample document(s).

Tip: It is considered a best practice to regularly test each Automated Indexing form to ensure that it does not require updating. All Automated Indexing forms should be tested after they are created and after any modifications have been made to them.

The test can be performed at the document level or at the batch level. The document level test is likely to be faster than the batch level test because only one document is being evaluated, but the batch level test can be useful because the OCR engine automatically attempts to match an existing form to all of the documents in the batch, and can help you determine which documents require a new or modified Automated Indexing form.

Note: All documents in the batch are listed in the **Batch Documents** window, in the top-right corner of the **Automated Index Configuration** window. A green check mark is displayed next to documents in the batch that have been matched to a Automated Indexing form, and a red X is displayed next to documents that are unclassified.

To test a Automated Indexing form:

1. From the **Automated Index Configuration** window, ensure that the document and its associated Automated Indexing form are displayed in the Document Viewer.
2. To test a single document, click **Process Current Document** in the Tools window.

To test an entire batch, click **Process Entire Batch** in the Tools window.

During processing, click **Cancel** at any time to cancel the test. However, if the test is cancelled before it is complete, any documents in the batch that had not been processed remain listed as unclassified or they are matched to the Automated Indexing form that they were matched to during pre-processing or the last test process.

3. If you are testing a single document, the **Test Result Verification** window is displayed once the test is complete.

Result Verification

- [-] Form type: HU Transcript 3 (121)
 - AD - Transcript
 - Calculated registration offset (1, 0) pixels
 - Field counts: 6 identified, 0 suspect
 - Default suspect level: 75
- [-] Field results
 - [-] Last Name, First Name, Middle Name at (0.64, 1.87) inches on page(s): 1 (absolute) - [Person Name/Separated]
 - Document page: 1
 - [-] Keyword Type: Last Name
 - Text value: WILLIAM
 - Calculated suspect level: 20, tolerance: 75
 - Sequence ID: 142-0
 - [-] Keyword Type: First Name
 - Text value: CLINTON
 - Calculated suspect level: 20, tolerance: 75
 - Sequence ID: 142-0
 - [-] <Line item extraction/Multiple keyword types> at (0.83, 3.02) inches on page(s): 1 (absolute) - [Line Item Extraction]
 - Document page: 1
 - XML Node: Classes
 - [-] Keyword Type: Class Name
 - Text value: COMPOSITION I
 - Calculated suspect level: 70, tolerance: 75
 - Sequence ID: 143-0
 - XML Node: Class
 - [-] <Line item extraction/Multiple keyword types> at (0.87, 4.29) inches on page(s): 1 (absolute) - [Line Item Extraction]
 - Document page: 1
 - XML Node: Classes

```

>>Engine - version 16.4 initialized
>>Form Identity: begin new document operation [document type = AD - Transcript (137)]
>>Form Identity: begin identification test on form [HU Transcript 3 (121)] at page 1
>>Form Identity: copy document page 1 to engine
>>Form Identity: beginning identification chain group 1
>>Form Identity: identity 117 zone = [149, 21] - [385, 73]
>>Form Identity: zone text on identity 117 [HYLAND UNIVERSITY]
>>Form Identity: exact text match test for [HYLAND UNIVERSITY]
>>Form Identity: match test succeeded at position 0
>>Form Identity: registration offset for identity 117 is [1, 0]
>>Form Identity: matched form ID = HU Transcript 3 [121]
>>Form Identity: averaged registration offset = [1, 0]
>>Field recognition: copy document page 1 to engine
>>Field recognition: field ID [142] zone = [64, 187] - [360, 214]
>>Field recognition: field ID [142] last name = 'William', first name = 'Clinton', middle name = '', suspect level = 20
>>Field recognition: field ID [143] zone = [83, 302] - [276, 368]
>>Field recognition: field ID [143] recognized line item matrix 1 x 1
>>Field recognition: field ID [143] [row = 1, column = 1] raw result = 'Composition I', suspect level = 70
>>Field recognition: field ID [144] zone = [87, 429] - [257, 512]
>>Field recognition: field ID [144] recognized line item matrix 1 x 1
>>Field recognition: field ID [144] [row = 1, column = 1] raw result = 'Beg Spanish I', suspect level = 70
>>Field recognition: field ID [145] zone = [88, 570] - [275, 658]
>>Field recognition: field ID [145] recognized line item matrix 1 x 1
>>Field recognition: field ID [145] [row = 1, column = 1] raw result = 'Principles of Sociology', suspect level = 70
>>Field recognition: field ID [154] zone = [88, 731] - [241, 817]
>>Field recognition: field ID [154] recognized line item matrix 1 x 1
>>Field recognition: field ID [154] [row = 1, column = 1] raw result = 'icial record.', suspect level = 70

```

The **Test Result Verification** window is divided into two panes:

- The top pane contains a summary of information about the Automated Indexing form that was matched to the document. Information regarding Keyword Values that did not meet their respective fields' configured Suspect Level is displayed in red.
- The bottom pane contains information returned from the OCR engine during processing.

If you are testing an entire batch, the **Test Result Verification** window is not displayed. Instead, the results are only displayed in the **Result Verification** window within the **Automated Index Configuration** window.

4. Review the information in the **Test Results Verification** window or the Result Verification window in order to determine what, if any, changes need to be made to the Automated Indexing form(s).

Configuring Automated Indexing to Use Intelligent Character Recognition

If your solution is licensed for Intelligent Character Recognition (ICR), you can configure Automated Indexing to read and capture Keyword Values from hand-written text.

ICR can only be configured to be performed in Text Field Data Field Zones (not Line Item Extraction, Grouped Line Item Extraction, Optical Mark Field, or Signature Detection Field Data Field Zones). In order to perform ICR, the Data Field Zone must be configured to identify hand-written text instead of machine-printed text.

To configure a Data Field Zone to read hand-written text (i.e., perform ICR processing):

1. From the **Data Field Zone** dialog box, ensure that the Text Field tab is displayed. ICR processing can only be performed on Text Field Data Field Zones.

Note that the **Advanced recognition** drop-down select list is enabled. This option is only enabled when your solution is licensed for ICR.

3

The screenshot shows the 'Data Field Zone' dialog box with the 'Text Field' tab selected. The 'Keyword type' is set to '<None>'. Under 'Data Application', 'Entire zone value' is selected. Under 'Page Location(s)', 'Absolute page:' is selected with the value '1'. The 'Advanced recognition' dropdown at the bottom is circled in red and set to 'Machine print text/OCR'. Other options include 'Remove line breaks', 'First line only', 'Table style', 'Find by tag/expression', 'Person name extraction', 'Remove all whitespace from result', 'Filtering' (Numerals, Uppercase, Lowercase, Punctuation, Miscellaneous, MICR data, Exclude Asian), 'Advanced filtering' (VB script, Suspect level), and 'Specific allowed characters'.

Data Field Zone

Grouped Line Item Extraction | Optical Mark Field | Signature Detection Field

Text Field | Line Item Extraction

Keyword type: <None>

Data Application

- ☒ Entire zone value
- ☐ Remove line breaks
- ☐ First line only
- ☐ Table style
- ☐ Find by tag/expression
- ☐ Person name extraction

☐ Remove all whitespace from result

Page Location(s)

- ☒ Absolute page: 1
- ☐ Relative page(s):

Field only appears on a single page; always the same absolute page number (one-based)

- ☐ First page
- ☐ First page (not only page)
- ☐ Interior pages
- ☐ Last page
- ☐ Last page (not only page)

Filtering

Character search hints:

- ☒ Numerals
- ☒ Punctuation
- ☒ Uppercase
- ☒ Miscellaneous
- ☒ Lowercase
- ☐ MICR data
- ☐ Exclude Asian

Advanced filtering:

VB script: <None>

Suspect level: 75 (0-99) Form Default = 75

Specific allowed characters:

Advanced recognition: Machine print text/OCR

Save Cancel

2. Using the **Advanced Recognition** drop-down select list, select one of the OCR/ICR recognition options:
 - **Machine print text/OCR.** The OCR engine will perform optical character recognition in order to read machine-printed text.
This option is selected by default.
 - **Handwriting/ICR -numerals/grouping punctuation (North American style).** The OCR engine will perform intelligent character recognition to read hand-written text.
This option will only attempt to recognize numeric characters written in the North American style (e.g., the 7 character is not crossed).
 - **Handwriting/ICR -numerals/grouping punctuation (European style).** The OCR engine will perform intelligent character recognition to read hand-written text.
This option will only attempt to recognize numeric characters written in the European style (e.g., the 7 character is crossed).
 - **Handwriting/ICR - alphanumeric/punctuation.** The OCR engine will perform intelligent character recognition in order to read hand-written text.
This option will attempt to recognize all alphanumeric characters.
3. If one of the **Handwriting/ICR** options is selected, the **Remove vertical/horizontal lines from zone before processing** check box is displayed below the Advanced Recognition drop-down select list.

Select this option to automatically strip vertical or horizontal lines (i.e., constraint boxes) from the zone before it is processed.
4. Once you have selected the OCR/ICR recognition option for the Data Field Zone, click **Save** to close the **Data Field Zone** dialog box.

Limitations of Intelligent Character Recognition

Automated Indexing has the following limitations when attempting to perform ICR:

- ICR cannot be performed using Point and Shoot Indexing.
- A single Data Field Zone cannot be configured to perform both OCR and ICR (i.e., both hand-written text and machine-printed text cannot be identified from with the same Data Field Zone).

- Only characters belonging to single byte character-based alphabets (e.g., English, French, Spanish, German, etc.) can be recognized via ICR processing. Characters belonging to double-byte character-based alphabets (e.g., Traditional Chinese, Japanese, Korean, etc.) cannot be recognized.
- ICR processing is available for only Text Field Data Field Zones. It is not available for any of the other types of Data Field Zones (Line Item Extraction, Group Line Item Extraction, etc.).

Improving the Accuracy of Intelligent Character Recognition

It is difficult to predict the accuracy of an ICR process due to the variability in the quality of the hand-written text. The following tips will help to improve the accuracy of ICR processing.

- The hand-written characters should be clearly separate from one another (i.e., they should not touch or intersect).
- The hand-written text should be unconstrained. If the characters are being entered into constraint boxes, you should configure your OCR process format to automatically drop them by selecting the **Remove vertical/horizontal lines from zone before processing** check box.

Constrained Text:



Name

Unconstrained Text:

A single line of handwritten text reading 'John Adams' without any grid constraints.

Name

- The ICR numeric text processor (**Handwriting/ICR - numerals/grouping punctuation (North American style)**, **Handwriting/ICR - numerals/grouping punctuation (European style)**) is significantly more accurate than the ICR alphanumeric text processor (**Handwriting/ICR - alphanumeric/punctuation**) because the character set is much smaller.

If you are attempting to process numeric data only from a Data Field Zone, you will have more accurate results if you use the numeric text processor.

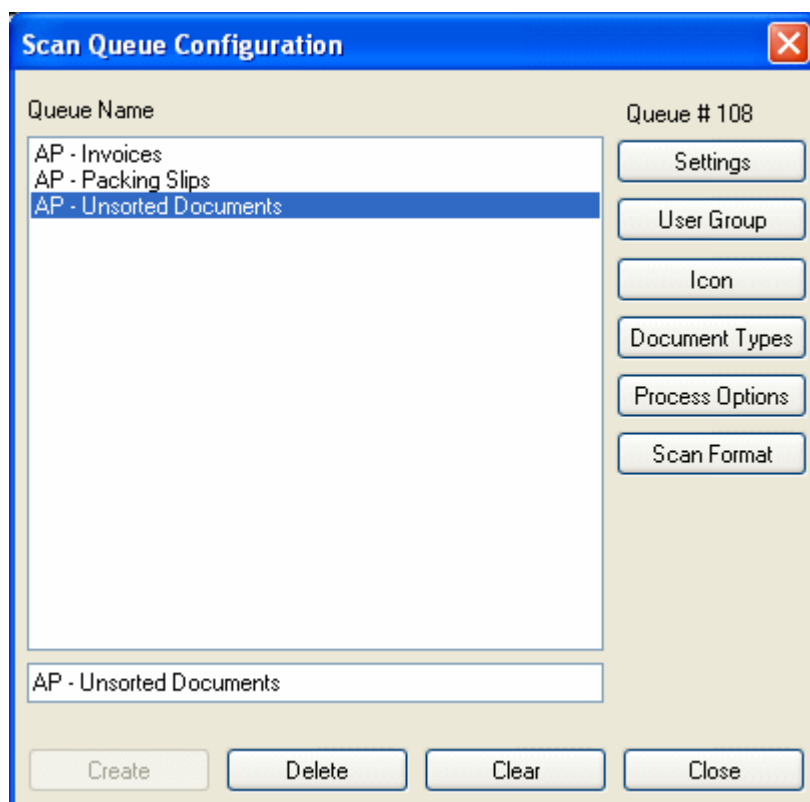
Configuring a Scan Queue for Automated Indexing

Scan queues associated with the batches that will undergo Automated Indexing must be configured to route those batches to the **Awaiting Automated Index** batch status queue.

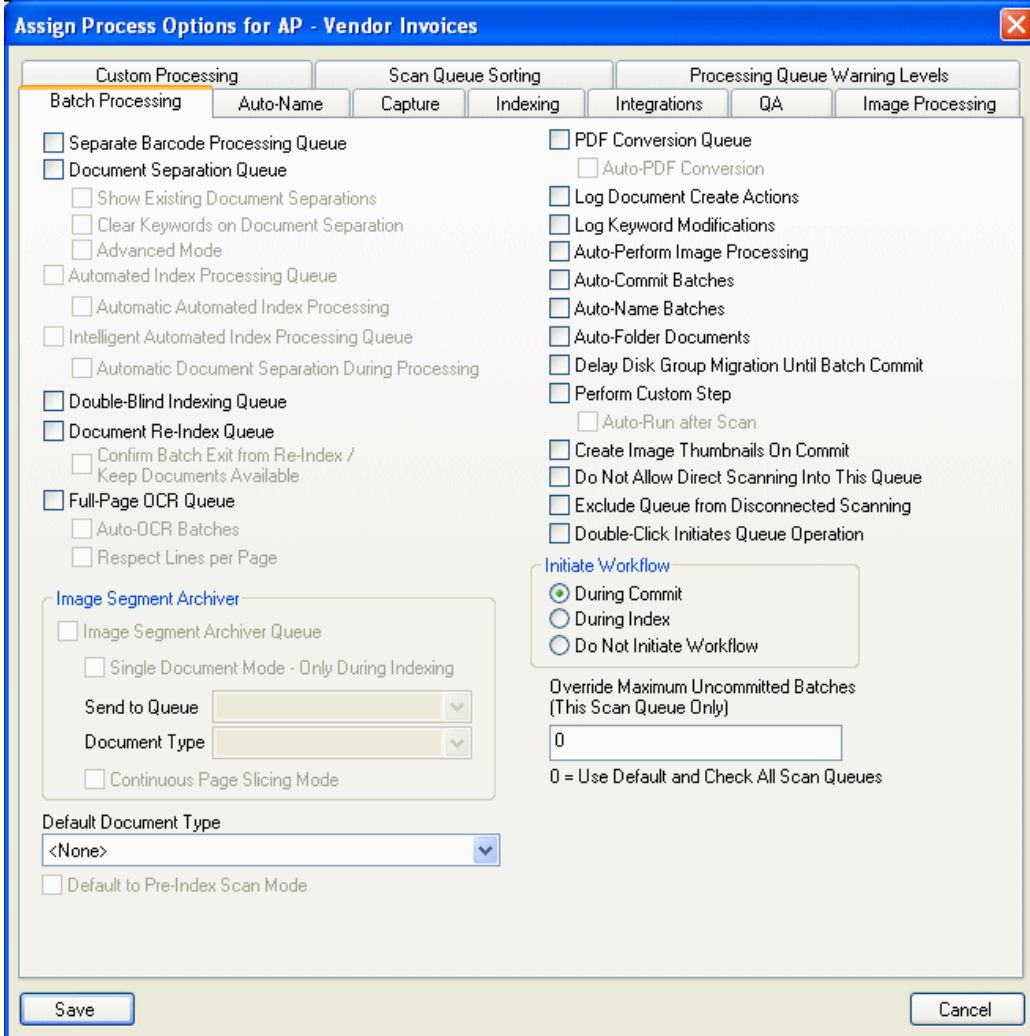
Note: For information on creating and configuring scan queues, see the Document Imaging documentation.

To configure a scan queue for Automated Indexing:

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.



2. Select the scan queue to be configured for Automated Indexing from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.



The dialog box titled "Assign Process Options for AP - Vendor Invoices" contains several tabs: Custom Processing, Scan Queue Sorting, and Processing Queue Warning Levels. The Custom Processing tab is active and contains sub-tabs: Batch Processing, Auto-Name, Capture, Indexing, Integrations, QA, and Image Processing. The Batch Processing sub-tab is selected, showing a list of checkboxes for various processing options. The "Automated Index Processing Queue" checkbox is checked. The "Image Segment Archiver" section has a checkbox for "Image Segment Archiver Queue" which is unchecked. The "Default Document Type" is set to "<None>". The "Initiate Workflow" section has a radio button for "During Commit" selected. The "Override Maximum Uncommitted Batches" field is set to 0.

Assign Process Options for AP - Vendor Invoices

Custom Processing | Scan Queue Sorting | Processing Queue Warning Levels

Batch Processing | Auto-Name | Capture | Indexing | Integrations | QA | Image Processing

☐ Separate Barcode Processing Queue

☐ Document Separation Queue

☐ Show Existing Document Separations

☐ Clear Keywords on Document Separation

☐ Advanced Mode

☐ Automated Index Processing Queue

☐ Automatic Automated Index Processing

☐ Intelligent Automated Index Processing Queue

☐ Automatic Document Separation During Processing

☐ Double-Blind Indexing Queue

☐ Document Re-Index Queue

☐ Confirm Batch Exit from Re-Index / Keep Documents Available

☐ Full-Page OCR Queue

☐ Auto-OCR Batches

☐ Respect Lines per Page

Image Segment Archiver

☐ Image Segment Archiver Queue

☐ Single Document Mode - Only During Indexing

Send to Queue:

Document Type:

☐ Continuous Page Slicing Mode

Default Document Type:

☐ Default to Pre-Index Scan Mode

☐ PDF Conversion Queue

☐ Auto-PDF Conversion

☐ Log Document Create Actions

☐ Log Keyword Modifications

☐ Auto-Perform Image Processing

☐ Auto-Commit Batches

☐ Auto-Name Batches

☐ Auto-Folder Documents

☐ Delay Disk Group Migration Until Batch Commit

☐ Perform Custom Step

☐ Auto-Run after Scan

☐ Create Image Thumbnails On Commit

☐ Do Not Allow Direct Scanning Into This Queue

☐ Exclude Queue from Disconnected Scanning

☐ Double-Click Initiates Queue Operation

Initiate Workflow

☒ During Commit

☐ During Index

☐ Do Not Initiate Workflow

Override Maximum Uncommitted Batches (This Scan Queue Only)

0 = Use Default and Check All Scan Queues

Save Cancel

3. Select the **Automated Index Processing Queue** check box. Batches imported via this scan queue will now be routed through the **Awaiting Automated Index** batch status queue.

4. If you would like batches associated with this scan queue to automatically undergo Automated Indexing as soon as the batch enters the **Awaiting Automated Index** batch status queue, select the **Automatic Automated Index Processing** check box.

If the **Automatic Automated Index Processing** check box is not selected, batches will reside in the **Awaiting Automated Index** batch status queue until they are processed manually; via a scheduled process, if your OnBase Client workstation is configured to do so; or until the workstation polls the batch status queue and processes any batches residing within it.

5. Click **Save** to save your scan queue configuration and close the **Assign Process Options for <Scan Queue Name>** dialog box.

Disabling Images of Keyword Values from Displaying During Indexing

When a user is performing index verification and selects a Keyword Type field that has been populated with a Keyword Value via Automated Indexing, by default, an image of the Keyword Value identified by the engine is displayed above the Keyword Type field.

The screenshot shows a document interface. At the top, 'Document Date' is 07/01/2010. Below it, a large '1001' is displayed. Under the 'Keywords' section, there are four fields: 'Invoice #' with value '1001', 'Invoice Date' with value '07/01/2010', 'Vendor' with value 'COMPUTERS R US', and 'P.O. Number' with value '4279'. Each field has a lock icon to its right.

If desired, your scan queue can be configured to not display this image of the Keyword Value above the Keyword Type field.

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.
2. Select the scan queue to be modified from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.
3. Select the Indexing tab.
4. Select the **Disable Floating Image Snippets Above Keywords** check box.
5. Click **Save**.

Disabling Automatic Keyword Value Zoom

When a user is performing index verification and selects a Keyword Type field that has been populated with a Keyword Value via Automated Indexing, by default, the Working window will zoom to the area of the document that the value was extracted from.

If desired, your scan queue can be configured to not automatically zoom to this area.

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.
2. Select the scan queue to be modified from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.
3. Select the Indexing tab.
4. Select the **Disable Automatic Keyword Value Zoom to Image Zone** check box.
5. Click **Save**.

Navigating Only Between Suspect Keyword Values

By default, if a user presses the **Tab** key when performing index verification, the next Keyword Value, regardless of it is suspect or not, is selected in the **Indexing** dialog box.

If desired, you can configure the **Indexing** dialog box to only navigate between Keyword Values that have been identified as suspect when a user presses the **Tab** key (i.e., when a user presses the **Tab** key, the next suspect Keyword Value is selected; all non-suspect Keyword Values between the last-selected Keyword Value and the newly-selected Suspect Keyword Value are skipped).

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.
2. Select the scan queue to be modified from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.
3. Select the Indexing tab.
4. Select the **TAB Cycles Through Suspect Keywords** check box.
5. Click **Save**.

When this option is selected, a user may navigate between all Keyword Values in the **Indexing** dialog box (suspect or non-suspect) by pressing the **Up Arrow** or **Down Arrow** keys.

Note: If the **TAB Cycles Through Suspect Keywords** check box and the **Tab to Next Field** radio button is selected in either the Main Enter Key or the Numeric Enter Key section, then pressing the **Enter** key will select the next suspect Keyword Value in the **Indexing** dialog box (i.e., non-suspect Keyword Values are skipped).

Configuring Automatic Automated Indexing Processes

Automating Indexing can be initiated on a batch manually, or it can be automated in several ways.

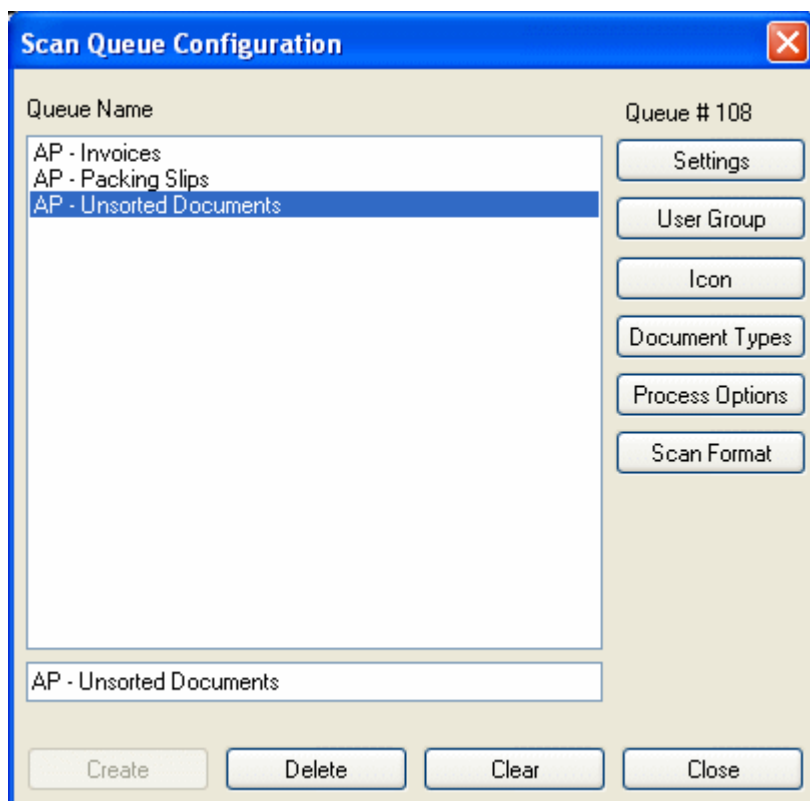
Note: For information on manually initiating Automated Indexing on a batch, see Performing Batch Automated Indexing on page 22.

Automating Automated Indexing at the Scan Queue Level

Scan queues can be configured to automatically process batches once they enter the **Awaiting Automated Index** batch status queue.

To configure a scan queue to automatically process a batch once it enters the **Awaiting Automated Index** batch status queue:

1. From the OnBase Configuration module, click **Import | Scan Queues**. The **Scan Queue Configuration** dialog box is displayed.



2. Select the scan queue to be configured to automatically process batches in the **Awaiting Automated Index** batch status queue from the Queue Name list and click **Process Options**. The **Assign Process Options for <Scan Queue Name>** dialog box is displayed.

Assign Process Options for AP - Vendor Invoices

Custom Processing | Scan Queue Sorting | Processing Queue Warning Levels

Batch Processing | Auto-Name | Capture | Indexing | Integrations | QA | Image Processing

☐ Separate Barcode Processing Queue

☐ Document Separation Queue

☐ Show Existing Document Separations

☐ Clear Keywords on Document Separation

☐ Advanced Mode

☐ Automated Index Processing Queue

☐ Automatic Automated Index Processing

☐ Intelligent Automated Index Processing Queue

☐ Automatic Document Separation During Processing

☐ Double-Blind Indexing Queue

☐ Document Re-Index Queue

☐ Confirm Batch Exit from Re-Index / Keep Documents Available

☐ Full-Page OCR Queue

☐ Auto-OCR Batches

☐ Respect Lines per Page

Image Segment Archiver

☐ Image Segment Archiver Queue

☐ Single Document Mode - Only During Indexing

Send to Queue:

Document Type:

☐ Continuous Page Slicing Mode

Default Document Type

<None>

☐ Default to Pre-Index Scan Mode

☐ PDF Conversion Queue

☐ Auto-PDF Conversion

☐ Log Document Create Actions

☐ Log Keyword Modifications

☐ Auto-Perform Image Processing

☐ Auto-Commit Batches

☐ Auto-Name Batches

☐ Auto-Folder Documents

☐ Delay Disk Group Migration Until Batch Commit

☐ Perform Custom Step

☐ Auto-Run after Scan

☐ Create Image Thumbnails On Commit

☐ Do Not Allow Direct Scanning Into This Queue

☐ Exclude Queue from Disconnected Scanning

☐ Double-Click Initiates Queue Operation

Initiate Workflow

☒ During Commit

☐ During Index

☐ Do Not Initiate Workflow

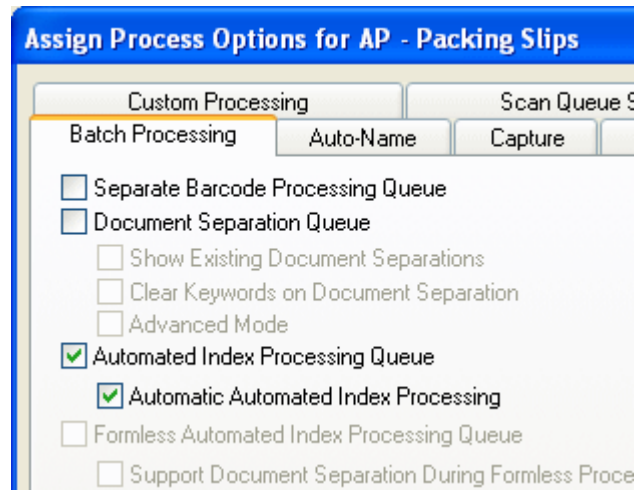
Override Maximum Uncommitted Batches (This Scan Queue Only)

0

0 = Use Default and Check All Scan Queues

Save Cancel

- On the **Batch Processing** tab, select the **Automatic Automated Index Processing** check box.



Note: The **Automated Index Processing Queue** check box must be selected in order for the **Automatic Automated Index Processing** check box to be enabled.

- Click **Save** to save the configuration and exit the **Assign Process Options for <Scan Queue Name>** dialog box.

Automating Automated Indexing via a Scheduled Process

Automated Indexing can be configured as a scheduled process for an OnBase Client registered for Automated Indexing.

For more information, see Scheduling on page 245.

Automating Automated Indexing via a Command Line Switch

If your workstation is registered for Automated Indexing, you can add a command line switch to your OnBase Client shortcut to automatically poll the **Awaiting Automated Index** batch status queue and automatically process the batches it finds there.

For more information, see your solution provider. Solution providers, see Command Line Switches on page 233.

Exiting the Automated Indexing Form Configuration Interface

Once you have finished configuring your Automated Indexing form(s), exit the Automated Indexing configuration interface and return to the **Document Imaging** window. To exit:

- Close the **Automated Index Configuration** window by clicking the **X** in the upper-right corner.
- Click **Exit Configuration** in the **Tools** window.

Tip: Prior to exiting the Automated Indexing configuration interface, be sure to save your changes by clicking **Save Configuration** in the **Tools** window.

INSTALLATION

REQUIREMENTS

OCR Processing Workstation Supported Operating Systems

- Windows XP SP3 or later service pack
- Windows Server 2003 SP2 or later service pack
- Windows Vista SP1 or later service pack
- Windows Server 2008 SP1 or later service pack
- Windows Server 2008 R2 RTM or later service pack
- Windows 7 RTM or later service pack

Note: As of release 7.2, OnBase no longer supports the use of the Windows® NT 4.0 and Windows® 98 operating systems. Microsoft retired its support of Windows NT 4.0 on January 1, 2005 and Windows® 98 on July 11, 2006. Microsoft no longer offers technical support or security updates for these operating systems.

Note: As of release 10.0.0, OnBase no longer supports the use of the Windows 2000 Professional Edition and Windows 2000 Server operating systems. Microsoft's extended support for these operating systems was retired on July 13, 2010. Microsoft will no longer offer technical support or security updates for these operating systems after that date. Additionally, these operating systems do not support Microsoft .NET Framework 4.0, which is an OnBase requirement as of version 10.0.0. If you are using either Windows 2000 operating system, you should not upgrade to OnBase 11.0.0 until you have upgraded to a Windows operating system supported by OnBase.

Operating System Requirements

The above Client module supported operating systems require both:

- Microsoft .NET Framework 4.0 (Client Profile)
- Microsoft Visual C++ 2010 Redistributable Package

Note: The Microsoft Visual C++ 2010 Redistributable Package (x86) is installed using the **vcredist_x86.exe** delivered with OnBase.

OCR Processing Workstation Hardware Requirements

Hardware	Minimum	Recommended
CPU	1 GHz	2 GHz or faster
Memory (RAM)	512 MB	1 GB or greater (2 GB or greater for Windows Vista and later versions of Windows) Note: Additional RAM may be needed if working with grayscale or color images.
Free Hard Disk Space	60 MB for the Hyland OCR Engine for OnBase 200 MB for the system files and software	500 MB
Screen Resolution	1024 x 768 (1280 x 800) Note: Using a lower resolution may result in a loss of functionality.	1280 x 1024 (1440 x 900 widescreen)

Databases Supported

The following tables list the databases supported in OnBase 11.0.0.

Microsoft SQL Server

Microsoft SQL Server	Additional Information
Microsoft SQL Server™ 2000 (SP4 recommended) Microsoft SQL Server 2005 (SP2 or later recommended) Microsoft SQL Server 2008 (RTM and SP1; SP1 recommended) Microsoft SQL Server 2008 R2	<p>Microsoft SQL Server 2000 and 2005 must be running in compatibility mode 7 or greater. Running in compatibility mode 6.5 or lower will result in errors during the upgrade process.</p> <p>SQL Server 2005 drivers must be upgraded to the Feature Pack for Microsoft SQL Server 2005 - December 2008 or a later feature pack.</p> <hr/> <p>Note: On January 11, 2011 Microsoft discontinued technical support for Microsoft SQL Server 7.0. As of release 11.0.0, Hyland Software no longer supports SQL Server 7.0.</p> <hr/> <p>Note: You must ensure that your SQL Server database client software version matches or exceeds the database server version. For example, if your database server is SQL Server 2005, verify that the database client is SQL Server 2005 (or later). Running a previous client version, such as SQL Server 2000, will result in system instability and memory issues. For instructions on determining your server and client versions, see Database Client / Server Version Compatibility on page 228.</p> <hr/>

Oracle

Note: If you are using an Oracle database, it is strongly recommended that you have a certified Oracle Database Administrator on staff.

Oracle	Additional Information
Oracle v 8.0.5.0 or later	Oracle version 8.0.5.0 can be used, but it is not recommended due to potential memory leaks. If Oracle 8.0.5.0 is used, a third-party ODBC driver is recommended.
Oracle 8i: 8.1.7.7 or later	ODBC drivers should be 8.1.7 or later. 8.1.6.x drivers have known issues and are not supported.

Oracle	Additional Information
Oracle 9i: Release 1 and Release 2 (9.2) Oracle 10g: Release 1 and Release 2	<p>Oracle driver version 10.2.0.3 is recommended.</p> <p>An issue has been observed with the Oracle ODBC drivers where Unicode characters (e.g., Japanese characters) retrieved from a CLOB data type will be truncated, and could potentially cause errors in OnBase. The data remains intact in the database; however the results when retrieving the data will be truncated. This has specifically been observed in two areas of OnBase that use the CLOB data type to store large amounts of data: VBScripts and License Certificates. This behavior may apply to other areas of the software that use this data type as well.</p> <p>To ensure that Unicode characters retrieved from a CLOB data type are not truncated, the Oracle 10g R2 ODBC drivers (which are backward compatible) should be installed, as well as the latest patchset (version 10.2.0.3) for these drivers.</p>
Oracle 11g: Release 1 and Release 2	All Oracle 11g drivers can be used.

Sybase SQL Anywhere

Sybase SQL Anywhere	Additional Information
Sybase SQL Anywhere™ 5.5.04, 7.0.0-7.0.3, 8.0.1, 9.0, 10 and 11	<p>No 6.0.X releases.</p> <hr/> <p>Note:As of release 8.2.0, Sybase Adaptive Server Enterprise (ASE) is no longer supported.</p> <hr/>

Database Client / Server Version Compatibility

Due to critical issues that have been reported to Hyland Software, Hyland Software strongly recommends that:

- your database client software version matches or exceeds the database server version and
- you are running the most recent version of the database client.

This will help to reduce compatibility issues and minimize troubleshooting time when issues do occur.

Your database administrator can determine the database server version and identify the most-recent version of the database client software. The ODBC driver number indicates which version of the database client software you are using. For example, if your database server software is Oracle 10 Release 2, verify that the Oracle Client software is Oracle 10 Release 2 (or later). The same is true of SQL databases. For example, if your database server is SQL Server 2005, verify that the database client is SQL Server 2005 (or later).

To check your database client version, perform the following steps from the workstation or server where the ODBC connection is configured:

1. Open your ODBC Data Source Administrator, and click on the **Drivers** tab.
2. Select the driver you are using to connect to your OnBase database.
 - If your database server software is Oracle 10 Release 2, the version number should appear as **10.2.[#.#.#]** (or later), where **10.2** is the version number and **[#.#.#]** represents the service pack.
 - If your database server software is SQL Server 2005, the version number should appear as **2005.[##.####.##]** (or later), where **2005** is the version number and **[##.####.##]** represents the service pack.

The above descriptions are examples of two commonly used database version numbering schemes. Ensure that the supported database you use adheres to the database client/server recommendation.

About Virtual Environments

Hyland Software develops, tests, and supports the OnBase suite of products on specific Operating Systems, not specific hardware configurations. When OnBase is operated in a virtual environment (such as Citrix, VMware, Hyper-V, or Windows Remote Desktop) there may be limitations or subtle differences imposed by the environment. The customer and the virtual environment vendor are responsible for any interactions or issues that arise at the Hardware or Operating System layer as a result of their use of a virtual environment.

When it appears that an OnBase performance-related issue is either caused by (or is unique to) the virtual environment, organizations may be asked to validate that the issue occurs in a non-virtual environment. Hyland Software will make this request if there is reason to believe that the virtual environment is a contributing factor to the issue.

Each OnBase site is unique. Hyland Software depends on the customers who deploy OnBase in virtual environments to do so only after careful design and adequate planning (that takes into account the workloads of your organization), and in accordance with recommendations provided by the virtual environment's vendor. As with any implementation, Hyland Software strongly recommends that any customer deploying an OnBase solution in a virtual environment thoroughly test the solution before putting it into production.

For information about using OnBase in a Citrix and Microsoft Windows Remote Desktop environment, please see the **OnBase in a Citrix and Microsoft Windows Remote Desktop Environment** reference guide, available from your solution provider.

64-Bit Support Statement

The OnBase suite of products is tested on 64-bit systems and is capable of being deployed on 64-bit systems using the Windows 32-bit on Windows 64-bit Emulator (WOW64) layer. However, OnBase modules that integrate with third-party applications may not be able to be used with the 64-bit versions of these applications. For these modules, only the 32-bit versions of these third-party applications are currently supported by the OnBase integrations. Consult the module-specific requirements section in each module reference guide for complete requirements details.

Supported database versions that are deployed on a 64-bit database server are also supported. For more information, contact your solution provider.

4

Licensing

See Licensing on page 4 for licensing requirements.

PRE-INSTALLATION

In order to use Automated Indexing, your OnBase solution must be properly licensed and configured to import batches of image or PDF documents into OnBase scan queues.

A number of OnBase modules, such as Document Imaging, Disconnected Scanning, or Document Import Processor (DIP), can be used to import documents into batches via scan queues.

Hyland OCR Engine for OnBase

OnBase Intelligent Automated Indexing, Automated Indexing, Batch OCR, Payee Extraction, and Automated Redaction use the Hyland OCR Engine for OnBase as their recognition package.

The Hyland OCR Engine for OnBase (version 16.5 or later) is required for these modules. This software is included with the purchase of these modules, and is available from your OnBase solution provider. It does not require separate licensing or registration.

Additional Information

- You must install the Hyland OCR Engine for OnBase on each workstation that will perform Automated Indexing, Batch OCR processing, Payee Extraction, and Automated Redaction.
- Intelligent Automated Indexing requires the OCR engine be installed on the server hosting the Intelligent Automated Indexing NT Service.
- The Hyland OCR Engine for OnBase does not support multi-threaded access for Automated Indexing, Batch OCR, Automated Redaction, and Payee Extraction. Multi-threaded access is supported for Intelligent Automated Indexing.

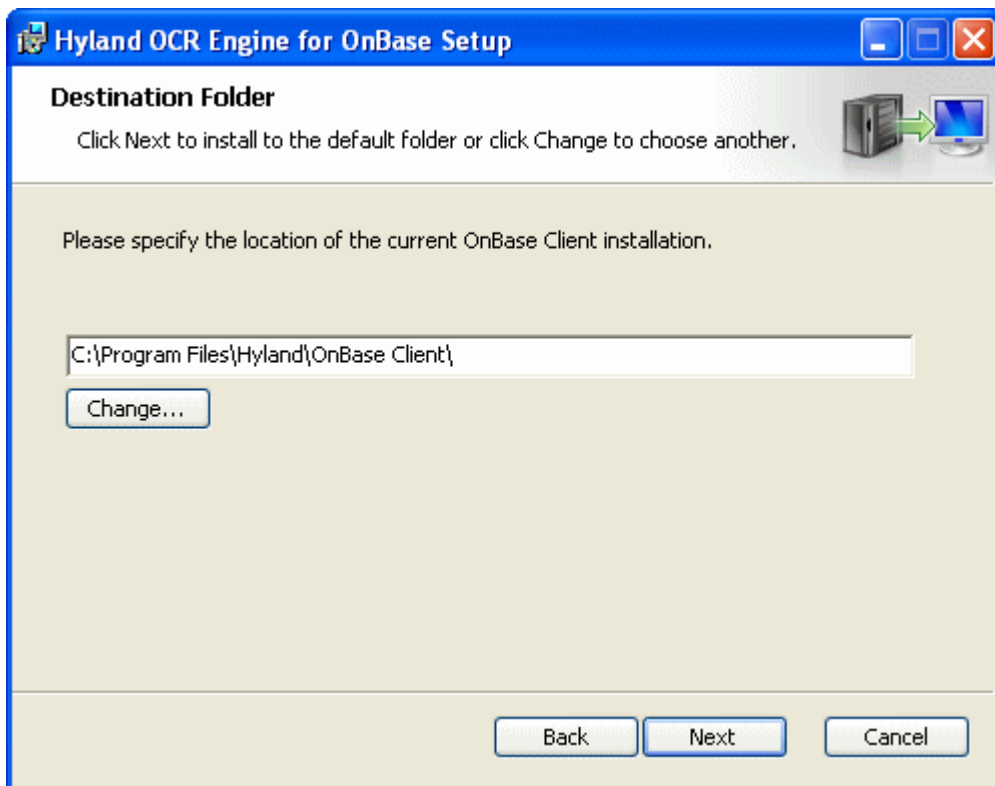
INSTALLATION

With the exception of the installation of the Hyland OCR Engine for OnBase, no additional installation is required for the Automated Indexing module.

To install the Hyland OCR Engine for OnBase:

Obtain the Hyland OCR Engine for OnBase installer from Technical Support and copy it to each machine that will be performing OCR processing.

1. Double-click the Hyland OCR Engine for OnBase executable to launch the installer. The setup wizard's **Welcome** screen is displayed.
2. Click **Next**. The **End-User License Agreement** screen is displayed. In order to proceed with the installation, you must agree to the terms of the EULA. If you agree, select the **I accept the terms in the License Agreement** check box and click **Next** to proceed to the **Destination Folder** screen.



3. On the **Destination Folder** screen, you must specify the location of the OnBase Client executable on the Automated Indexing processing workstation.

The default installation location (**C:\Program Files\Hyland\OnBase Client**) is automatically selected. To select a different location, manually enter the path in the **Destination Folder** field or click **Change** to browse to the location.

When the destination folder has been selected, click **Next**.

4. The **Ready to install Hyland OCR Engine for OnBase** screen is displayed. Click **Install** to begin the installation.
5. Once the installation is complete, click **Finish**.

COMMAND LINE SWITCHES

The following command line switches can be used in conjunction with Automated Indexing:

- **-SCHED**. This switch is used if the workstation will be performing scheduled Automated Indexing processes.

Auto-Processing Switches

-SCANAUTOINDEX

When this switch is used, the workstation will poll the **Awaiting Automated Indexing** queue and automatically process batches of documents found within it.

-SCANAUTOQUEUE

This switch is used in combination with any of the auto-processing command line switches to direct the processing workstation to only automatically process batches from a specific scan queue. The scan queue(s) to be processed are identified by their Scan Queue Number.

For example, to automatically perform OCR processing on batches scanned using scan queue #101, you would add the following switches, in this order, to the Client module's shortcut:

-SCANAUTOOCR -SCANAUTOQUEUE:101

Multiple scan queues can be specified by separating the Scan Queue Numbers by a comma (i.e., **-SCANAUTOQUEUE:101,102,103**).

Tip: The Scan Queue Number of a scan queue is displayed when the scan queue is selected in the **Scan Queue Configuration** dialog box (click **Import | Scan Queues** from the OnBase Configuration module).

AutoProcessDelay INI Setting

A timer is used to poll the batch status queues affected by these switches to search for new batches to process. By default, this timer is set to 300 seconds (5 minutes), but this value can be changed by modifying the **AutoProcessDelay** value in the onbase32.ini file.

For more information on this setting, see the information on the **AutoProcessDelay** setting in the **INI Settings** section below.

INI OPTIONS

The following INI settings are related to Automated Indexing:

Scan

The **[Scan]** section contains the following INI setting:

INI Setting Name	Description
AutoProcessDelay	This value sets the number of seconds allowed to pass between each polling of batch status queues when the - SCANAUTOINDEX command line switch is used. By default, this setting is set to 300 .

AutomatedIndex

The **[AutomatedIndex]** section contains the following INI setting:

INI Setting Name	Description
Diagnostic	When enabled (Diagnostic=1), this option will add the information from the Result Verification window of the Automated Indexing Configuration window to the Verification Report. When this option is not enabled (Diagnostic=0), the Result Verification information is not included in the Verification Report.
AnalysisMin	This option sets the minimum threshold allowed for Automated Indexing form order analysis. By default, it is set to 100. (AnalysisMin=100). Note: This number refers to the number of documents that have been matched to a form, not the number of documents available in your system. Note: If your solution does not contain more documents than the minimum specified by the AnalysisMin setting, you will not be able to perform form order analysis.

INI Setting Name	Description
AnalysisMax	<p>This option sets the maximum threshold allowed for Automated Indexing form order analysis. By default, it is set to 30,000. (AnalysisMin=30000).</p> <hr/> <p>Note: This number refers to the number of documents that have been matched to a form, not the number of documents available in your system.</p> <hr/> <p>Note: If your solution contains more documents than the maximum number of allowed documents, you are able to perform form order analysis; however, the analysis is only performed on the number of documents specified by the AnalysisMax setting (the excess documents are not included in the analysis).</p> <hr/>

INI File

INI files (initialization files) are plain-text files that contain configuration information. These files are used by Windows and Windows-based applications to save and access information about your preferences and operating environment. OnBase uses an initialization file named onbase32.ini. If a user does not have rights to access the onbase32.ini file, that user will be unable to use the Client or Configuration modules.

The onbase32.ini file is primarily used to store settings specified in the Client or Configuration module. For example, when a user selects a default data source in the OnBase Client's Workstation Options dialog box, this selection is saved to the onbase32.ini file. The onbase32.ini file is also used to make modifications to OnBase modules that cannot be made through the module's interface.

Previous File Location/File Name

Every version of the OnBase Client prior to 8.2.0 used an INI file named OnBase.ini. In OnBase 8.2.0 and subsequent versions, the INI file was moved to a new location to be consistent with changes Microsoft has made to Windows. Since the location has changed, the name of the file has also been changed to alleviate some confusion between the needs of OnBase 8.2.0 and installations of older executables. The new file name is onbase32.ini.

Location

The table below shows the default location of the onbase32.ini for supported operating systems.

Operating System	Default Location
Windows XP	C:\Documents and Settings\All Users\Application Data\Hyland Software
Windows Server 2003	C:\Documents and Settings\All Users\Application Data\Hyland Software
Windows Vista	C:\ProgramData\Hyland Software
Windows Server 2008	C:\ProgramData\Hyland Software
Windows Server 2008 R2	C:\ProgramData\Hyland Software
Windows 7	C:\ProgramData\Hyland Software

Note: To maintain backwards compatibility with previous versions of OnBase, OnBase will check the workstation's **C:\Windows** folder for the OnBase INI file if it is not found in the folder specified above. If the OnBase INI file is found in the **C:\Windows** folder, OnBase will copy the file to the new location. The previously-existing version of the OnBase INI file will remain in the **C:\Windows** folder, but will no longer be used by OnBase.

Your onbase32.ini file may reside in a different location, if that location is specified by the following command line switch on the OnBase Client shortcut target.

-INIFILE= "full path\filename", where **full path** and **filename** are replaced by the specific path and file name.

If this command line switch is not used and you move or rename your onbase32.ini file, OnBase will recreate the file in the default folder and ignore the newly created file.

INI Considerations in a Citrix and Microsoft Windows Remote Desktop Environment

In remote desktop environments, a remote session is established in which the user is running applications that are not installed locally. This presents a challenge when an application, such as OnBase, requires a user-specific INI file to establish unique settings. In a remote desktop environment, you must ensure that each user has a single, unique INI file to make sure any user-specific settings are consistent for that user.

Note: The default location of the OnBase INI file is not unique in a remote desktop environment.

To ensure that the INI file is accessible by OnBase and unique to each user in a remote desktop environment, the **-INIFILE** command line switch must be applied to the OnBase Client shortcut and be set to a unique location for the INI file.

Note: Additional details regarding the deployment of OnBase in a remote desktop environment is discussed in detail in the **OnBase in a Citrix and Microsoft Windows Remote Desktop Environment** module reference guide, available from your first line of support.

Editing the INI File

Users with the **Configuration** Product Right can open the onbase32.ini file from the OnBase Client by selecting **Admin | Utilities | Edit INI File**. When multiple onbase32.ini files exist, opening the onbase32.ini file from the OnBase Client ensures that a user is editing the correct onbase32.ini file instance. In most cases, this will be the onbase32.ini file residing in the default directory described above. If an alternate location for the onbase32.ini file is specified by the **-INIFILE** command line switch, the file in the specified location will be opened.

BACKUP/RECOVERY

Backup

Configuration

The Automated Indexing configuration is stored in the OnBase database. A proper backup of the database will contain all Automated Indexing configuration information, including all Automated Indexing forms.

Registry Settings

No registry settings apply to Automated Indexing.

External Files

Ensure that the Hyland OCR Engine for OnBase software has been properly backed up. You must also backup your onbase32.ini file.

Module-Related INI Options

Use the following chart to track the current settings of all related INI options for Automated Indexing.

Section	Setting	Current Value
AutomatedIndex	Diagnostic	
	AnalysisMin	
	AnalysisMax	

Additional Steps

There are no additional steps required to backup Automated Indexing.

Recovery

Configuration

All Automated Indexing configuration, including all Automated Indexing forms, is stored in the OnBase database. Restoring the database will restore any Automated Indexing configuration.

Registry Settings

No registry settings apply to Automated Indexing.

External Files

Ensure that your onbase32.ini file is properly restored to the correct location by the backup copy.

The onbase32.ini file can be restored from the backup if the recovery machine is intended to be used for exactly the same purpose as the original machine. If this machine will be used for other modules, thorough testing may be required to ensure that there are no unintended consequences from adding/modifying INI settings on the existing workstation.

Restore the Hyland OCR Engine for OnBase by re-installing it on the processing workstation using the **OCRSetup** executable.

Module-Related INI Options

The onbase32.ini file can be restored from the backup if the recovery machine is intended to be used for exactly the same purpose as the original machine. If the recovery machine will be used for other modules, you may need to recover only the listed INI setting(s) from the table above.

Registration

Migrate the registration of Automated Indexing from the original workstation to the recovery workstation. The registration may need to be revoked from the original machine and then added to the recovery machine.

Additional Steps

No additional recovery steps are required for Automated Indexing.

TROUBLESHOOTING

Common Issues

When I attempt to extract Keyword Values from an image using Point and Shoot Indexing, the text values that are returned are consistently incorrect and/or scrambled.

Most likely, the cause of this problem is the quality and/or resolution of your image. Try performing Point and Shoot Indexing on a “cleaner” image file.

It is considered a best practice to ensure that an image has a resolution of at least 240 dpi in order for acceptable results to be returned when perform Automated Indexing or Point and Shoot Indexing upon it.

I have been having trouble with the Automated Indexing engine incorrectly mistaking some alphabetic characters for numeric characters (i.e., reading INVO1CE instead of INVOICE). In order to prevent these indexing errors, I enabled the Allowed character types filtering option to prevent numeric characters from being recognized in fields that are only supposed to contain alphabetic characters. However, instead of routing the documents for manual indexing as I expected, the Automated Indexing engine now correctly reads the value.

When the **Allowed character types** filter options are used, the Automated Indexing engine is better able to “guess” what character SHOULD be read.

For example: you are using Automated Indexing to read the value **INVOICE** from a document, but the engine occasionally has trouble distinguishing the **I** from a **1**, and returns the value as **INVO1CE**. If the filter options are used to prevent Automated Indexing from reading numeric characters for that value, the engine will favor the **I** character over the **1** character, and will be more likely to return the value correctly as **INVOICE**.

Note: If the Automated Indexing engine, however, is confident that the value is, in fact, a **1** and not an **I**, then the **1** character will be replaced with a tilde (~) and the value will be marked as suspect.

But, if you are not using the filter options to prevent Automated Indexing from reading numeric characters, the engine will have no reason to favor the **I** character over the **1** character, other than its normal character recognition logic. Only if the engine is unsure of, and gives a high suspect level to, both the **I** and the **1** characters will the value itself be marked as suspect.

My Automated Indexing forms correctly identify shorter Keyword Values, but longer Keyword Values are consistently rejected and the documents are marked as suspect.

If the Keyword Value identified by the Automated Indexing engine exceeds that maximum length of the Keyword Type it is to be assigned to, then the Keyword Value is rejected and the document is marked as suspect.

How can I configure the Date Format for Date Keyword Values?

Automated Indexing always assumes that the Date Format of Date Keyword Values captured by the Automated Indexing engine will match the workstation's regional format.

For example, if your Automated Indexing workstation is set to use **English (United States)** as its workstation locale, then all dates captured by the Automated Indexing engine must be in the MM/DD/YYYY format.

If you are attempting to capture Date Keyword Values in a format that does not match the workstation's regional format, then you must configure a VB script for the Data Field Zone configured to parse the value into the correct format.

I noticed that when PDF documents undergo OCR processing, the results are less accurate than they previously were.

It is a known issue with the Hyland OCR Engine for OnBase version 16.3 that when PDF documents undergo OCR processing, the resulting text rendition is less accurate than in previous versions of the Hyland OCR Engine for OnBase.

It is recommended that you upgrade to the Hyland OCR Engine for OnBase version 16.5 or later to correct this issue.

The same issue does not affect image documents.

Process Locks

Another administrator is configuring Automated Indexing forms for a scan queue, and I am trying to configure Automated Indexing templates for a different scan queue. We have sufficient Automated Indexing licenses available, but I keep getting a message that I am locked out.

I am attempting to configure Automated Indexing forms, but I keep getting a locked message even though no other users are currently logged on.

Only one OnBase administrator can access the Automated Indexing configuration interface at a time, regardless of the number of Automated Indexing licenses that are available. If another administrator is currently using the Automated Indexing configuration interface, then you will not be able to access it until that user has exited it.

You can check to see if another user is currently configuring Automated Indexing by viewing the Process Locks in the OnBase Client by clicking **Admin | Utilities | Process Lock Administration**.

If you are getting the locked message and you are sure that no other users could be using the Automated Indexing configuration interface, it is possible that a process lock was not removed when a user exited the Automated Indexing configuration interface.

For information on process locks, including information on deleting unnecessary process locks, see the System Administration documentation.

Automated Indexing & AutoFill Keyword Sets

I am attempting to store AutoFill Keyword Set data from documents indexed via Automated Indexing process using the scan queue's Store Keyword Set Data option, but the AutoFill Keyword Set data is saved inconsistently.

Instances of AutoFill Keyword Sets can automatically be created using data applied to a document from a Batch Automated Indexing process if the associated scan queue has the **Store Keyword Set Data** option enabled. AutoFill Keyword Set data is only stored for documents that are indexed correctly and marked as fully-indexed; AutoFill Keyword Set data will not be stored for any document that is unidentified or has any values marked as suspect.

AutoFill Keyword Set information cannot be stored for any document indexed via an Ad-Hoc Automated Indexing process because the document does not have a connection to a scan queue and, therefore, the **Store Keyword Set Data** scan queue option.

I am using Automated Indexing to identify a Keyword Value that is used as the primary Keyword Value to trigger an AutoFill Keyword Set. However, when the primary Keyword Value is associated with more than instance of an AutoFill Keyword Set, I'm not given the option to select the instance of the AutoFill Keyword Set(s) I want.

If you are using Ad-Hoc Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:

- If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, no instance of the AutoFill Keyword Set is assigned to the document (even though the primary Keyword Value is assigned to the document). This is the default behavior.
- If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

Regardless of the setting of the **Expand All Matching Instances** option you are not given the opportunity to select the instance of the AutoFill Keyword Set or, depending on your configuration, select multiple instances of the AutoFill Keyword Set to assign to the document.

If you are using Batch Automated Indexing to extract a Keyword Value that is the primary Keyword Value for an AutoFill Keyword Set, be aware that if multiple instances of the AutoFill Keyword Set exist with the same primary Keyword Value, the behavior of the AutoFill Keyword Set depends on how it was configured:

- If the **Expand All Matching Instances** option was not selected for the AutoFill Keyword Set, the batch is routed to the **Index in Progress** batch status queue so that you can select the instance, or instances, of the AutoFill Keyword Set to assign to the document. This is the default behavior.
- If the **Expand All Matching Instances** option was selected for the AutoFill Keyword Set, all instances of the AutoFill Keyword Set are assigned to the document.

I am using Point and Shoot Indexing to extract a Keyword Value that is used as the primary Keyword Value to trigger an AutoFill Keyword Set. However, after I obtain the Keyword Value, I have to tab to the next Keyword field to trigger the Autofill Keyword Set expansion.

When using Point and Shoot Indexing, be aware that the focus remains on the field containing the just-extracted Keyword Value.

If you are using Point and Shoot Indexing to extract a Keyword Value that is the primary value for an AutoFill Keyword Set, the AutoFill Keyword Set is not triggered until you set the focus to another Keyword field.

CONTACTING SUPPORT

When contacting your solution provider, please provide the following information:

- The OnBase module where the issue was encountered.
- The OnBase version and build (Example: 11.0.0.571) and/or the Core Services version and build (Example: 11.0.0.6).
- The type and version of the connected database, such as Microsoft SQL Server 2008 or Oracle 11g, and any Service Packs that have been installed.
- The operating system that the workstation is running on, such as Windows XP or Windows Server 2008, and any Service Packs that have been installed. Check the supported operating systems for this module to ensure that the operating system is supported.
- The name and version of any application related to the issue.

- The version of Internet Explorer, and any Service Packs that have been installed, if applicable.
- A complete description of the problem, including actions leading up to the issue.
- Screenshots of any error messages.

Supplied with the above information, your solution provider can better assist you in correcting the issue.

SCHEDULING

SCHEDULING OVERVIEW

Processing incurs a large drain on your OnBase solution's resources. Scheduling processing for off-hours is an automated way to conserve system resources. Processing can be accelerated if the process is run from the database server.

Note: Purging documents from Document Maintenance can also be scheduled. For more information, see the System Administration documentation.

Two types of processing activities may be scheduled with the Scheduler: a Process Format or a Process Job.

- A Process Format is used in processing modules and in scanning modules to specify how OnBase processes data being imported into OnBase. A Process Format is, basically, one individually-configured process.
- A Process Job is one or more Process Formats that have been configured to run sequentially. A Process Job does not have to consist exclusively of a single type of Process Format; it can contain multiple Process Formats from any module that allows scheduling.

Note: Process Formats created from Document Imaging sweep or scan from disk processes cannot be included in a Process Job.

CONFIGURING & USING THE SCHEDULER

Requirements for Configuring/Running a Scheduled Process

To configure a scheduled process, either a Process Format or a Process Job, a user must belong to a user group with the **Client** and **Scheduler** product rights, and he/she must have rights to use the appropriate processing module. A scheduled process can be configured on any OnBase Client workstation, not just the processing workstation or a workstation running with the **-SCHED** command line switch.

A

To run a scheduled process, OnBase must be running with the **-SCHED** or **-SCHEDINST** command line switch on the processing workstation in order for the scheduled process to be executed at the configured time. The user account logged onto OnBase at this time needs only the **Client** product right in order for the process to be performed.

Using the -SCHED and -SCHEDINST Switches

This section explains the difference between the **-SCHED** and **-SCHEDINST** command line switches.

-SCHED

Some process formats or jobs can be scheduled to run automatically. The **-SCHED** switch causes the Client to queue these scheduled process formats and jobs for later processing; if the machine running the OnBase Client in Scheduler mode (i.e., running the OnBase Client with the **-SCHED** command line switch applied) is also the processing workstation, then the process formats or jobs will run at their scheduled times.

In order for the scheduled process format or job to be run, OnBase must be running in Scheduler mode on the processing workstation. If OnBase is not running, or if OnBase is not running in Scheduler mode, then the scheduled processes will not run.

A process format or job can be scheduled from any OnBase Client workstation by a user with the proper rights.

-SCHEDINST

The **-SCHEDINST** command line switch is very similar to the basic **-SCHED** switch. When you apply the **-SCHEDINST** switch to a Client shortcut, you can specify that the selected instance of the OnBase Client should only process jobs assigned to that Client instance's specific instance name.

The format of the switch is **-SCHEDINST="MyProcName"**, where **MyProcName** is the name of a specific processing instance. The OnBase Client that this switch is applied to will be unable to process any scheduled jobs that are not configured with a **Specific Processing Instance** of **MyProcName**.

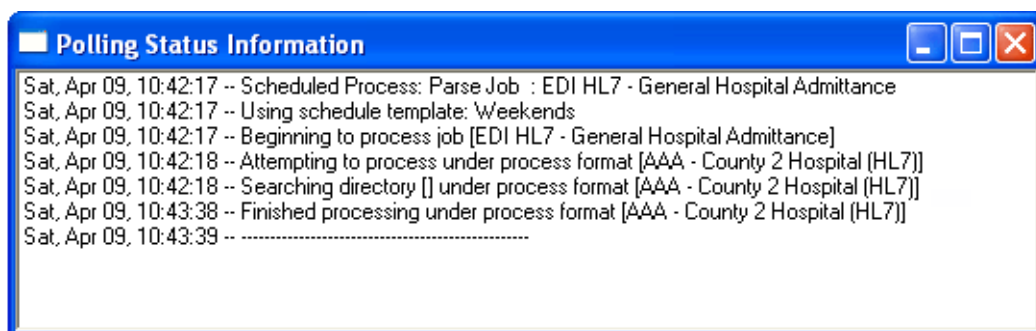
A process format or job can be scheduled from any OnBase Client workstation by a user with the proper rights.

Verifying the Scheduler is Running

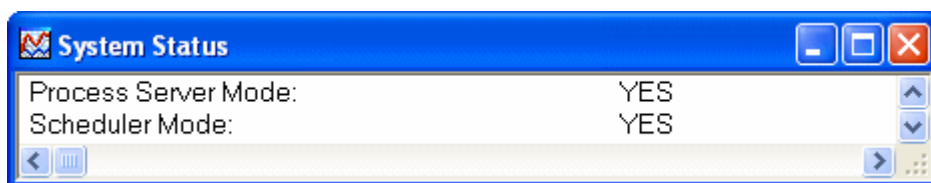
To verify that the Scheduler is running on the processing workstation, click **Window | Polling Status Information** in the OnBase Client.

Note: The **-SCHED** or **-SCHEDINST** command line switch must be applied to the Client shortcut to use this option.

The **Polling Status Information** window is displayed. Information about scheduled processes is displayed in it as the process is run. If this window exists, the Scheduler is running.



Another way to verify the Scheduler is running is to select **Window | System Status**. Both **Process Server Mode** and **Scheduler Mode** will be displayed as **YES**.



Running Multiple Scheduled Processes

Tip: Attempting to run more than one process job or format at once in the same session will result in a dramatic drop in all processing speeds. It is recommended to run a single automated process at a time.

If multiple jobs are configured, they can be performed sequentially in one OnBase Client session on the same workstation. Multiple sessions of the OnBase Client can be run simultaneously on one workstation to process these jobs in parallel; these sessions will coordinate processing tasks to ensure that each job is processed and that a job is not processed more than once.

A

In order to process jobs in parallel on multiple sessions of the OnBase Client, each session must be OnBase version 8.3.0.552 or later. If any one of the sessions is running an earlier version of OnBase, then none of the other sessions will perform any processing while it is processing.

Tip: It is recommended that all sessions use at least OnBase 8.3.0.552 or later if you plan on processing jobs in parallel using multiple session of the OnBase Client.

Scheduled Process Configuration Reports

A user belonging to a user group with the proper rights can run a Scheduled Processes Configuration Report.

This report provides information on all of the scheduled processes (process formats and process jobs) that have been scheduled to run. It is organized by processing workstation, and displays a weekly, monthly and end-of-month schedule, with jobs listed in order by starting time. Once run, this report is stored in OnBase as a document belonging to the **SYS Configuration Reports** Document Type.

Tip: It is considered a best practice to run a new Scheduled Process configuration report each time a new process (process format or process job) is scheduled. With the information stored in this report, troubleshooting and communications with Technical Support are greatly improved. Additionally, Configuration Reports are stored in OnBase, so there is a historical record of the structure of your OnBase solution.

For more information on Configuration Reports, including the Scheduled Processes Configuration Report, see the System Administration documentation.

A

Working With Process Formats

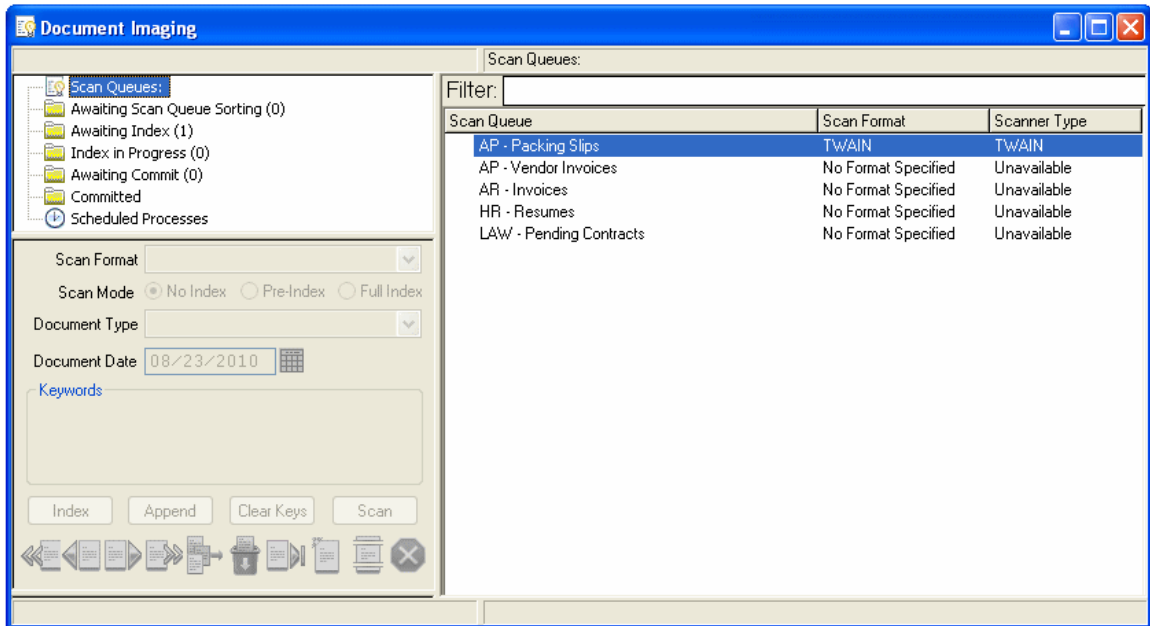
A Process Format is used in processing modules and in scanning modules to specify how OnBase processes data being imported into OnBase. A Process Format is, basically, one individually-configured process.

Creating a Scheduled Process Format

You can add a format to the Scheduler from its process queue by selecting the process format and selecting a scheduling option from the right-click menu.

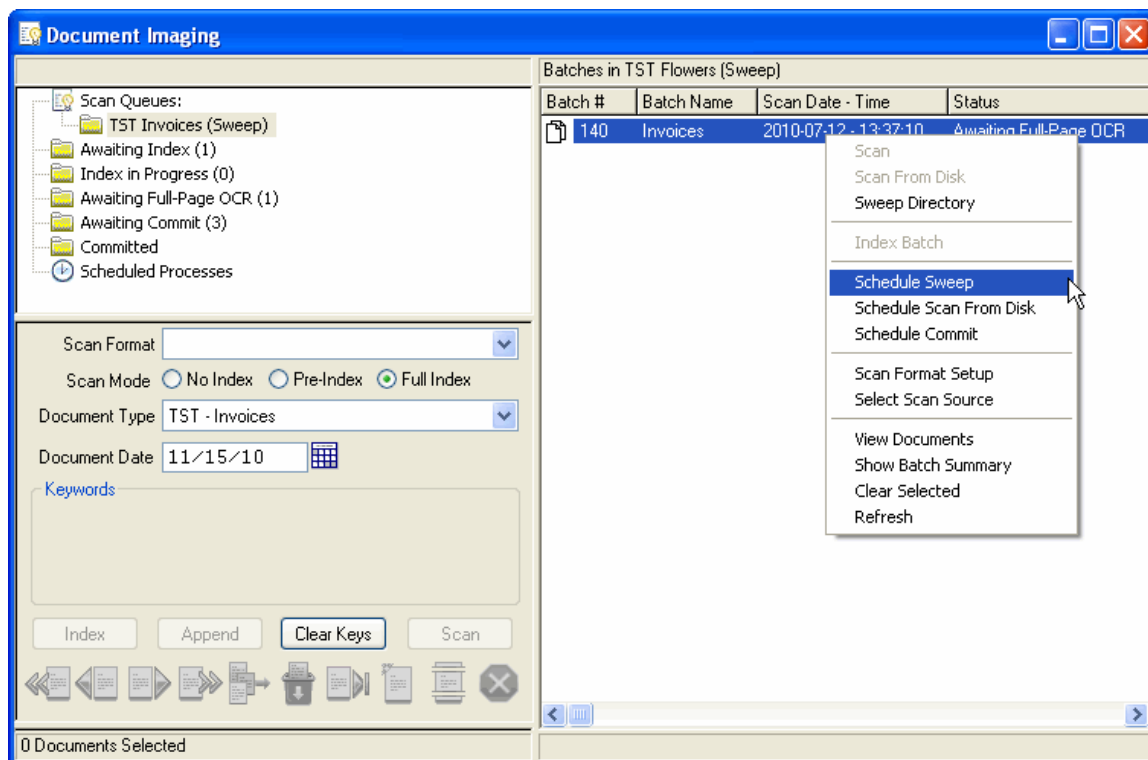
For example:

In the OnBase Client, click **Processing | Scan/Index**. The **Document Imaging** window is displayed.



Double-click a process format you would like to add to the Scheduler. A list of available batches is displayed.

Select the batch you want to schedule, then right-click and select **Schedule Sweep**, **Schedule Scan From Disk**, or **Schedule Commit**.



The **Process Scheduling** window is displayed.

Name	Scheduled Item	Action	Workstation
Parse Format : COLD Proce...	COLD Process	Parse Format	DEVm006904 (AUTO)

Run Now
Suspend
Remove

☐ Hide other process types

Schedule Configuration | Processing Options

Name: Parse Format : COLD Process

Processing Workstation: DEVm006904 (AUTO)

☐ Specific Processing Instance:

Schedule Template: Weekends

Weekly

Sun Mon Tue Wed Thu Fri Sat

• • • • • •

Default Daily Schedule | Selected Day

Time Range: Start Time: 8:00 AM End Time: 5:00 PM

Processing Frequency

☐ Once then Suspend
☒ Once per Day
☐ Once Every Minutes

Apply OK Cancel

A new Process Format is added to the **Scheduled Items** box. It is automatically selected. By default, all scheduled Process Formats (e.g., COLD Process Formats, DIP Process Formats, etc.) are displayed in the **Scheduled Items** box when scheduling a new Process Format. For information on viewing only the Process Formats for the currently-selected process type, see Viewing Scheduled Processes on page 260.

SCHEDULE CONFIGURATION

A

The first options that must be configured for the scheduled process are the Schedule Configuration options on the **Schedule Configuration** tab. This tab is displayed by default.

1. In the **Name** field, enter a name for the scheduled process.
2. Using the **Processing Workstation** drop-down, select the workstation that will be used to run the scheduled process.

Note: This workstation will need to be running with the **-SCHED** or **-SCHEDINST** command line switch in order to run the scheduled process.

3. If you always want the scheduled process to be run from a specific instance of the OnBase Client, select the **Specific Processing Instance**, then enter the name of the instance in the **Specific Processing Instance** text field.

Note: If you select the **Specific Processing Instance** option but leave the **Specific Processing Instance** text field blank, the scheduled process can be run from any instance of the OnBase Client.

4. Using the **Schedule Template** drop-down, select one of the schedule templates for the process or select **<Custom Schedule>** to manually configure the schedule for this process.

Note: For information on creating a **Custom Schedule** or **Schedule Template**, see below.

5. Select how often you would like the scheduled process to run by selecting one of the Processing Frequency radio buttons.
 - **Once then Suspend.** The scheduled item will be processed once, then the scheduled process is suspended.
 - **Once per Day.** The scheduled item will be processed once per day.

Note: If the scheduled item is modified, the process may be run again on the same day.

- **Once every "" Minutes.** The scheduled item is processed in the interval (measured in minutes) entered in the field. The maximum number of minutes that can be entered is 99999.

6. When you are finished setting the **Schedule Configuration** options, click **Apply**.

Calendar

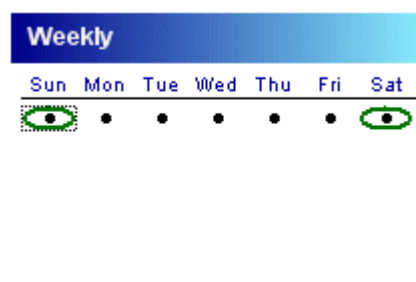
The calendar is used to select the day(s) on which a scheduled process should be run.

Note: The calendar is displayed based on your Workstation Regional Settings and the OnBase language DLL that you are using.

To change the view of the calendar, click the calendar heading (in the example above, **Weekly**) to display a menu. Select one of the following options to display a different calendar for configuration:

- **Weekly.** Allows you to configure a process to run on a certain day of the week (i.e., Thursday).
- **Monthly.** Allows you to configure a process to run monthly, on a particular date (i.e., the 1st and 15th of the month).
- **Monthly (Day-Relative).** Allows you to configure a process to run on a relative day of the month (i.e., the first Saturday of the month, the 2nd Wednesday of the month).
- **Annual.** Allows you to configure a process to run on a certain day of the year (i.e., June 30).
- **Full Calendar.** Allows you to configure a process to run on specified days of specified years (e.g., August 10, 2011 and/or July 17, 2012).

To select days that you would like to run a scheduled process, double-click the day on the calendar. The selected day is circled.



Note: In the example above, two days are selected but **Sunday** is the currently-selected day.

To deselect a day, double-click it.

Default Daily Schedule

The **Default Daily Schedule** tab allows you to configure the processing configuration for all days that do not have a **Selected Day** tab configuration.

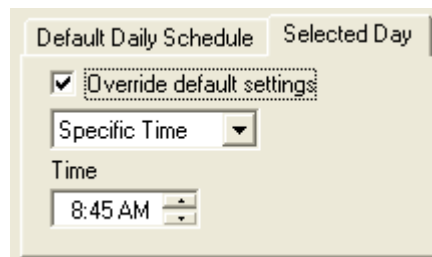
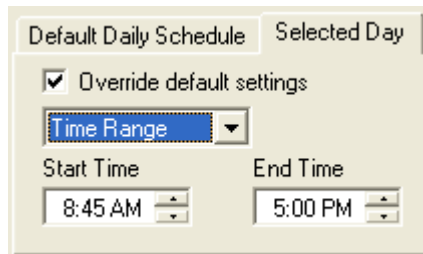
The image displays two screenshots of the 'Default Daily Schedule' configuration window. The top screenshot shows the 'Time Range' option selected in the drop-down menu, with 'Start Time' set to 8:00 AM and 'End Time' set to 7:59 AM. The bottom screenshot shows the 'Specific Time' option selected in the drop-down menu, with 'Time' set to 8:00 AM.

The drop-down select list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time that you want your job or format to process. If you select **Specific Time**, a **Time** box is displayed. Select the time you want the job or format to be processed.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Selected Day

The **Selected Day** tab allows you to specify settings for the selected day that differ from the settings specified in the **Default Daily Schedule** tab. In order for the **Selected Day** tab to be enabled, you must click a day to select it and you must select the **Override default settings** check box.



The drop-down select list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time that you want your job or format to process. If you select **Specific Time**, a **Time** box is displayed. Select the time when you want the job or format to be processed.

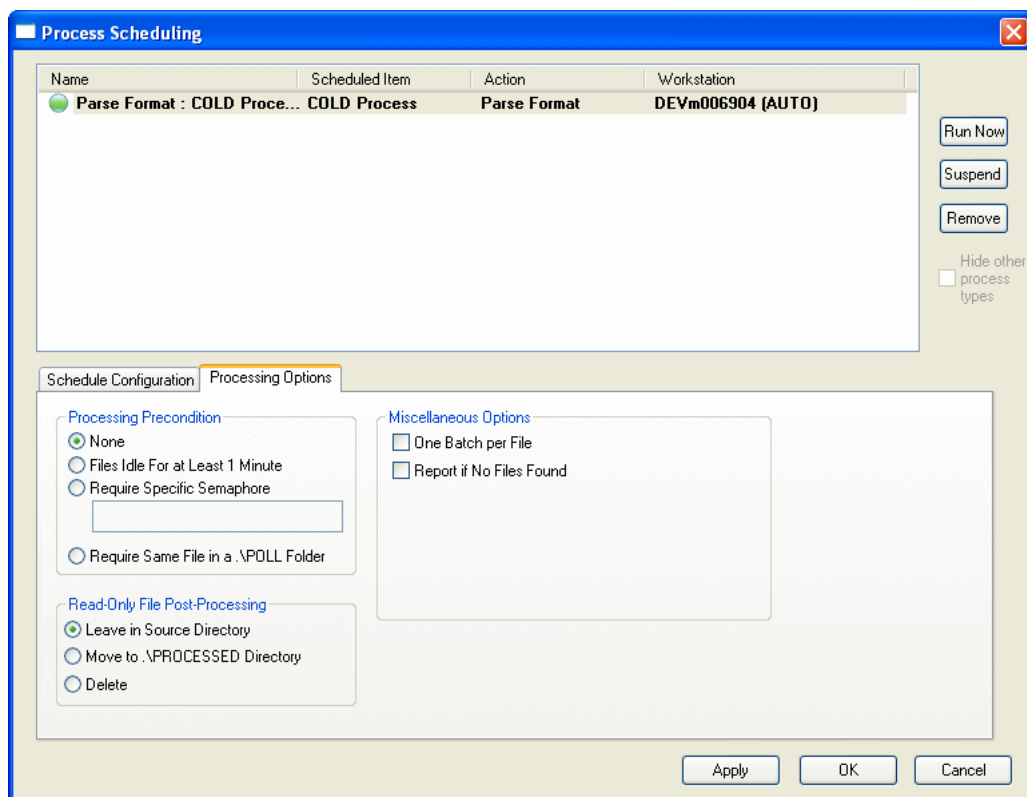
Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

A

PROCESSING OPTIONS

After the Schedule Options are configured on the Schedule Configuration tab, you must configure the Processing Options.

1. From the **Process Scheduling** window, click the **Processing Options** tab to display the Processing Options.



If you are scheduling a Full-Page OCR or an Automated Indexing process, the following options are displayed on the **Processing Options** tab.

Schedule Configuration Processing Options

OCR Options

- ☒ Full-Text OCR
- ☐ Automated Index
- ☐ Process Ad Hoc OCR Documents

2. Set the following Processing Options.

Option	Description
Processing Precondition	<p>The Processing Precondition options allow you to specify the conditions that must be met before processing can begin.</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Automated Indexing processes, Full-Text OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • None. If this option is selected, no processing precondition is necessary. • Files Idle For at Least 1 Minute. Select to indicate that processing must begin after the file indicated in the Default File Name of the processing format has been idle for at least one minute. • Require Specific Semaphore. Select to indicate that processing must begin after a trigger file is detected. The trigger file can be any file type/size/label and can be written to any location on the network. OnBase will only begin processing the processing file indicated in the Default File Name of the process format after the trigger file has been detected. How processing is triggered (definition of the file location and/or time variable) is defined by a semaphore. A semaphore is a technique for coordinating or synchronizing polling activity. A maximum of 255 characters can be entered in this field. • Require Same File in a .\POLL Folder. Select to indicate that processing must begin after a POLL file has been written to a specifically-configured POLL folder. The POLL file must appear in a folder labeled POLL, and the POLL folder must be created as a subfolder of the Default Directory of the process format. The name of the POLL file must be exactly identical to the name of the file to be processed. The value in the Default File Name field will be used to locate the POLL file. When OnBase locates the POLL file, the processor will attempt to process any file with that same name in the Default Directory. For example: The Default File Name is *.txt, and the Default Directory is C:\ProcessFiles. The file to be processed is stored in this directory. For this example, the file is named pf11x74.txt. The POLL file should be placed in C:\ProcessFiles\POLL, and named exactly the same as the process file (pf11x74.txt). OnBase will search C:\ProcessFiles\POLL for a file that matches the Default File Name of *.txt. Upon finding the pf11x74.txt file, the processor will return to the C:\ProcessFiles directory and search for the file named pf11x74.txt. This is the file that will be processed.

Option	Description
Read-Only File Post-Processing	<p>The Read-Only File Post-Processing options allow you to specify how files marked as read-only are processed.</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Automated Indexing processes, Full-Text OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • Leave in Source Directory. Select to leave processed read-only files in the folder they originated in. • Move to .\PROCESSED Directory. Select to move all processed files, regardless of read-only status, to an OnBase-generated PROCESSED folder located in the same folder the read-only files were originally in. • Delete. Select to delete the read-only files from the folder they originated in. <hr/> <p>Note: The Delete option is not available for Scheduled Sweeps or Scan from Disk processes.</p> <hr/>
Miscellaneous Options	<p>The Miscellaneous Options options allow you to specify special scheduling options. Not all options are available for all processes.</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Automated Indexing processes, Full-Page OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • One Batch per File. Select to process each file as one batch when multiple files are being processed at once. • Report if No Files Found. Select to create a Verification Report if no files are found when a scheduled format or job is run. <hr/> <p>Note: The Report if No Files Found option is only available when the None radio button is selected for the Processing Precondition. It is not available for Scheduled Sweeps or Scheduled Scan from Disk processing.</p> <hr/> <ul style="list-style-type: none"> • Document Type. Use the drop-down to select the Document Type of processed documents. <hr/> <p>Note: The Document Type drop-down is available only for scheduled sweep processes.</p> <hr/>

Option	Description
OCR Options	<p>The OCR Options options allow you to specify the configuration options for a scheduled Automated Indexing or Full-Text OCR process.</p> <hr/> <p>Note: These options are only available when scheduling an Automated Indexing or Full-Page OCR process.</p> <hr/> <ul style="list-style-type: none"> • Full-Text OCR. Select this radio button if you are scheduling a Full-Text OCR process. • Automated Indexing. Select this radio button if you are scheduling an Automated Indexing process. • Process Ad Hoc OCR Documents. Select this radio button if you would like to perform Automated Indexing or Full-Text OCR on documents in the ad-hoc batch status queues (i.e., Awaiting Ad Hoc Automated Index or Awaiting Ad Hoc OCR).

3. When you are finished configuring the Process Options, click **Apply**.

Viewing Scheduled Processes

By default, only scheduled process formats and jobs of the currently-selected process type will be displayed in the **Process Scheduling** window. To view scheduled process formats and jobs of all process types, deselect the **Hide other process types** check box.

To open the **Process Scheduling** window, perform one of the following actions:

- Click **Processing | Scheduler | Schedule Management**.
- Open the **Scheduled Processes** queue and double-click on a scheduled process
- Right-click on a process format in its process queue and select **Schedule Format**.

Note: Additional Product Rights are required to view a scheduled purge process. For more information, see the System Administration documentation.

Modifying a Scheduled Process Format

Once a scheduled process has been created, it can be modified as needed.

To modify an existing scheduled process:

1. Open the **Process Scheduling** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select the process to be modified from the **Scheduled Items** box.

3. Modify the settings on the **Schedule Configuration** and **Process Options** tabs as needed.
For more information on the options on these tabs, see *Schedule Configuration* on page 251 and *Processing Options* on page 256.
4. Once you have finished modifying the scheduled process, click **Apply**.

Deleting a Scheduled Process Format

Caution: If you delete a process format or process job that is scheduled, it will be deleted from the list of scheduled jobs.

Scheduled processes can be deleted from the **Process Scheduling** window.


1. Open the **Process Scheduling** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select the scheduled process you would like to delete from the **Scheduled Items** box and click **Remove**.
3. Click **Apply**.

Running/Suspending a Scheduled Process Format




From the **Process Scheduling** window, a scheduled process can be run immediately or it can be suspended.

1. Open the **Process Scheduling** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select a scheduled process from the **Scheduled Items** box.
 - To run the process now, click **Run Now**. The process is run the next time the processing workstation is polled.
 - To suspend the process, click **Suspend**. To resume a suspended process, click **Resume**.

An icon is displayed next to each scheduled process in the **Scheduled Items** box that indicates its status.

Icon	Description
	Run Now - Indicates that the user has clicked the Run Now button to cause the process to execute now instead of waiting for its scheduled time to run.

A

Icon	Description
	Suspend - Indicates a suspended process. The process will not run until a user selects it and clicks Resume .
	Active - Indicates an active scheduled process. An active process may be waiting to run or it may have already run at its scheduled time.
	Error - Indicates a process with a configuration error.

3. Click **Apply**.

Working With Process Jobs

A Process Job is one or more Process Formats that have been configured to run sequentially. A Process Job does not have to consist exclusively of a single type of Process Format; it can contain multiple Process Formats from any module that allows scheduling.

A few notes about Process Jobs:

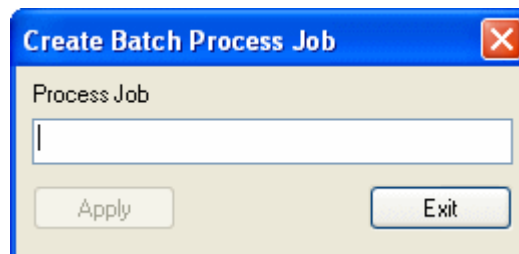
- Process formats must be created before they can be added to a job.
- AutoFill Keyword Import Processors can be scheduled from any **Process Job Queue**.
- Process Formats created from Document Imaging sweep or scan from disk processes cannot be included in a Process Job.

Creating a Job

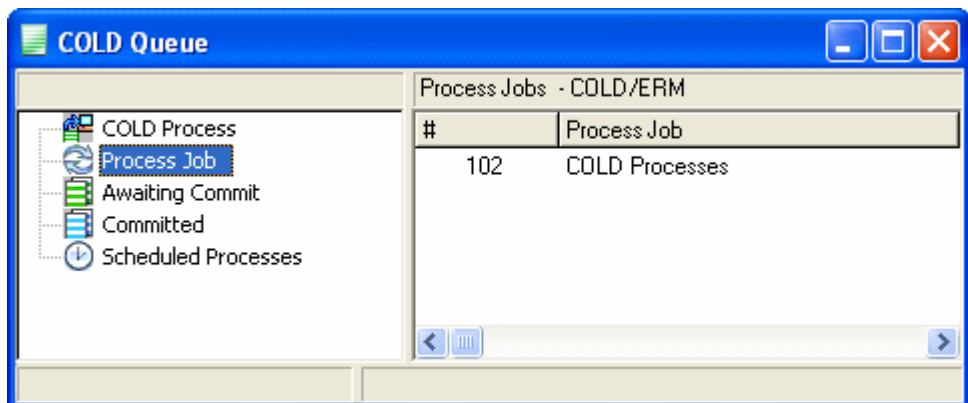
You can add a job to the Scheduler from a process queue (i.e., the COLD Queue, the EDI Queue, etc.). To create a job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on the window and select **Create New Job**.

Or, from the process queue, select **Process Job** and right-click in the **Process Jobs** window and select **Create New Job**. The **Create Batch Process Job** dialog box is displayed.



2. Enter a name for the job in the **Process Job** field and click **Apply**. The job is added to the process queue and is listed in the **Process Jobs** window.



Note: The process name must be 75 characters or fewer.

Configuring a Job

To configure a job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Configure Job**.

Or, select the job to be configured from the **Process Jobs** window in the process queue, right-click and select **Configure Job**.

The **Process Job Configuration** window is displayed.

Process Job Configuration

Process Format	Date	File Name	Alternate Path	Alternate Disk Group
----------------	------	-----------	----------------	----------------------

Up
Down

Process Format
Alternate Path
Alternate File Name
Alternate Disk Group
Language Conversion
< NO CONVERSION >
Processing Options
Store Document Indices
Store Document Data Files
Store Import File
Create Auto-Folder
Default Keywords
Date
Today
Yesterday
Last Week Day
Last Business Day
Specific Date
Disallow job date override
Add
Update
Delete
Clear
Close

2. Configure a process format to add to the job:

A

Process Job Parameter	Description
Process Format	Select the process format to be incorporated in the process job. All available process formats are listed.
Alternate Path	Enter an alternate path to the data to be processed (i.e., the Default Directory) to use instead of the Default Directory configured for the selected process format. If an alternate path is not specified, the process format's Default Directory is used.
Alternate Filename	Enter an alternate file name for the data to be processed (i.e., the Default File Name) to use instead of the Default File Name configured for the selected process format. If an alternate file name is not specified, the process format's Default File Name is used.
Alternate Disk Group	Enter an alternate Disk Group to store the data being processed instead of the Disk Group configured for the selected process format. If an alternate Disk Group is not specified, the process format's default Disk Group is used.
Language Conversion	Used to specify a language conversion if the source data file was created using a different ASCII code page. If a language conversion is not specified, the process format's Language Conversion setting is respected.
Store Document Indices	Select this option to store the processed documents in the database, along with their Keyword Values and document name. This option is enabled by default.
Store Document Data Files	Select this option to move the data file to the configured Disk Group after the process is complete. This option is enabled by default.
Store Import File	Select to store a copy of the index file used to import documents into OnBase for archive purposes.
Create Auto Folder	Select to provide the ability to Auto-Folder documents upon processing. See the Folders documentation for additional information regarding Auto-Foldering. Note: Not all processors offer the ability to Auto-Folder documents upon processing. See the configuration section of the Automated Indexing documentation for more information.

Process Job Parameter	Description
Default Keywords	<p>Click the Default Keywords button to select Keyword Types and Values that are displayed in the Batch Name for that Process Job when it is processed.</p> <p>These Keyword Types and Values are also displayed at the top of the Verification Report for that job.</p> <hr/> <p>Note: Only Keyword Types that have been configured for Document Types used in the Process Job are selectable.</p> <hr/> <p>Note: If a check process format is configured as part of the job, the Default Keywords button is disabled when the job is selected.</p> <hr/>
Disallow job date override	Select this check box to prevent users from overriding the specified job date.
Date	These settings allow a user-defined Document Date to be stored for the processed documents. This date is used as the %D parameter that appears in the document's Auto-Name string.

- Click **Add**.
- Repeat Step 2 for each process format that you would like to add to the job.

Process jobs are run in the order in which they display on the screen. Re-sequence a job by selecting it and clicking the **Up** or **Down** buttons.
- Once you've added all process formats to the job, click **Close**.

Scheduling a Job

Once you have created and configured a job, you must schedule it in order for it to automatically run. A job is scheduled in almost the same way that a process format is scheduled.

To schedule a job, you must first open the **Process Scheduling** window. To open it:

- From a process queue, select **Process Job** and then select the job to be scheduled in the **Process Jobs** window. Right-click and select **Schedule Job**.
- From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Schedule Job**.

SCHEDULE CONFIGURATION

The first options that must be configured for the scheduled job are the Schedule Configuration options on the **Schedule Configuration** tab. This tab is displayed by default.

1. In the **Name** field, enter a name for the scheduled process.
2. Using the **Processing Workstation** drop-down, select the workstation that will be used to run the scheduled job.

Note: This workstation will need to be running with the **-SCHED** or **-SCHEDINST** command line switch in order to run the scheduled job.

3. Using the **Schedule Template** drop-down, select a schedule template for the process or select **<Custom Schedule>** to manually configure the schedule for this process.

Note: For information on creating a schedule template, see below.

To create a custom schedule, you will need to use the **Calendar** to select the day(s) you would like the scheduled job to run on and then you will need to specify the time the scheduled job will run using the **Default Daily Schedule** and/or **Selected Day** tabs. For more information, see those sections below.

4. Select how often you would like the scheduled job to run by selecting one of the **Processing Frequency** radio buttons.
 - **Once then Suspend.** The scheduled item will be processed once, then the scheduled process is suspended.
 - **Once per Day.** The scheduled item be processed once per day.

Note: If the scheduled item is modified, the process may be run again on the same day.

- **Once every "" Minutes.** The scheduled item is processed in the interval (measured in minutes) entered in the field. The maximum number of minutes that can be entered is 99999.

5. When you are finished setting the Schedule Configuration options, click **Apply**.

Calendar

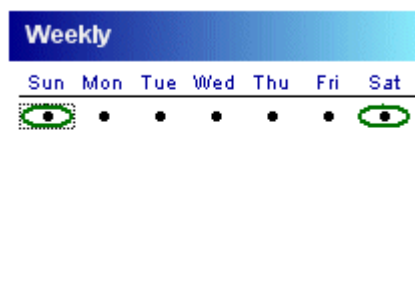
The calendar is used to select the day(s) on which a scheduled job should be run.

Note: The calendar is displayed based on your Workstation Regional Settings and the OnBase language DLL that you are using.

To change the view of the calendar, click the calendar heading (in the example above, **Weekly**) to display a menu. Select one of the following options to display a different calendar for configuration:

- **Weekly.** Allows you to configure a job to run on a certain day of the week (i.e., Thursday).
- **Monthly.** Allows you to configure a job to run monthly, on a particular date (i.e., the 1st and 15th of the month).
- **Monthly (Day-Relative).** Allows you to configure a job to run on a relative day of the month (i.e., the first Saturday of the month, the 2nd Wednesday of the month).
- **Annual.** Allows you to configure a job to run on a certain day of the year (i.e., June 30).
- **Full Calendar.** Allows you to configure a process to run on specified days of specified years (e.g., August 10, 2011 and/or July 17, 2012).

To select days that you would like to run a scheduled job, double-click the day on the calendar. The selected day is circled.



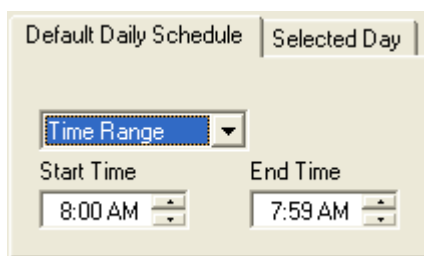
Once a day is selected, right-click on the calendar to display the currently-selected date in detail.

Note: In the example above, two days are selected but **Sunday** is the currently-selected day.

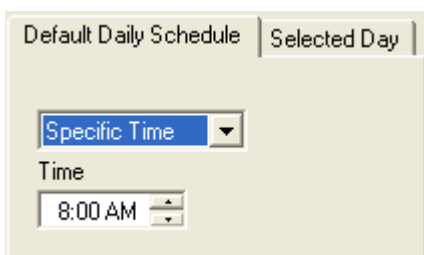
To deselect a day, double-click it.

Default Daily Schedule

The **Default Daily Schedule** tab allows you to configure the processing configuration for all days that do not have a **Selected Day** tab configuration.



The screenshot shows the 'Default Daily Schedule' tab selected. A dropdown menu is set to 'Time Range'. Below it, there are two time selection boxes: 'Start Time' set to '8:00 AM' and 'End Time' set to '7:59 AM'.



The screenshot shows the 'Default Daily Schedule' tab selected. A dropdown menu is set to 'Specific Time'. Below it, there is a single time selection box labeled 'Time' set to '8:00 AM'.

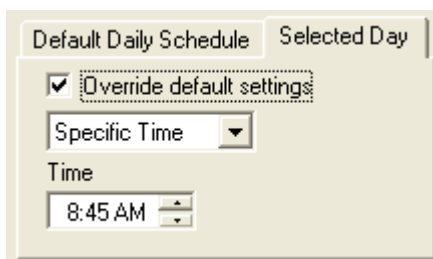
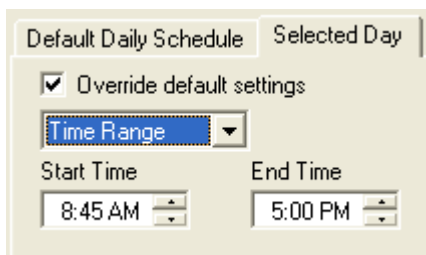
The drop-down select list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time that you want your job or format to process. If you select **Specific Time**, a **Time** box is displayed. Select the time you want the job or format to be processed.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Selected Day

A

The **Selected Day** tab allows you to specify settings for the selected day that differ from the settings specified in the **Default Daily Schedule** tab. In order for the **Selected Day** tab to be enabled, you must click a day to select it and you must select the **Override default settings** check box.



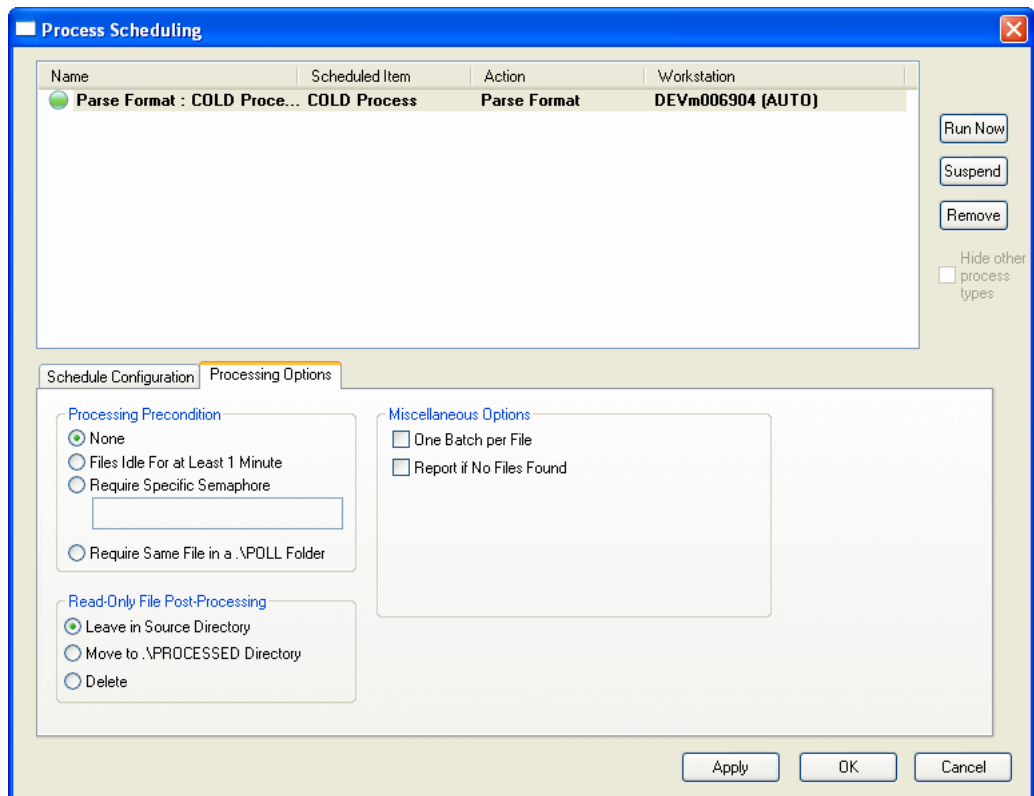
The drop-down select list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time that you want your job or format to process. If you select **Specific Time**, a **Time** box is displayed. Select the time when you want the job or format to be processed.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

PROCESSING OPTIONS

After the Schedule Options are configured on the **Schedule Configuration** tab, you must configure the Processing Options.

1. From the **Process Scheduling** window, click the **Processing Options** tab to display the Processing Options.



2. Set the following Processing Options.

Option	Description
Processing Precondition	<p>The Processing Precondition options allow you to specify the conditions that must be met before processing can begin.</p> <hr/> <p>Note: These options are not available for scheduled PDF conversions, Automated Indexing processes, Full-Text OCR processes or scheduled commits.</p> <hr/> <ul style="list-style-type: none"> • None. If this option is selected, no processing precondition is necessary. • Files Idle For at Least 1 Minute. Select to indicate that processing must begin after the file indicated in the Default File Name of the processing format has been idle for at least one minute. • Require Specific Semaphore. Select to indicate that processing must begin after a trigger file is detected. The trigger file can be any file type/size/label and can be written to any location on the network. OnBase will only begin processing the processing file indicated in the Default File Name of the process format after the trigger file has been detected. <p>How processing is triggered (definition of the file location and/or time variable) is defined by a semaphore. A semaphore is a technique for coordinating or synchronizing polling activity. A maximum of 255 characters can be entered in this field.</p> <ul style="list-style-type: none"> • Require Same File in a .\POLL Folder. Select to indicate that processing must begin after a POLL file has been written to a specifically-configured POLL folder. <p>The POLL file must appear in a folder labeled POLL, and the POLL folder must be created as a subfolder of the Default Directory of the process format. The name of the POLL file must be exactly identical to the name of the file to be processed.</p> <p>The value in the Default File Name field will be used to locate the POLL file. When OnBase locates the POLL file, the processor will attempt to process any file with that same name in the Default Directory.</p> <p>For example: The Default File Name is *.txt, and the Default Directory is C:\ProcessFiles. The file to be processed is stored in this directory. For this example, the file is named pf11x74.txt.</p> <p>The POLL file should be placed in C:\ProcessFiles\POLL, and named exactly the same as the process file (pf11x74.txt).</p> <p>OnBase will search C:\ProcessFiles\POLL for a file that matches the Default File Name of *.txt. Upon finding the pf11x74.txt file, the processor will return to the C:\ProcessFiles directory and search for the file named pf11x74.txt. This is the file that will be processed.</p>

Option	Description
Read-Only File Post-Processing	<p>The Read-Only File Post-Processing options allow you to specify how files marked as read-only are processed.</p> <ul style="list-style-type: none"> • Leave in Source Directory. Select to leave processed read-only files in the folder they originated in. • Move to .\PROCESSED Directory. Select to move all processed files, regardless of read-only status, to an OnBase-generated PROCESSED folder located in the same folder the read-only files were originally in. • Delete. Select to delete the read-only files from the folder they originated in.
Miscellaneous Options	<p>The Miscellaneous Options options allow you to specify special scheduling options. Not all options are available for all processes.</p> <ul style="list-style-type: none"> • One Batch per File. Select to process each file as one batch when multiple files are being processed at once. • Report if No Files Found. Select to create a Verification Report if no files are found when a scheduled job is run.

3. When you are finished configuring the Process Options, click **Apply**.

Viewing a Job

All scheduled process formats and jobs can be viewed in the **Process Scheduling** window.

By default, the **Hide other process types** check box is enabled, so only the selected process type's process formats or process jobs are displayed.

To open the **Process Scheduling** window:

- Click **Processing | Scheduler | Schedule Management** from the OnBase Client.
- From a process queue, select **Process Job** and then select a job in the **Process Jobs** window. Double-click on the job to display the process formats that compose it.
- From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed.

Modifying a Job

To modify an existing job:

From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Configure Job**.

Or, select the job to be modified from the **Process Jobs** window in the process queue, right-click and select **Configure Job**.

The **Process Job Configuration** dialog is displayed.

Note: For more information on configuring a process job, see [Configuring a Job](#) on page 264 and [Scheduling a Job](#) on page 266.

RENAMING A JOB

To rename an existing job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Rename Job**.

Or, select the job to be modified from the **Process Jobs** window in the process queue, right-click and select **Rename Job**.

The **Rename Process Job** dialog box is displayed.

2. Enter the new name for the job and click **OK**.

Deleting a Job

Caution: If you delete a process format or process job that is scheduled, it will be deleted from the list of scheduled jobs.

To delete an existing job:

1. From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Delete Job**.

Or, select the job to be modified from the **Process Jobs** window in the process queue, right-click and select **Delete Job**.

A confirmation message is displayed.





2. Click **OK**. The job is deleted.

Running/Suspending a Job

From the **Process Scheduling** window, a job can be run immediately or it can be suspended

1. Open the **Process Scheduling** window from the OnBase Client by clicking **Processing | Scheduler | Schedule Management**.
2. Select a job from the **Scheduled Items** box.
 - To run the job now, click **Run Now**. The job is run the next time the processing workstation is polled.
 - To suspend the job, click **Suspend**. To resume a suspended job, click **Resume**.

An icon is displayed next to each scheduled job in the **Scheduled Items** box that indicates its status.

Icon	Description
	Run Now - Indicates that the user has clicked the Run Now button to cause the job to execute now instead of waiting for its scheduled time to run.
	Suspend - Indicates a suspended job. The job will not run until a user selects it and clicks Resume .
	Active - Indicates an active scheduled job. An active job may be waiting to run or it may have already run at its scheduled time.
	Error - Indicates a job with a configuration error.

3. Click **Apply**.

A job can also be run immediately from the process format queue or the **Process Jobs** window.

From the OnBase Client, click **Processing | Process Jobs**. The **Process Jobs** window is displayed. Right-click on a job and select **Process Job**.

Or, from a process queue, select **Process Job** and then select the job to be run in the **Process Jobs** window. Right-click in the **Process Jobs** window and select **Process Job**.

Creating Schedule Templates

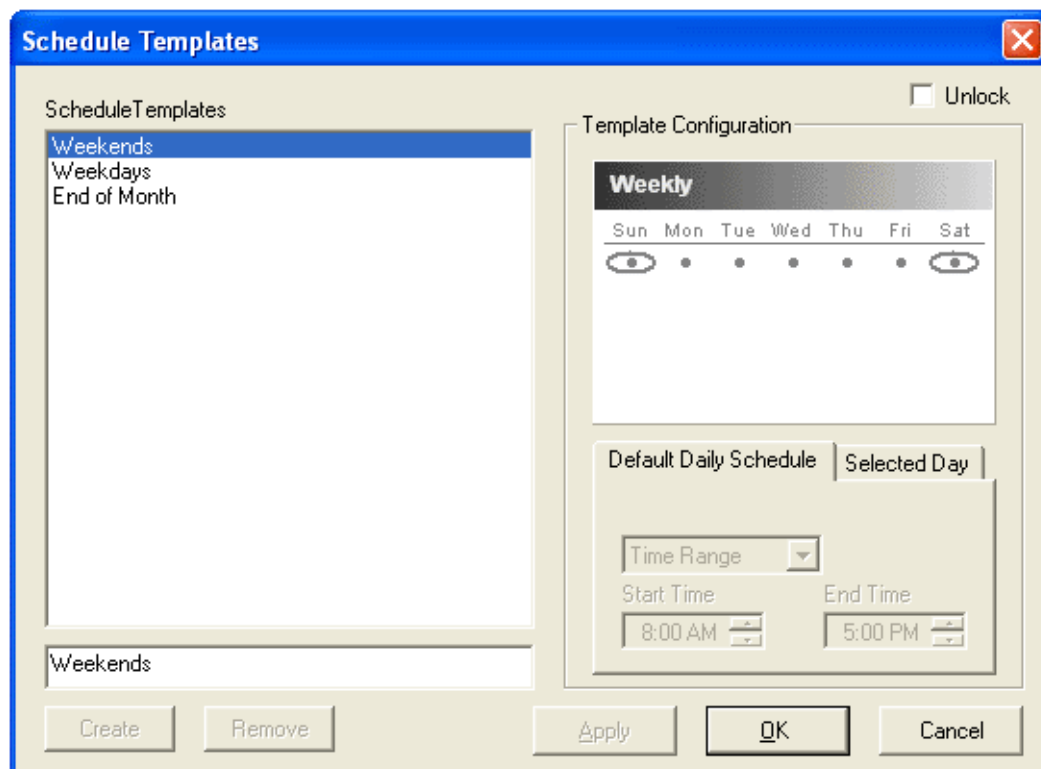
CREATING SCHEDULE TEMPLATES

A schedule template is used to create a processing schedule. These schedules can be used by multiple scheduled processes without having to be re-configured each time they are used.

Note: Any user with the Client and Scheduler product rights can create a schedule template. Once created, a schedule template is available to all users with Client and Scheduler product rights.

To create a schedule template:

1. From the OnBase Client, click **Processing | Scheduler | Schedule Templates**. The **Schedule Templates** window is displayed.



2. Enter a name for the new template and click **Create**.

Note: The maximum number of characters that can be used for a name is 80.

3. Configure the appropriate options. See the sub-sections below for more information on using the calendar, **Default Daily Schedule**, and **Selected Day** options under the **Template Configuration** area.
4. Once all Template Configuration options have been set, click **OK**.

To edit an existing template, select it from **Schedule Templates** list and select the **Unlock** check box. Once you have finished modifying it, click **OK**.

Calendar

The calendar is used to select the day(s) on which a scheduled process should be run.

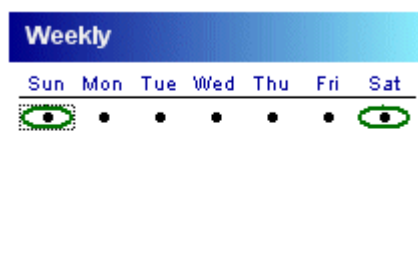
Note: The calendar is displayed based on your Workstation Regional Settings and the OnBase language DLL that you are using.

To change the view of the calendar, click the calendar heading (in the example above, **Weekly**) to display a menu. Select one of the following options to display a different calendar for configuration:

- **Weekly.** Allows you to configure a process to run on a certain day of the week (i.e., Thursday).
- **Monthly.** Allows you to configure a process to run monthly, on a particular date (i.e., the 1st and 15th of the month).
- **Monthly (Day-Relative).** Allows you to configure a process to run on a relative day of the month (i.e., the first Saturday of the month, the 2nd Wednesday of the month).
- **Annual.** Allows you to configure a process to run on a certain day of the year (i.e., June 30).
- **Full Calendar.** Allows you to configure a process to run on specified days of specified years (e.g., August 10, 2011 and/or July 17, 2012).

A

To select days that you would like to run a scheduled process, double-click the day on the calendar. The selected day is circled.

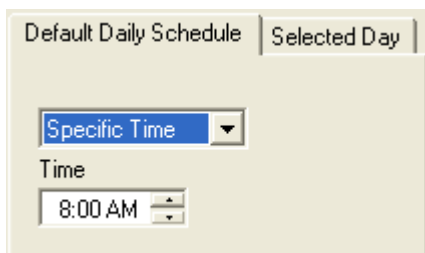
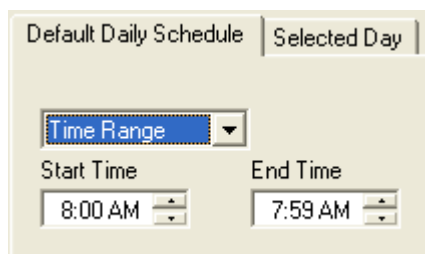


Note: In the example above, two days are selected but **Sunday** is the currently-selected day.

To deselect a day, double-click it.

Default Daily Schedule

The **Default Daily Schedule** tab allows you to configure the processing configuration for all days that do not have a **Selected Day** tab configuration.

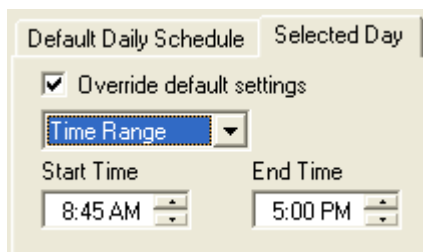


The drop-down select list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time that you want your job or format to process. If you select **Specific Time**, a **Time** box is displayed. Select the time you want the job or format to be processed.

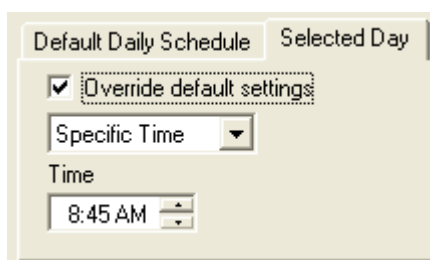
Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

Selected Day

The **Selected Day** tab allows you to specify settings for the selected day that differ from the settings specified in the **Default Daily Schedule** tab. In order for the **Selected Day** tab to be enabled, you must click a day to select it and you must select the **Override default settings** check box.



The screenshot shows the 'Default Daily Schedule' dialog with the 'Selected Day' tab active. The 'Override default settings' checkbox is checked. The drop-down menu is set to 'Time Range'. Below it, the 'Start Time' is set to 8:45 AM and the 'End Time' is set to 5:00 PM.



The screenshot shows the 'Default Daily Schedule' dialog with the 'Selected Day' tab active. The 'Override default settings' checkbox is checked. The drop-down menu is set to 'Specific Time'. Below it, the 'Time' is set to 8:45 AM.

The drop-down select list allows you to select **Time Range** or **Specific Time**. If you select **Time Range**, a **Start Time** box and an **End Time** box are displayed. Define the range of time that you want your job or format to process. If you select **Specific Time**, a **Time** box is displayed. Select the time when you want the job or format to be processed.

Tip: Specifying a **Time Range** and using the **Once Per Day** option will allow a scheduled process to run even if another process runs over its starting time, as long as the process is able to start within the specified range.

AUTOMATED INDEXING BEST PRACTICES

A

The following best practice recommendations have been assembled by a team of Automated Indexing subject matter experts. They represent the accumulation of years of experience installing and configuring scanning and automated indexing solutions.

The following recommendations are general in nature, and are applicable to most Automated Indexing solutions and network environments. Depending on your solution design and your organization's needs, not all of the best practice recommendations listed below may apply to, or be recommended for, your Automated Indexing solution.

Carefully consider the impact of making any changes, including those listed below, to your Automated Indexing solution prior to implementing them in a production environment.

GENERAL INFORMATION

Recommended Standards for OCR Processing

Because OCR processing is an integral part of the Automated Indexing process, it is considered a best practice to ensure that all documents that undergo Automated Indexing meet the recommended criteria for OCR processing. This will ensure that the OCR engine returns the most accurate results possible.

- It is considered a best practice to scan all documents as bi-tonal (i.e., black and white) images. Bi-tonal images require far less disk space, are easier to OCR process, and load faster than grayscale or color images.
- It is considered a best practice to store bi-tonal images using the TIFF-Group IV file format. Grayscale or color images should be saved using the JPEG file format.
- It is considered a best practice to scan documents at a resolution of 240 dpi. The resolution should always be set to a squared value (e.g., 100x100, 200x200, 300x300 dpi).

If poor or inaccurate OCR processing results are reported at 240 dpi, the resolution of the documents should be increased incrementally until acceptable results are achieved.

A

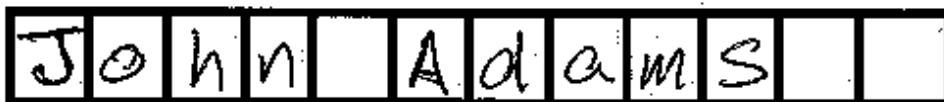
Recommended Standards for ICR Processing

If your solution is licensed for intelligent character recognition (ICR), you can use Automated Indexing to recognize and capture data from hand-written text.

It is difficult to predict the accuracy of an ICR process due to the variability in the quality of the hand-written text. The following tips will help to improve the accuracy of ICR processing.

- The hand-written characters should be clearly separate from one another (i.e., they should not touch or intersect).
- The hand-written text should be unconstrained. If the characters are being entered into constraint boxes, you should configure your OCR process format to automatically drop them by selecting the **Remove vertical/horizontal lines from zone before processing** check box.

Constrained Text:



Name

Unconstrained Text:

A single line of handwritten text in a cursive script that reads 'John Adams'. The letters are fluid and connected, with some overlapping.

Name

- The ICR numeric text processor (**Handwriting/ICR - numerals/grouping punctuation (North American style)**, **Handwriting/ICR - numerals/grouping punctuation (European style)**) is significantly more accurate than the ICR alphanumeric text processor (**Handwriting/ICR - alphanumeric/punctuation**) because the character set is much smaller.

If you are attempting to process numeric data only from a Data Field Zone, you will have more accurate results if you use the numeric text processor.

A

USAGE

It is recommended that the following best practice information be considered before scanning and processing documents in the Automated Indexing client.

Automated Indexing

Ad-Hoc Automated Indexing

Unlike batch Automated Indexing, ad-hoc Automated Indexing offers no indication of values that were not confidently identified by the Automated Indexing engine (i.e., values that did not meet their configured Suspect Level). It is considered a best practice to always compare the indexing information assigned to the document by the ad-hoc Automated Indexing process to the actual information displayed on the document to ensure that it is accurate.

Batch Automated Indexing

When scanning a batch of documents into a scan queue, it is considered a best practice to scan the batch as **Unindexed** or **Pre-Indexed** so that any documents that undergo batch Automated Indexing and are identified as having suspect Keyword Values are sent for user review in the **Index in Progress** batch status queue. If the batch is scanned as **Fully Indexed**, batches containing documents that are associated with suspect Keyword Values are not routed for manual review in the **Index in Progress** batch status queue because the batch is already marked as indexed.

Performing Point and Shoot Indexing

It is considered a best practice to always compare the values returned from the OCR engine to the actual values displayed on the document when performing Point and Shoot Indexing.

CONFIGURATION

A

It is recommended that the following best practice information be considered before configuring your Automated Indexing solution.

Automated Indexing Form Configuration

The following best practice information should be considered when creating or modifying an Automated Indexing form.

- By default, the Suspect Level threshold is set to **75** and the average score given to a processed field is 70. It is considered a best practice to set your Suspect Level to the default threshold of 75 to ensure that forms are correctly and consistently being matched to documents
- It is considered a best practice to save your Automated Indexing form after the form is created and after any configuration change is made (e.g., any Form Identification, Registration, or Data Field Zones are added, deleted, modified, moved, etc.).
- Because a limitless number of Automated Indexing forms can be configured for a single Document Type, let alone all Document Types configured for your Automated Indexing solution, it is considered a best practice to limit the number of Automated Indexing forms available to a Document Type as much as possible in order to save time and processing resources.
- It is considered a best practice to test each Automated Indexing form after it is created and after any modifications have been made to it. All Automated Indexing forms should also be regularly tested to ensure that they do not require updating.
- It is considered a best practice to periodically perform a form order analysis and re-order your forms. The frequency of this analysis should be based on the number of documents you are processing via Automated Indexing and the number of forms configured for your solution.