DocuShare OCR 2.0
User Guide

xerox®

Changes are periodically made to this document. Changes, technical inaccuracies, and typographic errors will be corrected in subsequent editions.

This document supports DocuShare Release 6.6.1 and higher.
Publication date: January 2014

Prepared by:
Xerox Corporation
Xerox Content Management
3333 Coyote Hill Road
Palo Alto, California 94304 USA

Contents

1 Using DocuShare OCR

1

Using DocuShare OCR

Overview

DocuShare OCR (optical character recognition) provides the ability to convert scanned documents and image files to a variety of popular document formats. After the documents are converted, you can then search for them, provided content indexing is enabled on your site. The add-on integrates an OCR converter within the DocuShare content rules feature, enabling you to create a content rule that converts documents scanned or uploaded to DocuShare.

Converting scanned documents

Using DocuShare OCR, you can convert scanned PDF and image documents to a variety of commonly used formats. To have DocuShare perform the conversion, you create a content rule for a collection. During the conversion, DocuShare creates another rendition of the document; you specify the rendition to create in the content rule.

To create a content rule that performs OCR conversion:

1. Locate a collection in which documents will be scanned.
2. Do one of the following:
   - Click the collection's **Content Rules** icon.
   - Click the collection's **Properties** icon. Then click the **Content Rules** link.
3. On the Content Rules page, click **Create a New Content Rule**.

   The Create a New Content Rule wizard appears. The following steps describe the required fields. For information about any of the optional fields in the wizard, click the appropriate field name.
4. On the Description page, enter a **Title** for the content rule and click **Next**.
5. On the Event Triggers page, select **Something added** and **Document**. Then click **Next**.
6. On the Content Property Conditions page, you can refine the event trigger by specifying property conditions. Then click **Next**.
7. On the Action Performed page, select **Convert** and click **Next**.
8. On the Action Settings page, do the following:
   a. Select **OCRConverter**.
   b. Under **Rendition**, select the file format in which to save the converted document.
   c. Under **Rendition detail**, select a specific rendition type.
   d. Select the **Make selected rendition the document's preferred display format (allows document to be searched)** checkbox if you want the document's preferred display format to be the file format you selected for **Rendition**. With this checkbox selected, the document is indexed, enabling users to search for it.
   e. Click **Next**.
9. On the Completion Settings page, you can choose to set a property value on the document when the content rule completes. In addition, you can append the property value to any existing property values. Then click **Done**.

   The View Properties page for the content rule appears and provides a summary of the content rule. When a document is added to the collection, the conversion you specified is run.

Supported file formats

DocuShare OCR supports the following file formats.

Adobe Portable Document Format

- Text PDF
- Edited PDF
- Searchable Image PDF
- Substituted Image PDF

Microsoft Excel Spreadsheet

- Excel 2007
- Excel 2000
- Excel 97

Microsoft Office Word

- Word 2007
- Word 2000
- Word 97
- WordPad

Microsoft PowerPoint

- PowerPoint 2007
- PowerPoint 97

Microsoft Publisher

Rich Text Format

- WordML (Word 2003)
- RTF 2000 ExactWord
- RTF Word 2000
- RTF Word 97
- RTF Word 6.0

Text Format

- Text
- Text PDF
- Text – formatted
- Text – with line breaks
- Text – comma separated
- Unicode text
- Unicode text – formatted
- Unicode text – with line breaks

- Unicode text – comma separated

WordPerfect Document

- WordPerfect 10
- WordPerfect 8

XML Document

- Word XML
- XML

xerox®