2.4 Managing PDF Scanned Elements

Contents

Recognizing Scanned Text in Acrobat

The electronic version of a document being remediated for accessibility is not always available. In these cases, the **Optical Character Recognition** (or OCR) function built into Adobe Acrobat must be utilized to identify all text elements of the document before any other editing processes are carried out. Having already covered the automated text recognition process of the **Edit PDF** tool in Module 1.3, 'Arranging Text and Images in Acrobat', we will extend on working with scanned documents with the **Enhance Scans** tool.

Using the Enhance Scans Tool

Adobe Acrobat has been designed for a user to manage all scanned elements through a single tool, **Enhance Scans**. Find and open this tool on the **Tools** page to begin remediating a scanned document. If the **Edit PDF** tool has not yet been opened to automatically recognize text in Acrobat, then OCR must be started to identify scanned text and image elements.

To Manually Run Optical Character Recognition in Acrobat:

1. With the **Enhance Scans** tool open, find and select the 'Recognize Text' dropdown menu that is found in the top **Enhance Scans** toolbar.

2. Select 'In This File' to run OCR for the current file, or 'In Multiple Files' to run OCR once other files have been selected.

3. An additional toolbar will appear below the **Enhance Scans** toolbar. Ensure the 'Pages' and 'Language' dropdown menus are set to 'All Pages' and 'English (US)' (as a note, this second language setting differs from the previously covered metadata language setting; it is designed to preset the language for the spelling of text recognition and screen reader pronunciation).

4. From the additional toolbar, select the 'Recognize Text' button. Acrobat will automatically run a 'Converting scanned page' progress bar and convert the page.

Correcting Recognized Text

Adobe Acrobat offers a feature that finds suspected OCR errors for the user to correct. Before making other improvements to the newly recognized text, it is advisable to use this automated feature to type any text that the text recognition software was unable to identify or meaningfully translate.

To Correct Errors in Recognized Text:

1. With the **Enhance Scans** tool open, find and again select the 'Recognize Text' dropdown menu that is found in the top **Enhance Scans** toolbar.

2. Select the 'Correct Recognized Text' option.

3. A dialogue window may appear that states 'The document does not contain any suspects. If this window does not appear, character recognition errors were found within the document.

4. An additional toolbar will appear below the **Enhance Scans** toolbar, with an image of the first text or word suspected for correction. In the text window located to the right of 'recognized as', edit or confirm the correct text is typed and press the 'Enter' key or the 'Accept' button.

5. The toolbar will automatically que the next text suspected of a mistake. Repeat this process until no suspects remain within the document.

Additionally, toggling the 'Review recognized text' checkbox (located in the 'Recognize Text' toolbar) while correcting errors will allow the user to examine undetected text errors within the document. By proof reading through the text generated by the OCR software and fixing these common errors, a content creator can ensure these hard-to-detect mistakes will not resurface with screen readers or other assistive technology.

Improving Scanned Elements in Acrobat

The quality of scanned images and text can be improved using an 'Enhance' feature found in the top **Enhance Scans** toolbar. After opening this feature and selecting 'Scanned Document', an additional 'Enhance' toolbar will open. Here, several optimization and filtering options can be adjusted in the 'Settings' menu. Open 'Settings' to view and adjust these options in an 'Enhance Scanned PDF' window.

Small Size vs. High Quality Slider

This slider tool allows a user to optimize a document's file size to scan quality ratio. Shifting the slider in the direction of 'Small Size' will further compress the document to a smaller file size. Alternatively, shifting the slider toward 'High Quality' will improve the scan quality at the expense of file size.

Deskew Filter

Under the 'Filters' section, select 'Edit' to find the 'Deskew' option. When 'Deskew' is set to 'On', enhancing the document will correct the angle of any misaligned scanned text. This is useful for text scanned in at a slight angle or text deformed by book or packet bindings.

Background Removal Filter

Also found when selecting 'Edit' in the 'Filters' section of the 'Enhance Scanned PDF' window, the 'Background Removal' option is defaulted to 'Off' but has a 'Low', 'Medium' and 'High' setting. The user can choose to remove unwanted scanned document background patterns and imperfections to varying degrees with this option. Select 'OK' in the 'Filters' window to change the default.

Finalizing and Applying Enhance Settings

To save all adjusted settings, select 'OK' in the 'Enhance Scanned PDF' window. Select the 'Enhance' button located in the top toolbar to apply the enhanced scan settings to the PDF document.

Conclusion

In this module, we've walked through the processes of converting and improving scanned documents into readable PDF documents. The next module, 'Creating and Editing PDF Tags', will work with readable text documents by managing structure tags for accessibility.