



USAID
FROM THE AMERICAN PEOPLE

Document Processing 101

Making Your Document A Searchable Part of
USAID's Online Archive

Presented by the Knowledge Services Center's Development Experience Clearinghouse

August 2010



USAID
FROM THE AMERICAN PEOPLE

Presentation Overview

The Development Experience Clearinghouse (DEC) is the online archive of USAID's technical and programmatic materials.

ADS 540 requires USAID staff and contractors to submit technical and programmatic materials to the DEC for safekeeping, transparency, and easy access by all.

Optical Character Recognition (OCR), also known as "Paper Capture," is the process by which documents submitted to the DEC are made searchable.

This presentation provides a 12-step orientation to the OCR process.

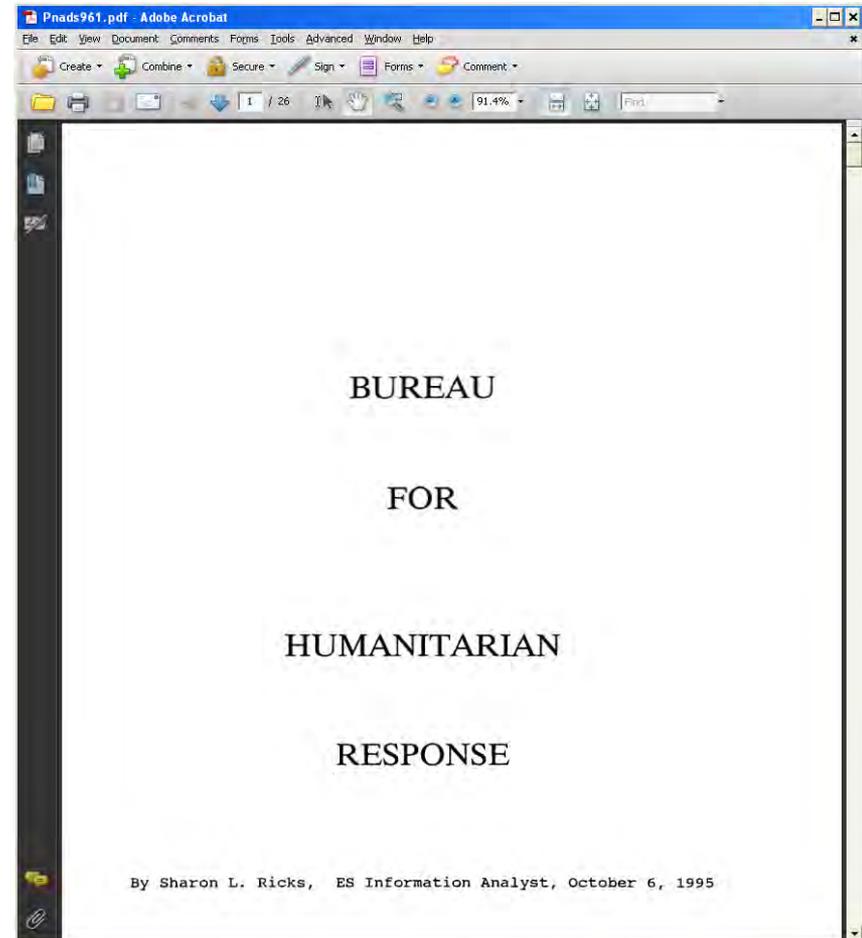


USAID
FROM THE AMERICAN PEOPLE

Step 1

- Open the PDF file in Adobe Acrobat 8 or later.

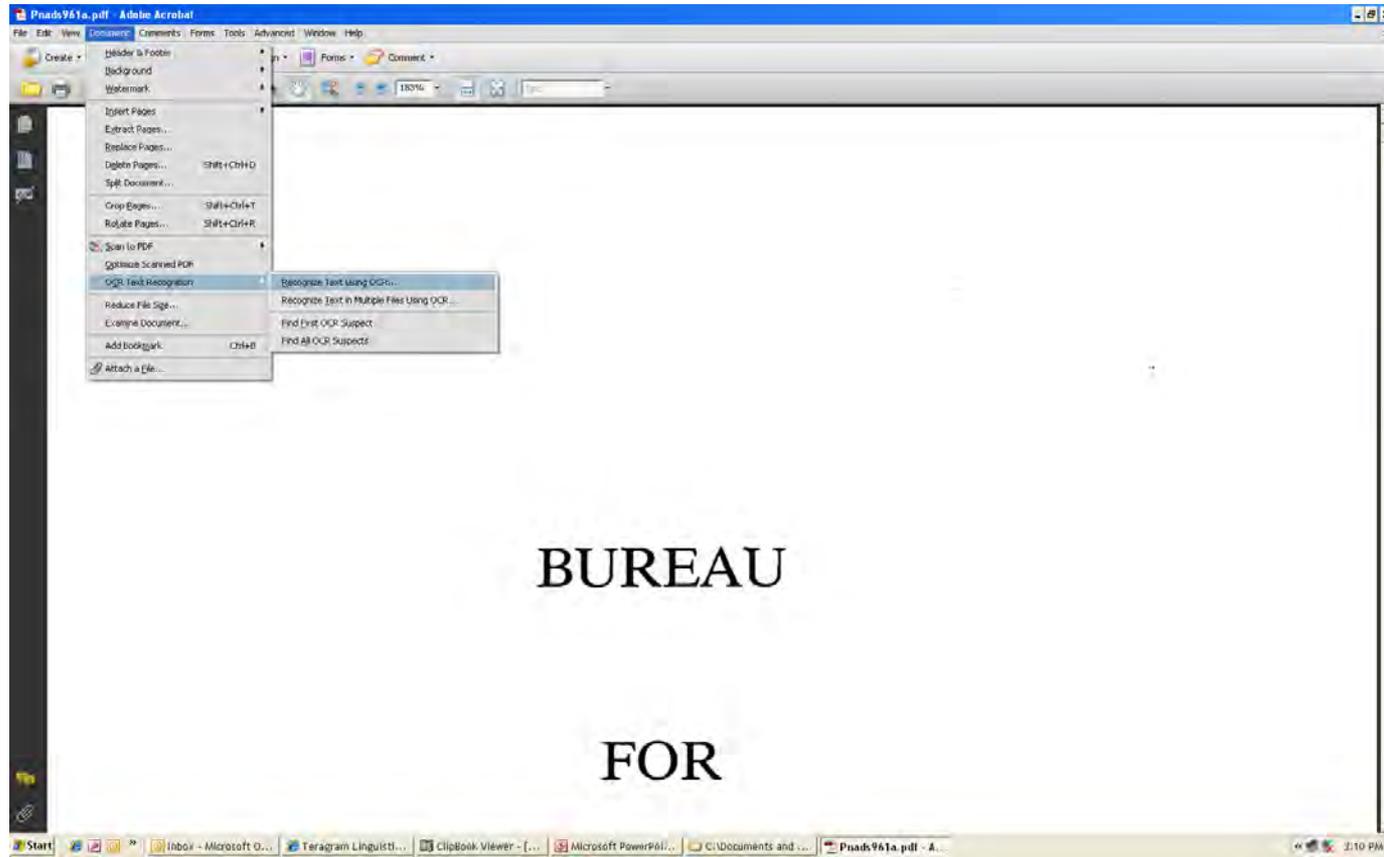
Note: The OCR should work regardless of the version of Adobe Acrobat Pro or Standard that you are working in.





Step 2

- From the **Document** dropdown menu, select **OCR Text Recognition > Recognize Text Using OCR.**





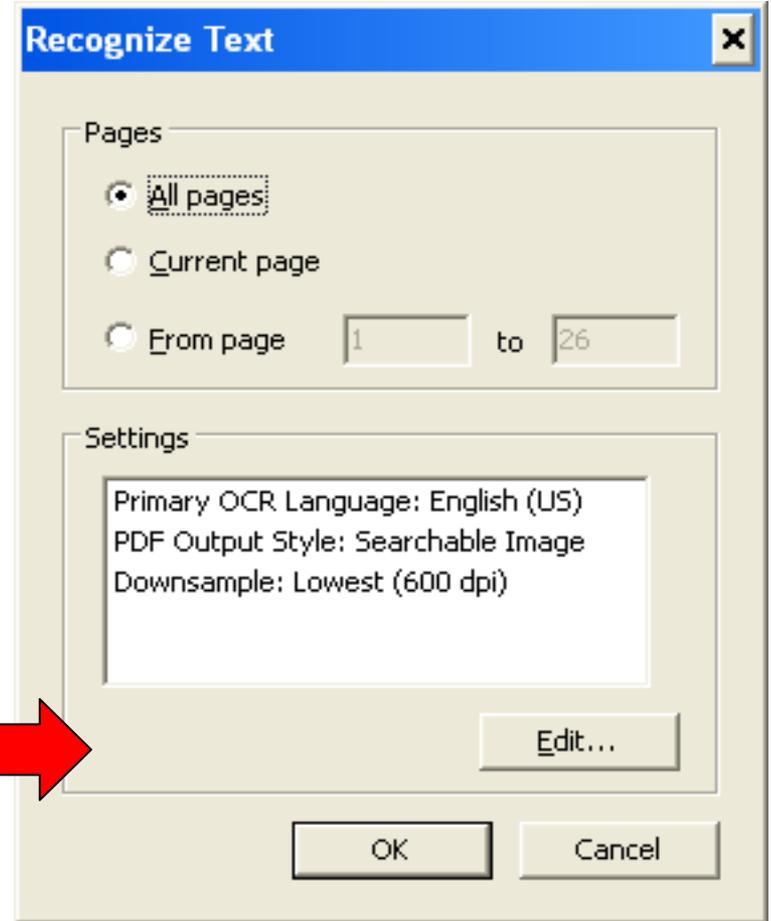
Step 3

- A **Recognize Text** dialog box will appear. Select **All pages**.

Use the following settings:

- Primary OCR Language: English (US)
Note: The language selected reflects the language the document is written in.
- PDF Output Style: Searchable Image (Exact)
- Downsample: None

- If a setting needs to be corrected, click **Edit**. (In this example, the *PDF Output Style* must be corrected.)

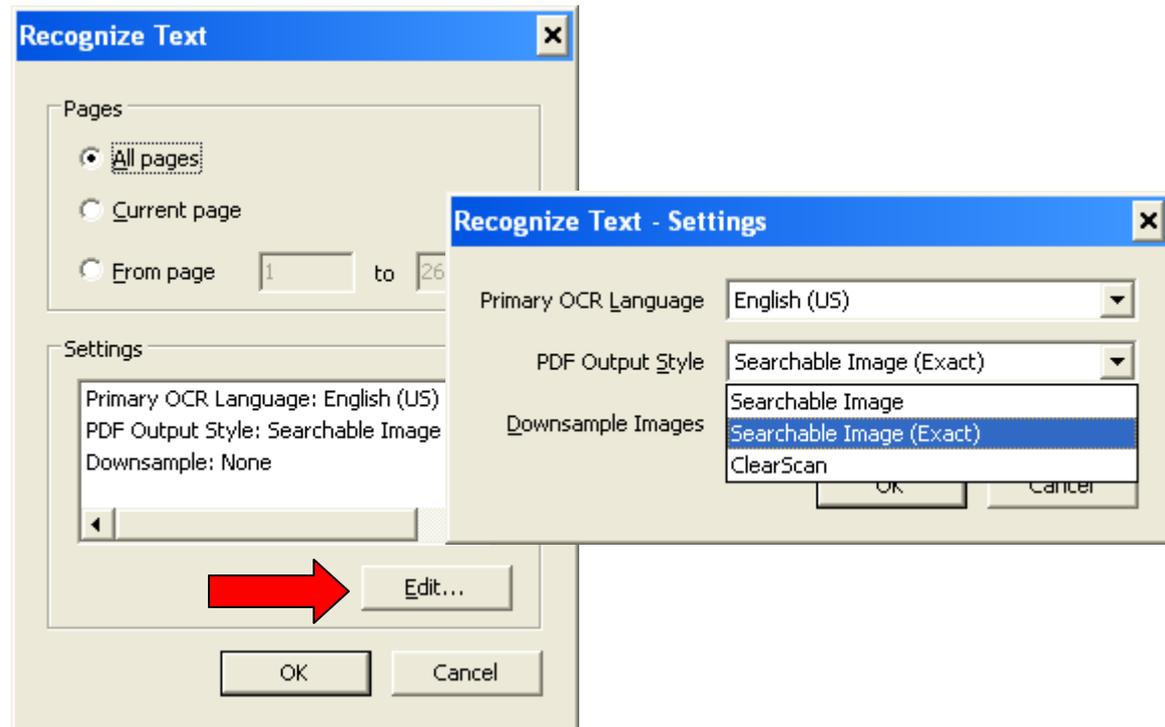




USAID
FROM THE AMERICAN PEOPLE

Step 4

- To correct a setting, click **Edit**.
- Click the dropdown menu next to the setting you want to change.
- Click to select the new setting. (Here, the *PDF Output Style* is changed.)
- Click **OK**.

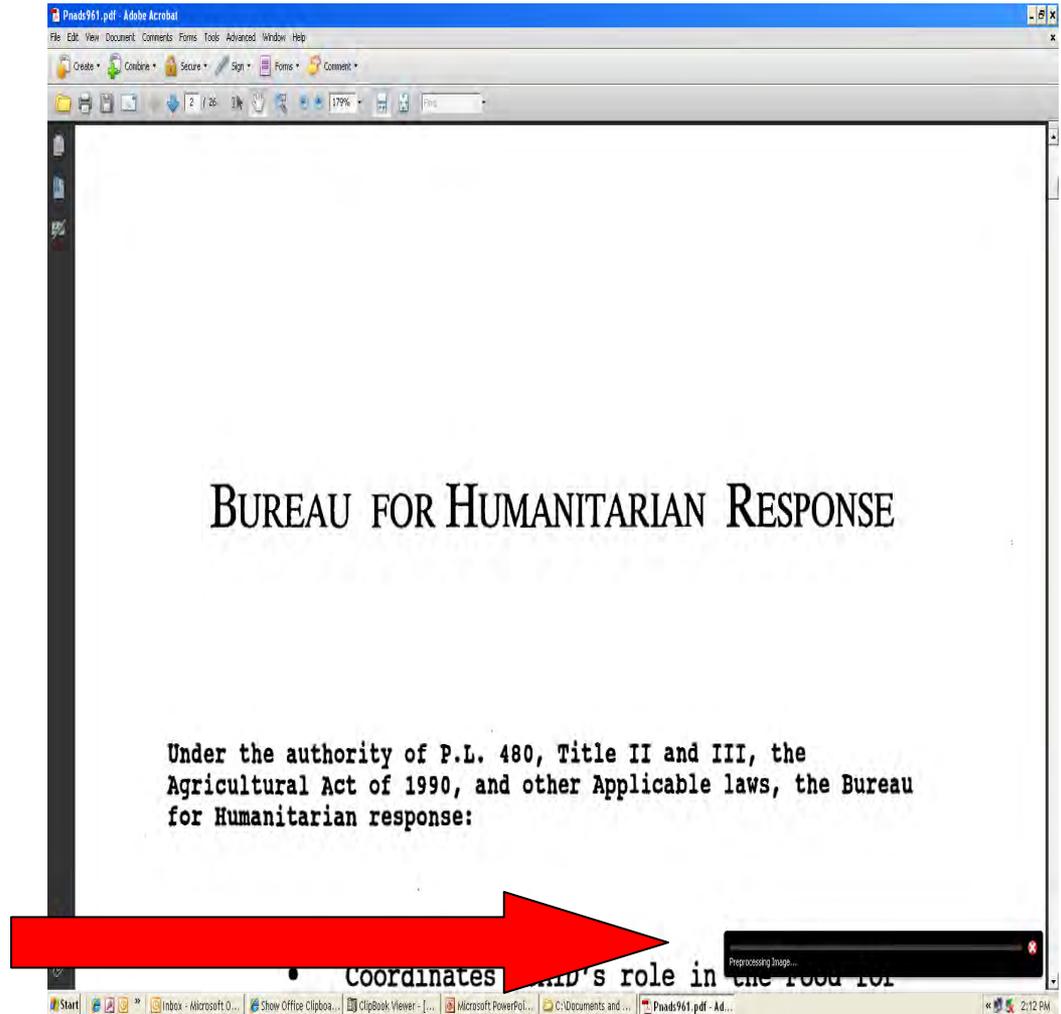




USAID
FROM THE AMERICAN PEOPLE

Step 5

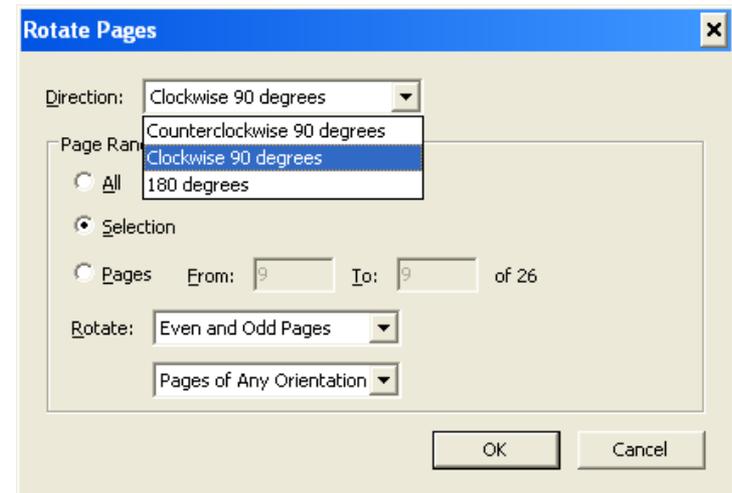
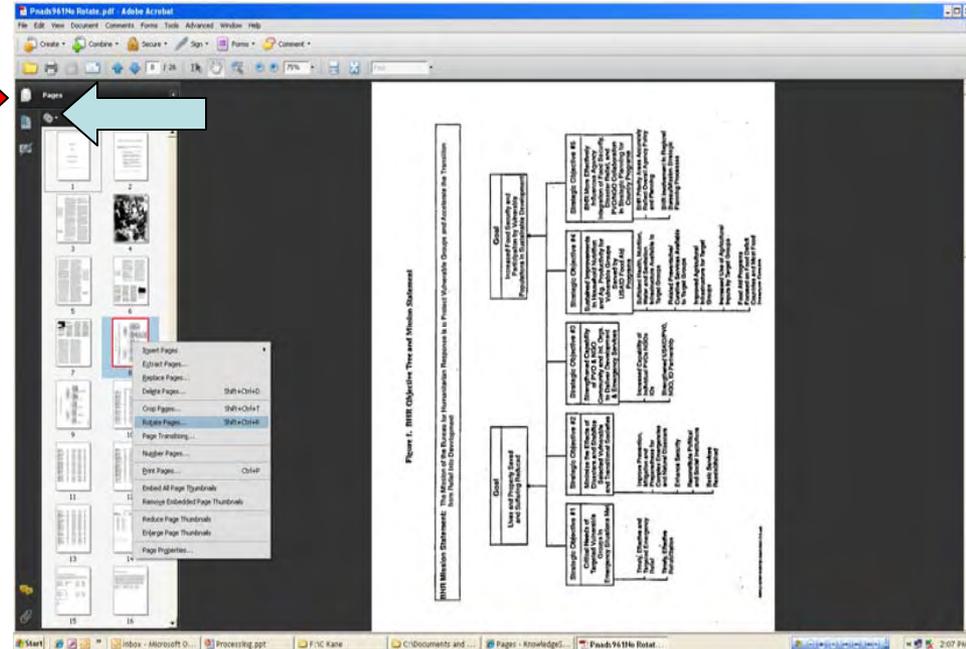
- When you click **OK** in the **Recognize Text** box, the OCR engine converts the PDF to searchable text.





Step 6

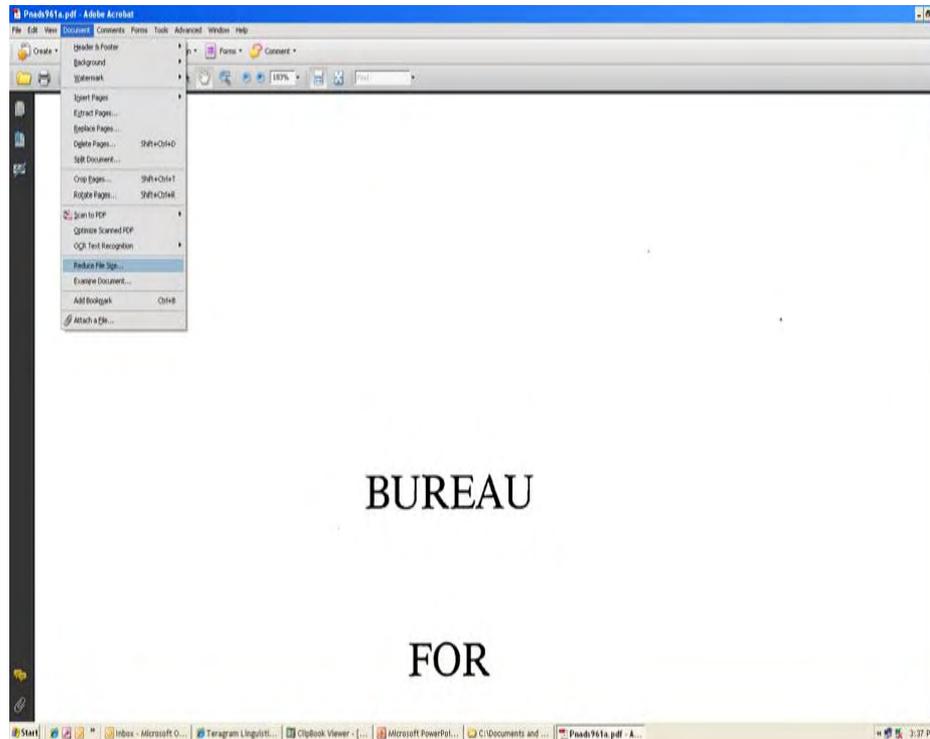
- After the conversion is complete, select **Pages** on the left side of the screen to verify that all of the frames are in their proper orientation.
- To change a page's orientation, select the **Options** icon (below **Pages**).
- Select **Rotate Pages**.
- Use the **Direction** dropdown to correct the orientation.
- Click the **Selection** button to apply the correction only to a single page.
- Click **OK**.



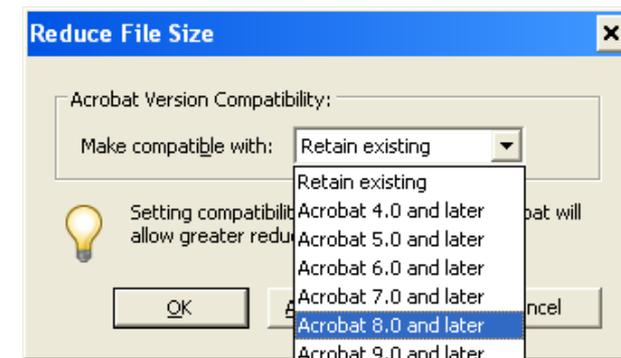


Step 7

- From the **Document** dropdown menu, select **Reduce File Size**.



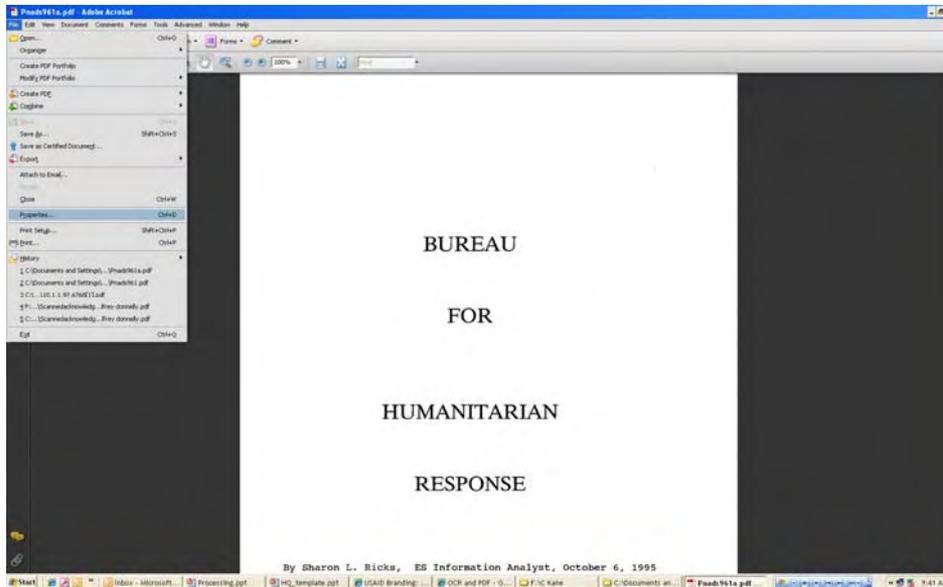
- From the **Make compatible with** dropdown menu, select **Acrobat 8.0 and later**.
- Click **OK**.



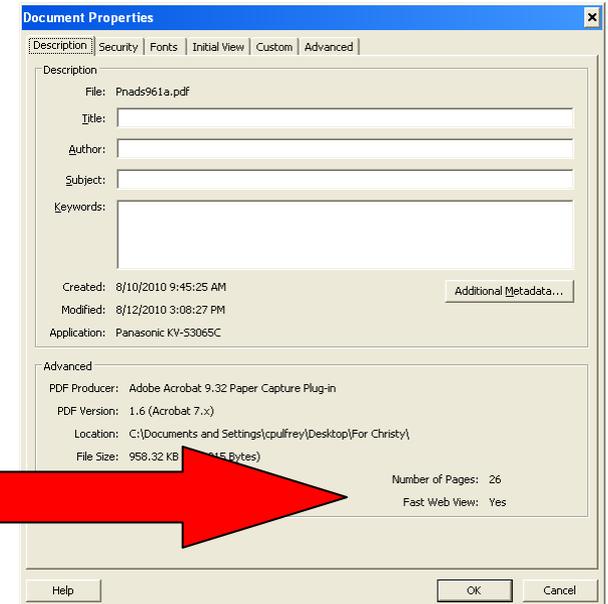


Step 8

- From the **File** dropdown menu, select **Properties**.



- A dialog box will appear. Confirm that the **Fast Web View** is **Yes**.



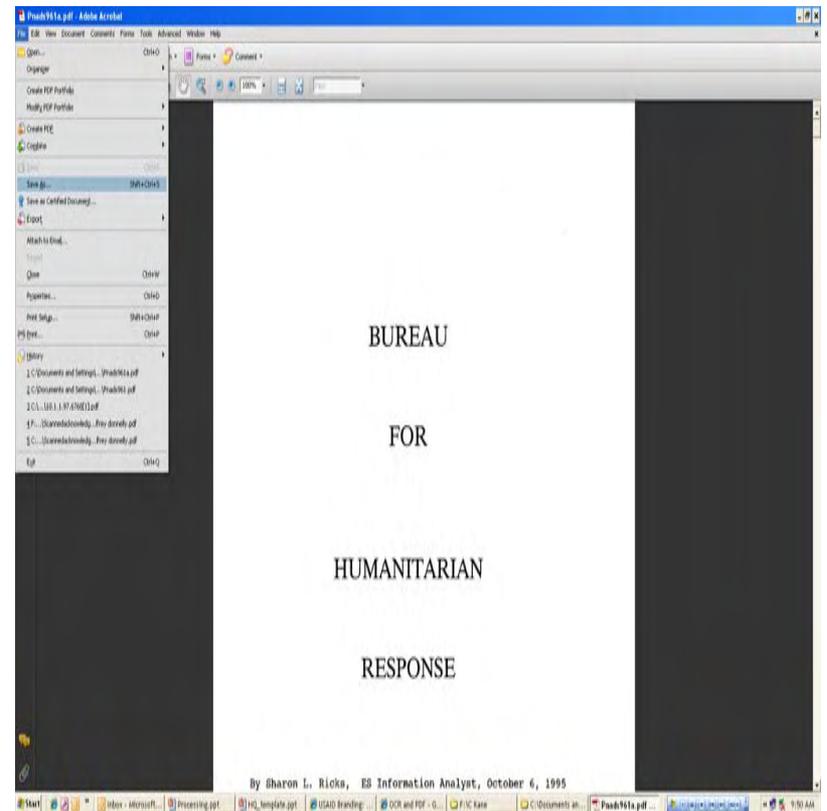


USAID
FROM THE AMERICAN PEOPLE

Step 8 (cont'd)

- If the **Fast Web View** is set to **No**, change it by selecting **Save as**.

Then return to **File > Properties** to make sure **Fast Web View** is now **Yes**.

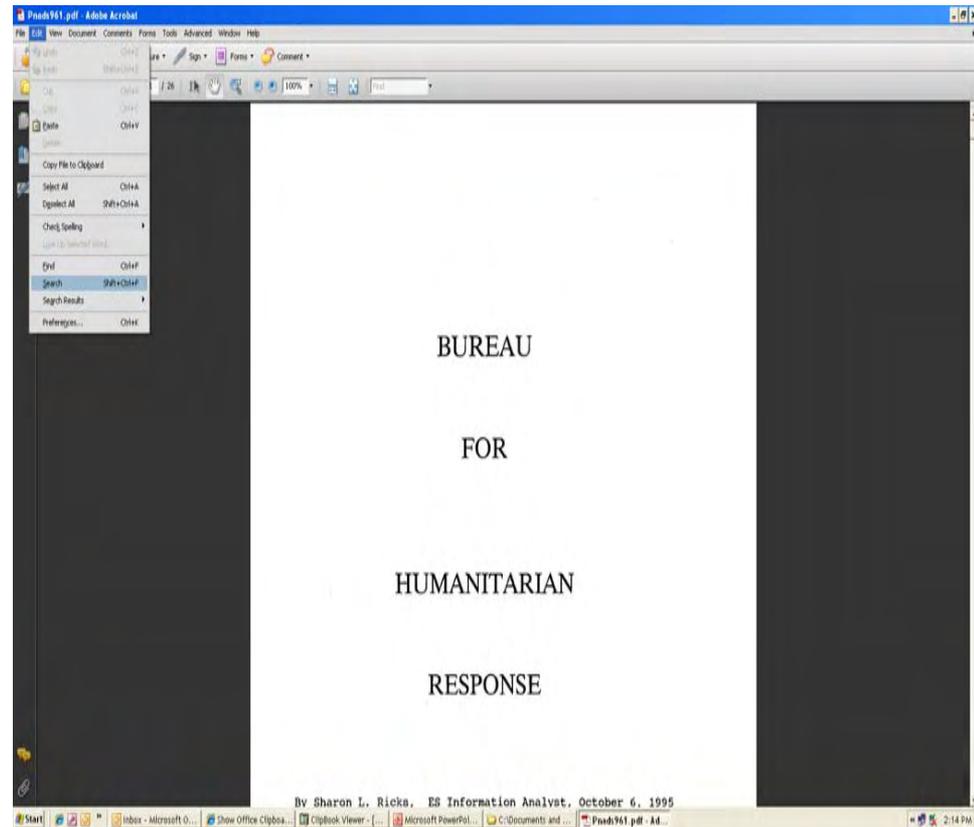




USAID
FROM THE AMERICAN PEOPLE

Step 9

- From the **Edit** dropdown menu, select **Search**.





USAID
FROM THE AMERICAN PEOPLE

Step 9 (cont'd)

- A dialog box appears with the question, “What word or phrase would you like to search for?” Type in the letter “e” and then click **Search**.

Note: E is used because it is the most frequently occurring letter in the English alphabet.





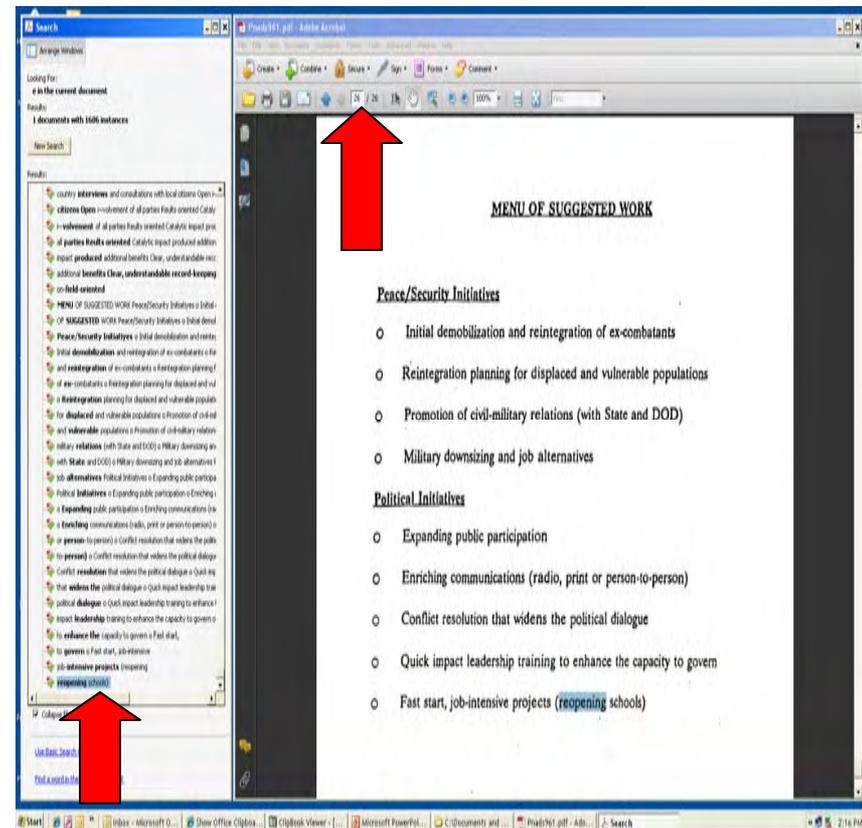
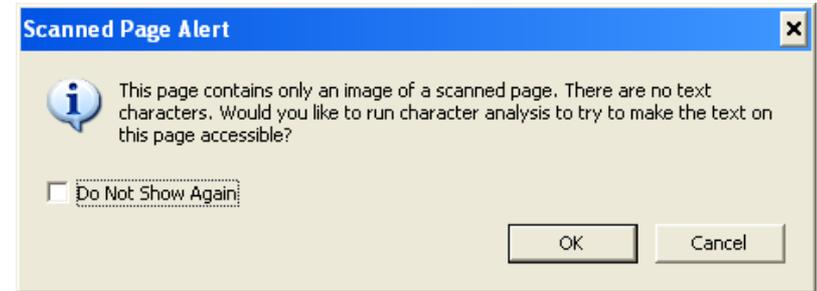
Step 10

If you receive a **Scanned Page Alert**, then the document was not converted. You'll need to try again, repeating the conversion process.

If the search finds words containing the letter “e” then your document has successfully been converted to searchable text.

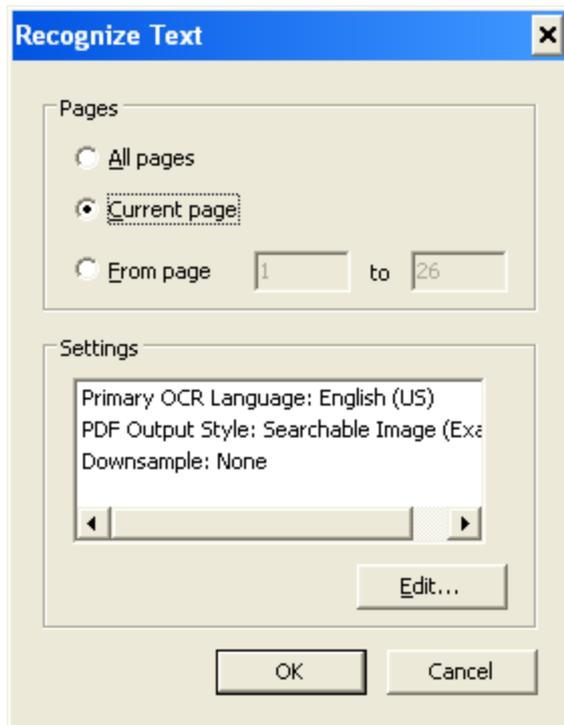
Note: Confirm that the document's first and final frames were converted to searchable text:

- Drag the scroll bar from the first frame to the final frame.
- Click on the final frame.
- Check that the page number count reflects the final frame as the last page.



Step 11

If the last page is not reflected in the total page count, it was not converted to searchable text and will need to be converted.



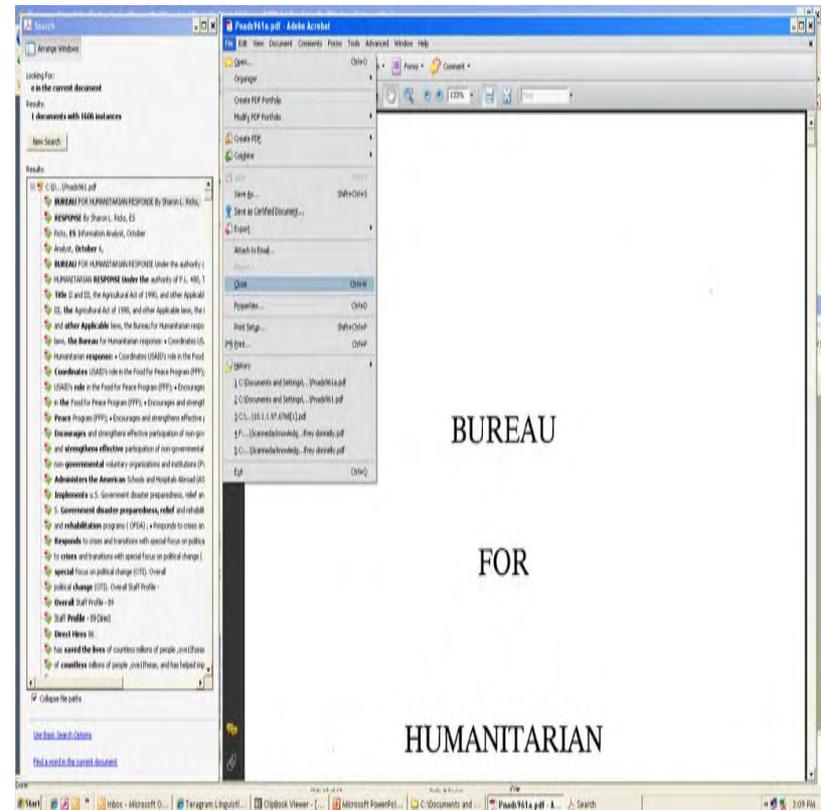
- Click on the last page.
- From the **Document** dropdown menu, select **OCR Text Recognition > Recognize Text Using OCR**.
- When you reach the **Recognize Text** dialog box, select **Current page** rather than **All pages**.
- Click **OK**.
- Select **File > Save As** to save the document.



USAID
FROM THE AMERICAN PEOPLE

Step 12

- Select **Close** to close the document and complete the conversion process.





USAID
FROM THE AMERICAN PEOPLE

Quick Review of the OCR process

1. Open the PDF file in Adobe.
2. From the **Document** dropdown menu, select **OCR Text Recognition > Recognize Text Using OCR**.
3. From the **Recognize Text** dialog box that appears, select **All pages**. Confirm the following settings:
 - Primary OCR Language: English (US)
 - PDF Output Style: Searchable Image (Exact)
 - Downsample: None
4. If necessary, click **Edit** to change settings. Click **OK**.
5. The OCR engine will convert the PDF to searchable text.
6. Confirm that the frames are all in their proper orientation. If necessary, correct the orientation by selecting **Pages**, then the **Options** icon, then **Rotate Pages**.
7. From the **Document** dropdown, select **Reduce File Size > Make compatible with Acrobat 8.0 or later**.
8. From the **File** dropdown, select **Properties**. Confirm that the **Fast Web View** is **Yes**. If **No**, select **Save as**.
9. From the **Edit** dropdown, select **Search**. Type in the letter “e” and then click **OK**.
10. When the search is complete, confirm that the first and final frames of the PDF have been converted to searchable text. If the final page has not been converted, click on the final page, then repeat Step 3.
11. In the **Recognize Text** dialog box, select **Current page**. Click **OK**. From the **Document** dropdown, select **Save as**.
12. Close the PDF.



USAID
FROM THE AMERICAN PEOPLE

The Knowledge Services Center's Development Experience Clearinghouse welcomes any questions or comments you may have: ksc@usaid.gov.