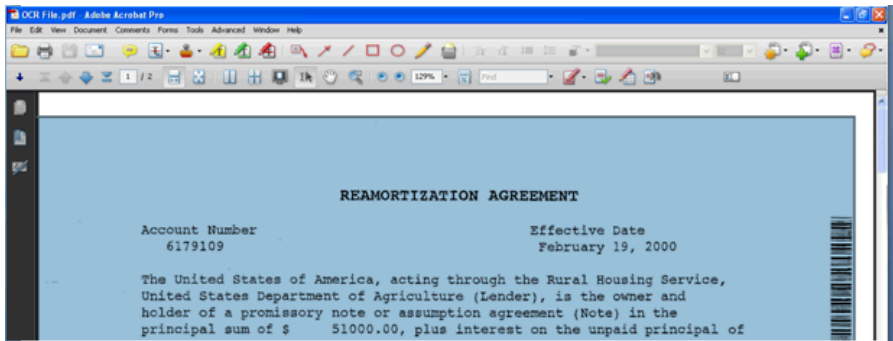# Converting Scanned Documents to Searchable Documents

Background

You probably know that our court's standard preference number two indicates a strong preference for most documents filed in ECF to be converted rather than scanned. The reason our judges issued this preference has to do with text searchability. Documents *scanned* to PDF for filing are not searchable–that is, you cannot search using Acrobat's Find tool (Ctrl+F or Edit>Find), nor can you cut and paste text. However, Adobe Acrobat features an **Optical Character Recognition (OCR)** tool which can, in many cases, convert a scanned PDF document to a searchable PDF document.

When is OCR Necessary?

The first step to understanding and using OCR is to recognize when a document is not already searchable. With a scanned document open and the Select tool selected, if you click anywhere in the text, an **unsearchable** document will **not** allow you to select any text OR it will turn blue, as shown here:

How to Run OCR

STEP	ACTION
1	With the scanned document open, click **Document** on the Acrobat menu bar.
2	Select **OCR Text Recognition**.
3	Select **Recognize text using OCR**.
4	The Recognize Text dialog box opens. Click **OK**. Warning: As the OCR process is underway, a black status bar will appear in the lower right corner of your Acrobat window. Do not click in the document until the status bar is gone. Interrupting the OCR process may lead to errors.

Watch for Conversion Errors

As helpful as OCR can be, it is not a perfect tool: **Some documents will not convert accurately**. For example, a PDF document heavily laden with graphics or a PDF document that has been copied, faxed, or scanned a number of times may become unreadable to OCR. Here is an excerpt from one such document:

Even though pages on a site may have been updated, you may be viewing old information if your browser's cache (pronounced "cash," a type of electronic memory) is not being cleared as frequently as it should be.

This is how OCR interpreted the same text:

Even though pages on a sire may have Peen updated, you ma)4:l,e viewIng old information If your browser's cache (pronounced "caSh: 11 type of electronic memory) IsmHletng cleared as ~. frequently as it st.ould be.

Always be sure to proofread all text if you have used the OCR tool.

Handwriting May Not Convert

While handwritten PDF documents will not usually translate well through OCR, typed text surrounding the handwriting will. Here is a scanned PDF document with a handwritten date:

ORDERS that the time in which SunTrust Bank may answer or otherwise respond to the

Complaint be extended through and including January 13, 2011.

This 29th day of Dee, 2010

This is the same text as interpreted by OCR:

ORDERS that the time in which SunTrust Bank may answer or otherwise respond to the

Complaint be extended through and including January 13,2011.

Thi~oflJ~ $J},i)

In this example, notice how **OCR was able to recognize the text** even though it could not accurately recognize the handwritten characters.

Converting Scanned Documents to Searchable Text Documents