

## Converting an Image-Only PDF to a Real Text PDF

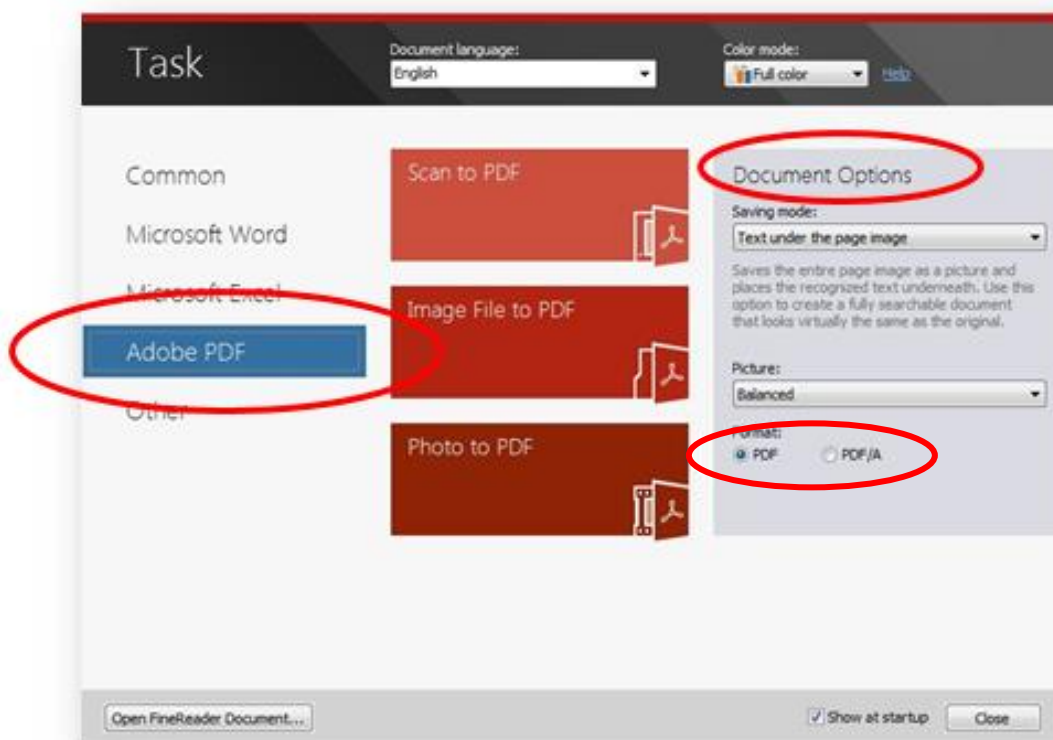
This document provides instructions to easily convert an image-only PDF into a real text PDF file using ABBYY Fine Reader 12.

### STEP 1: Open ABBYY Fine Reader

A task dialogue box will open automatically, providing access to some common tools.

### STEP 2: Select the Adobe PDF Tab

This tab can be found on the far left of the task box (see screenshot below.)



### STEP 3: Set the Document Options

The Document Options can be found in the far right side of the Task window. There are three options that need to be set (see screenshot above).

- Set *Saving Mode* to **Text under the page image** in the drop down menu.
- Set *Picture* to **Balanced** in the drop down menu.
- Set *Format* to **PDF** (not PDF/A).

### **STEP 4: Select the Image File to PDF Option**

Click on the **Image File to PDF** button in the central column of the Task window.

### **STEP 5: Browse for Image or PDF File to be Converted**

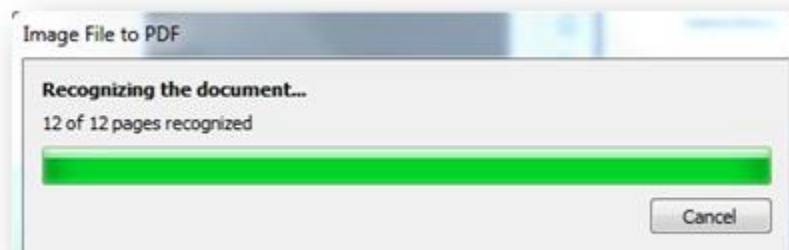
Another dialogue box will open. Use this window to browse to the correct folder where your file is and open the file.

### **STEP 6: Wait for the Document to be Recognized**

After opening the correct file in ABBYY Fine Reader, the software will read through the whole document and recognize each document feature. ABBYY has three primary document feature types.

- 1. Text areas**
- 2. Picture areas**
- 3. Table areas**

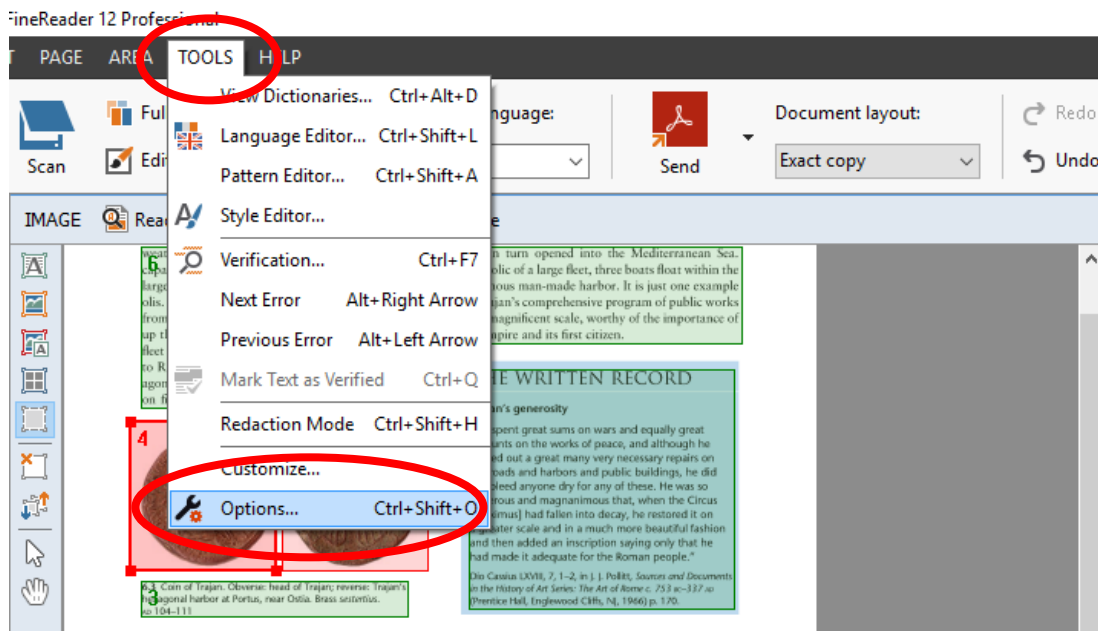
After initially scanning the whole document, ABBYY attempts to properly assign these three area categories to each element in the document so that it will be read and converted properly (see screenshot below).



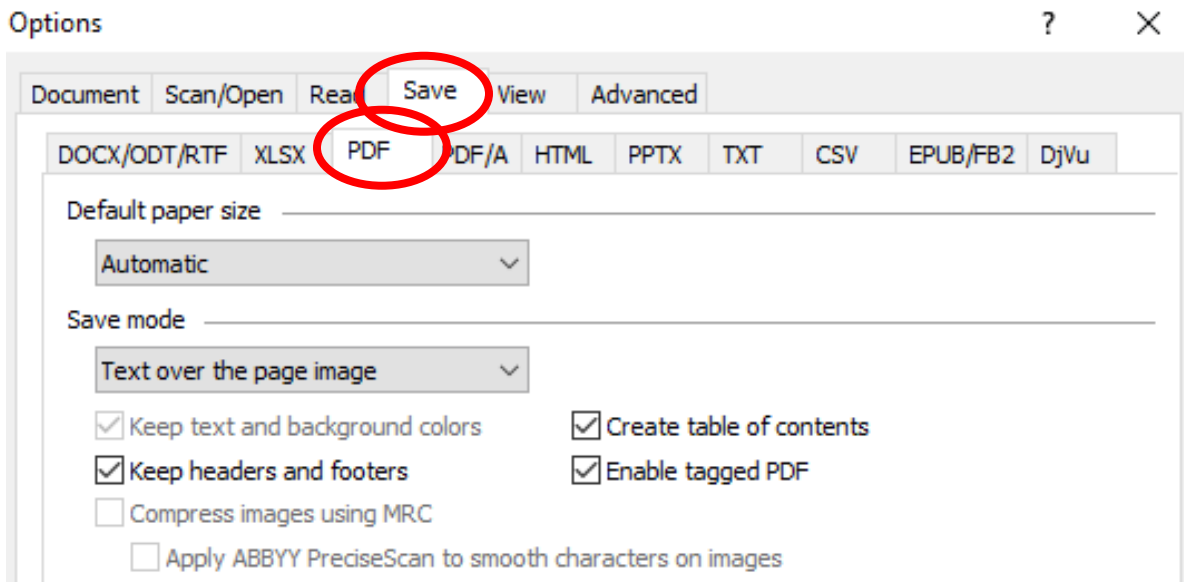
After ABBY has finished recognizing the document move on to the next step.

## STEP 7: Set Save Options

Click on **Tool** and then select **Options**, both outlined below, at the bottom of the drop down menu.

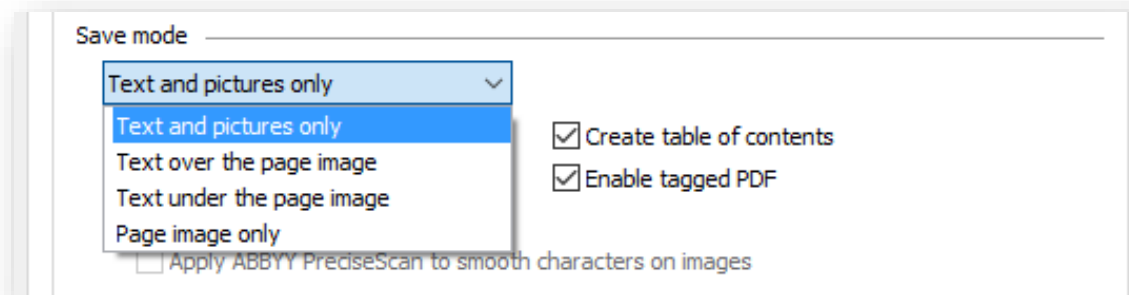


- In the options window ensure that **Save** and then **PDF** are selected as outlined below.



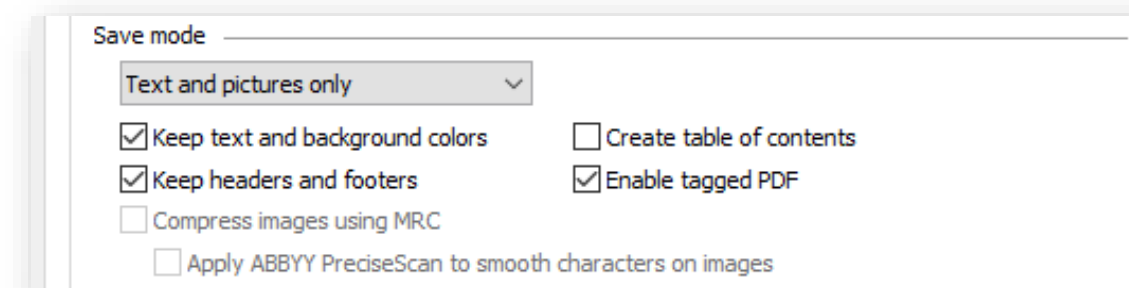
- In the Save PDF Window set **Default paper size** to **Automatic**.
- Set the **Save mode setting** according to the following guidelines.

## Save Mode: Drop Down Menu



- A. **Text and pictures only** should be used if the document needs to be fully accessible or if there *are* many graphical errors in the original PDF such as black streaks and blurry and unclear text.
- B. **Text over the page image** should be used if the text in the original PDF is visually unclear, if the document does not need to be fully accessible, or if there are *not* many graphical errors such as smears or streaks.
- C. **Text under the page image** should only be used if the text in the original PDF is clear, if there are *not* many graphical errors, and if the PDF does not need to be fully accessible.
- D. **Page image only** should never be used.

## Save Mode: Check Boxes



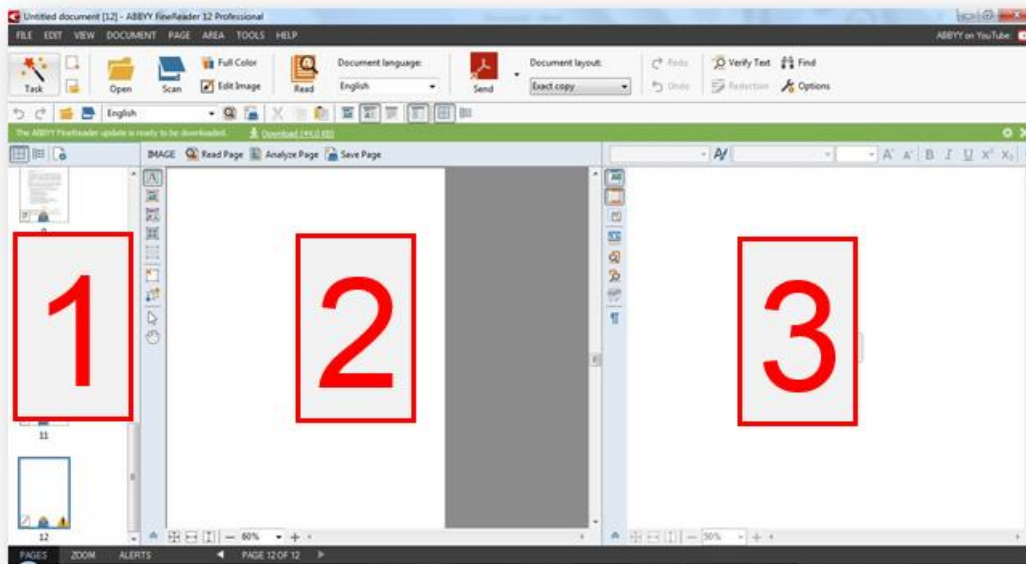
- A. **Keep text and background colors** should be checked if you want text and background coloration to be present.
- B. **Keep headers and footers** should be checked, unless the headers and footers of the original document contain non-vital content such as copyright information or watermarks.
- C. **Create table of contents** can be checked if the document is long and if headings are present.
- D. **Enabled tagged PDF** should be checked.

## STEP 8: Learning the Interface

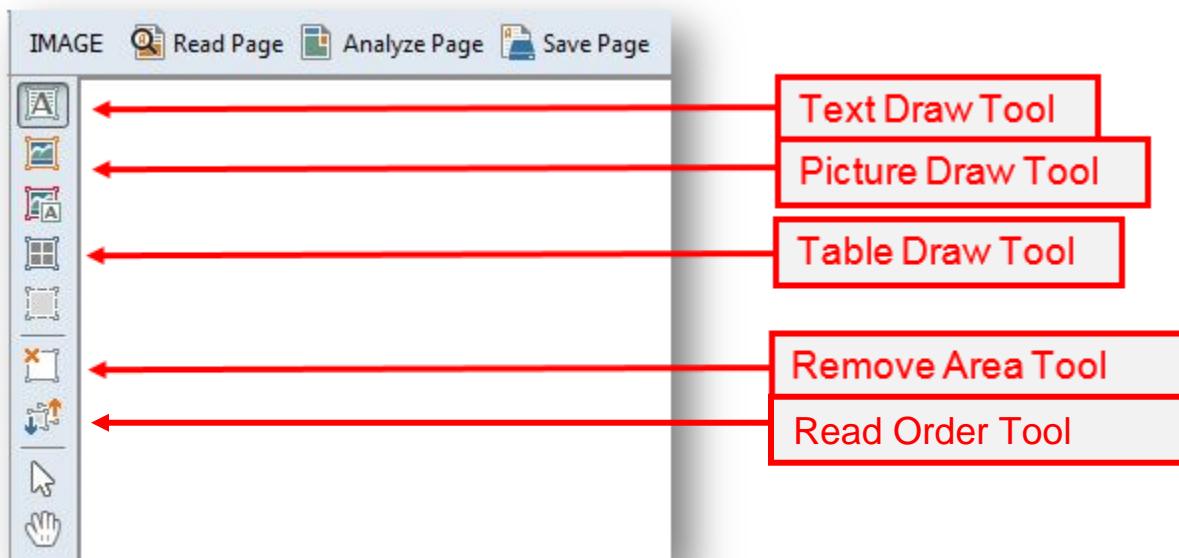
Before using ABBY to convert documents, there are a few things to learn first.

ABBYY's primary work window is broken into three parts:

1. Document Navigation partition on far left (**number 1 below**)
2. Document Recognition and Reading partition in the middle (**number 2 below**)
3. Document Output and Creation partition on the right (**number 3 below**)



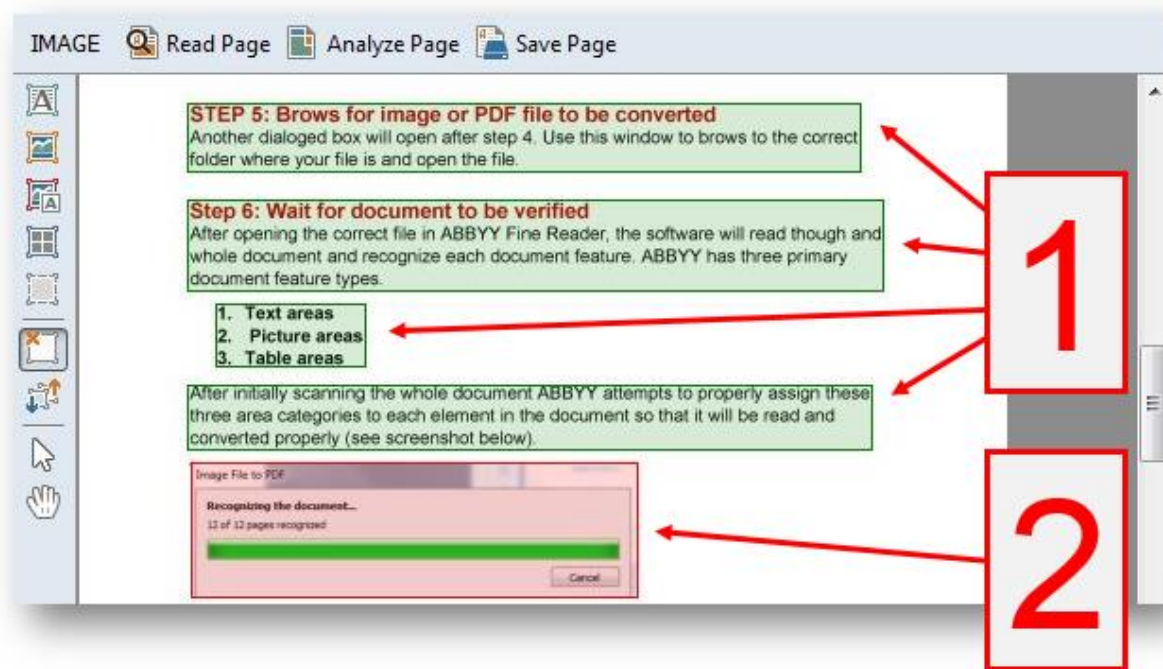
The tools for verifying and correcting the document recognition can be found in partition 2 above. On the upper left hand side of partition 2 are several buttons which allow for the area drawing tools to be selected (see screenshot below).



These buttons allow for the correct drawing tool to be selected in order to change or remove each area in the text. Remember, as mentioned in step 6, that these areas are **text**, **picture**, and **table**.

Look though the document in the Document Recognition and Reading partition and notice the boxes drawn over each document element. The **green boxes** are drawn over text, the **red boxes** are drawn over images and the **blue boxes** are drawn over tables  
**NOTE:** depending on your document, you may only see one or two of these areas drawn because the document may not have images or tables.

The screen shot below shows how ABBYY divides the document by areas. Text areas are labeled by number **1** and the image is labeled by number **2**.



## STEP 9: Verify and Fix the Document Recognition

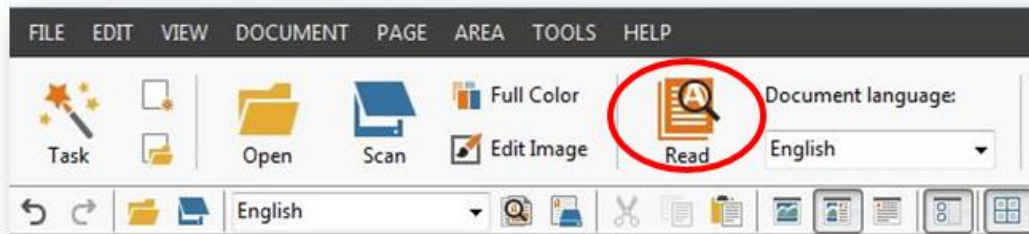
Check each page for correct area recognition. Navigate by using the navigation partition in the far left hand side of the main window.

- If an area is not properly recognized, delete it by selecting the **Remove Area Tool** and clicking on the desired area box to be removed.
- Select the correct draw tool and re-draw the box around the area.



## STEP 10: Read the Document

Click the **Read** button to re-read all the changes you have made to the document. This button can be found in the toolbar at the top of the screen (see screenshot below).

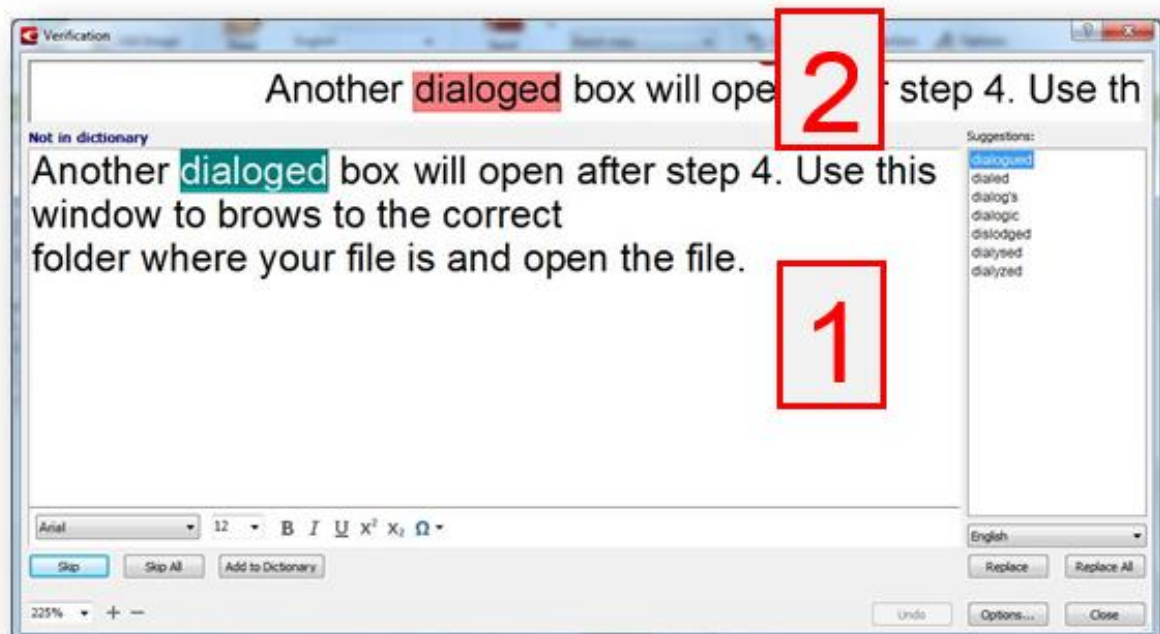


## STEP 11: Verify the Text

The last step is to verify the text in the document. To access this tool, click on the **Verify Text** button at the top of the screen (see screenshot below).



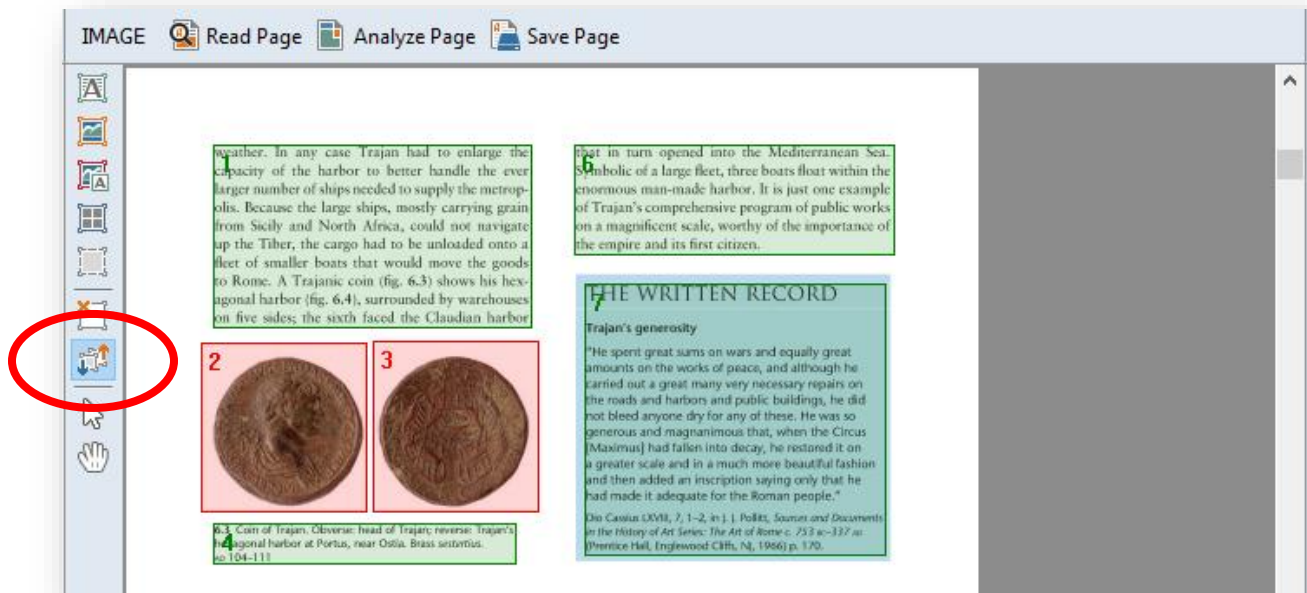
The Verify Text tool allows for the uncertain characters and words in the text recognition process to be manually verified. Each character or word will be outlined in a central partition, number 1 below, and the original image or text is provided at the top of the window, number 2 below.



- If the character outlined in the central partition, labeled 1 above, has been correctly recognized, click on the **Skip** button or press **Enter** on your keyboard, and the next uncertain element of the document will be outlined.
- If the character outlined in the central partition is not correct, correct it and then move to the next uncertain element by clicking on the **Confirm** button that appears or by pressing **Enter** on your keyboard.

## STEP 12: Correct the Reading Order

Select the reading order tool, outlined below, and check that each element on each page is numbered correctly so that the content is ordered as it should be read, starting at 1.



- If the reading order is incorrect, start from the beginning of each page and click on each element in the order that you would read though the page. You will notice that each box is reassigned to its correct ordering for proper reading.
- Check and correct each page.

## STEP 12: Send to PDF

Click the **Send** button at the top of the screen.

