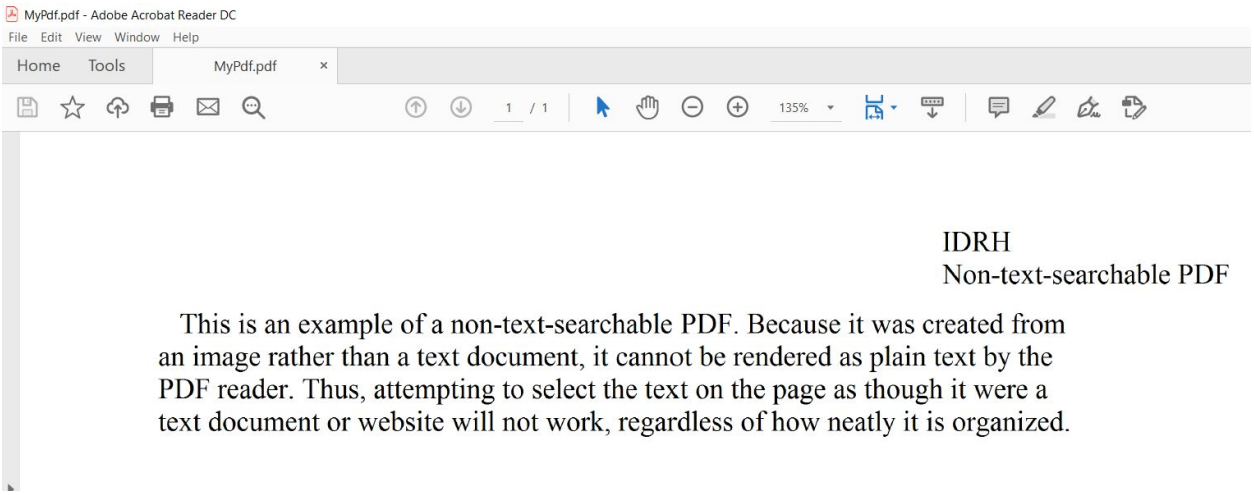
Testing a scanned and image-based PDF

The following includes instructions for testing an image-based PDF.

The image below is an example of a scanned PDF in which you cannot select or find text using traditional methods. While it can be challenging to read text from a scanned/image-based PDF, you can use Provar to test this type of documentation by creating a custom API. (Please refer to this support article, Creating custom APIs for more information.)



Prerequisites

You will need a Java-ocr and a PDF box jar. You can reference the Importing and Executing JAR files support article to learn more. You can also download the related files here:

- Java-ocr
- PDF box jar

You will also need to download TESS data from the following URL and store it within the root folder of the project.

- Tess data

Instructions

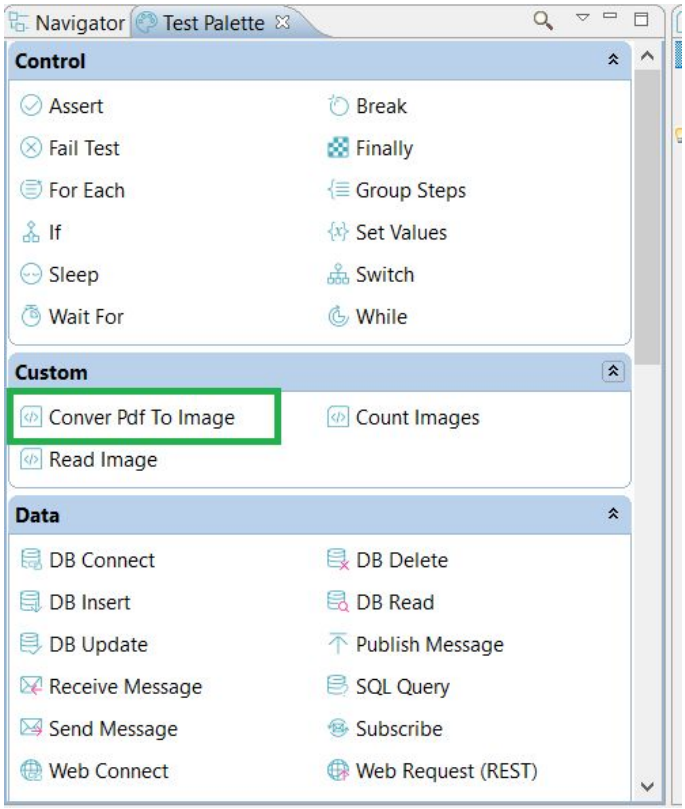
Step 1: Download and import the following files. Place them within the **src** > **customapis** folder.

- ConverPdfToImage.java
- CountImages.java
- ReadImage.java files

Note: *The* ***src folder*** *is located within the* ***Navigator*** *tab of Provar.*

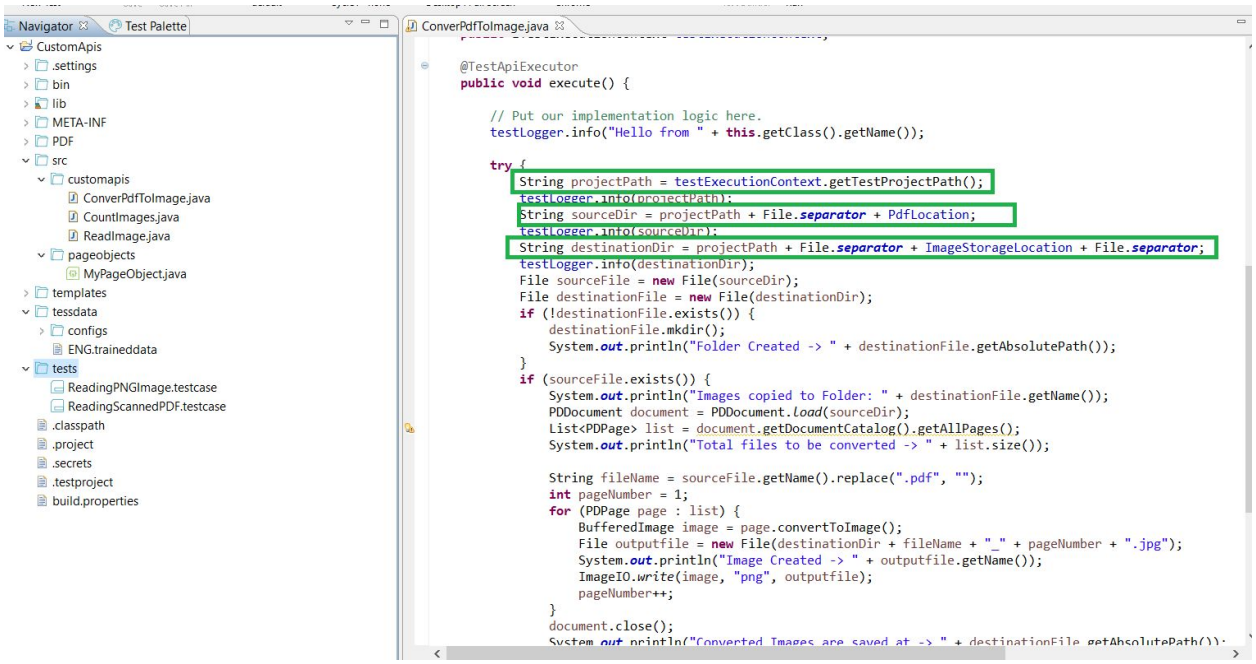
Step 2: Manually trigger a **Refresh and Recompile**. Please refer to this support article to learn more.

Step 3: From the Test Palette, drag **Convert PDF To Image** into an image API within your test case.



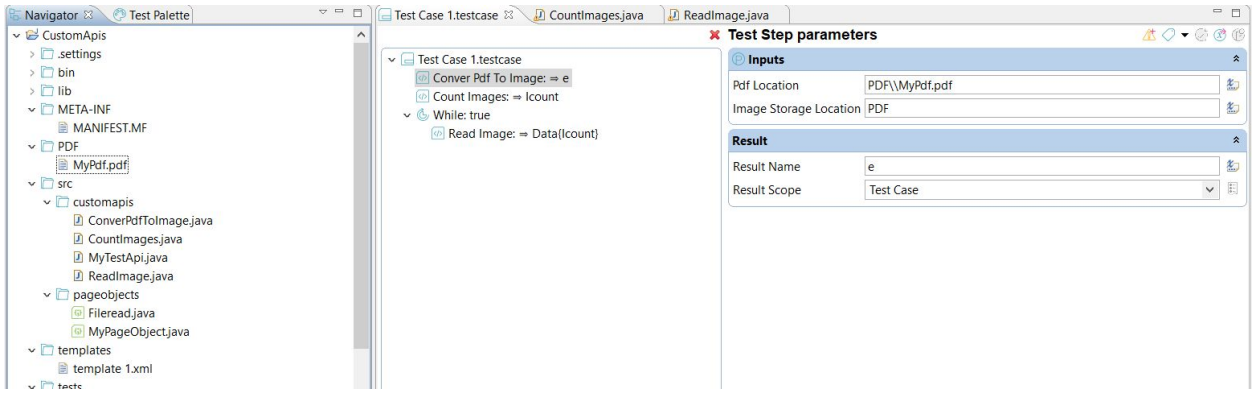
Step 4: Open the test step parameters screen. Within the **PDF Location** field, insert the location of the PDF. The location where the PDF is stored can be edited in a custom API Java file. Within the **Result Name** field, insert the name of the file.

Note: *Enter the name of the output variable in the* ***Result Name*** *field.*

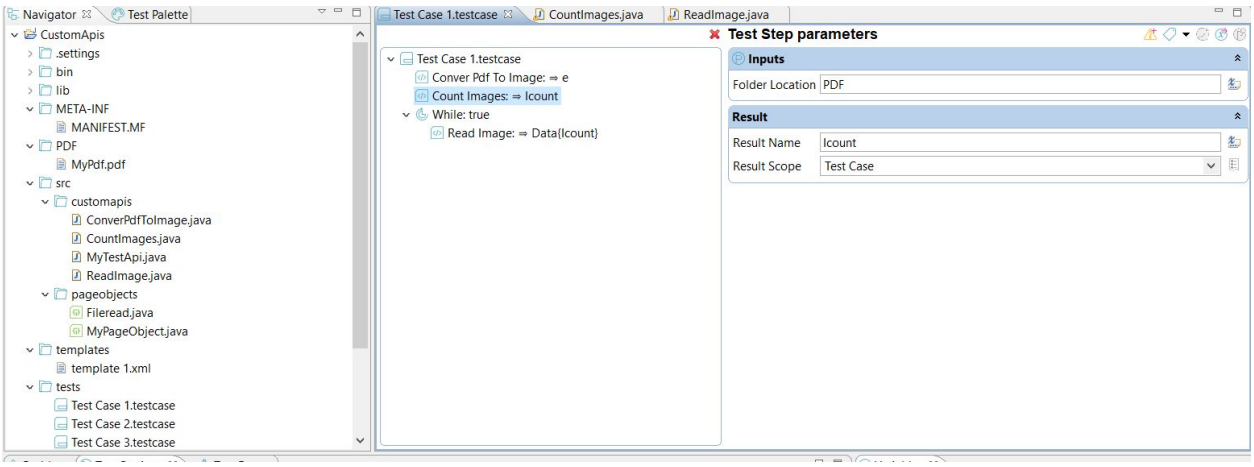


Note: *Users can update the location by changing the projectPath, sourceDir and destinationDir variables. These are highlighted in the image above.*

In the **Image Storage Location** field, specify the directory where you want the images to be stored. Based on the number of pages, one or more images will be created.



Step 5: In order to read the data from every image, you will need to count the images. Drag the second API, **Count Images,** from the Test Palette. Within the Test Step parameters display, enter the location where your images are stored within the **Folder Location** field. The results will be stored in the result variable **Icount**.



Step 6: In this step, you can read the data within the image. Drag the third API, **Read Image,** from the Test Palette. Within the **Image Location** field, enter the location of the images that you want to read. Provar will store the result in the **Result Name** variable based on the images you read. For example, the first image will be labeled Data1, the second image will be labeled Data2, etc.