

Aspire OCR

THIS ITEM IS BEING DEPRECATED.

1. Download and install **tesseract-ocr** from <http://code.google.com/p/tesseract-ocr/>
2. Create the Groovy Stage as shown below to OCR the Tiff Image.

```
<component name="OCRTiffImage" subType="default" factoryName="aspire-groovy">
  <config>
    <script>
      <![CDATA[
        tiff_filename = doc.getText("FILENAME");
        tiff_fullurl = doc.getText("COMPLETE_FILENAME");
        text_filename = doc.getText("FILENAME") + ".txt";
        text_file = new File(text_filename);
        if (!text_file.exists()) {
          try
          {
            // note tesseract.exe will append the .txt extension in output filename
            command1 = "tesseract.exe " + tiff_fullurl + " ./OCR_Files/" + tiff_filename;
            proc2 = command1.execute();
          }
          catch (Exception e) {
            println "Error Occured during OCR";
          }
        }
        parentdoc = job.getParentJob().getObject();
        if (text_file.exists()) {
          contents = text_file.getText();
          parentdoc.addCDATAElement ("DocumentContent", contents);
        }
      ]]>
    </script>
  </config>
</component>
```