## Requirements

* Create a word document or a pdf document given an image
* Detect and recognize Text in an input Image
* Detect sketches in the input image and draw a vector image of the sketch
* Improve the quality of the sketch (Smoothen the edges, remove noises, etc.)
* Reduce the size of the document
* Create document using images without user’s support

## Main Functionalities in the Application

### Input an image

The input image size should not be larger than A4 size (2480 x 3508). The software can not recognize words in the font size less than 8. The software will neglect any lines which do not have a length more than 10 pixels. The image can be color or B&W

### Prepare the image for processing

The input has to go through some processing in order to quality image. Erotion and dilation has be done on the image.

### Separate the text area and the sketch/diagram

This has to be done in order to recognize the text using tesseract. And the sketch/diagram will process separately.

### Digitalize the texts in the image

This can be done by writing an algorithm or using an platform like tesseract. Other than recognizing the text software should identify other properties as well(Font Size etc.)

### Recognize the specific patterns in the sketch

There are algorithms to recognize patterns (line detection) in a image. Hit and Miss transform, Hough transform. Since creating an algorithm from the scratch is unnecessary one of these already existing algorithms can be used.

### Vectorize the sketch

Given the special points of the sketch software can draw the sketch/diagram using MatPlotLib library. The sketch/diagram and text should be in the same order as the input image.