**ASSIGNMENT**

**Introduction**

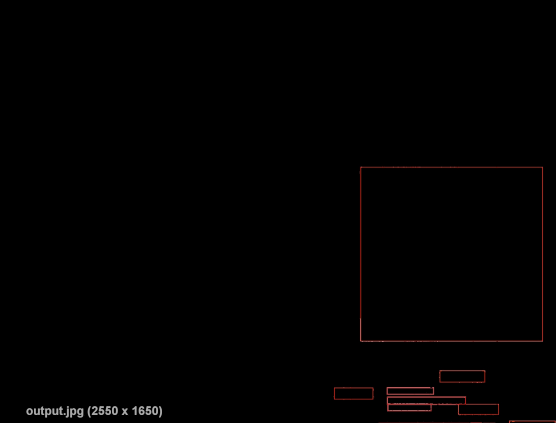
Identify and Extract the fields from the given pdf.

**Approach**

After seeing the image , the required fields are marked in red in colour. it is easy to identify red colour region with opencv techniques.

**Steps followed:**

* The given image contains red boxes and goal is to extract the the data from red boxes.
* Identify Red colour boxes and mask the non red colours part to zero.
* Using contours method identify rectangle and square boxes.
* Identified boxes coordinates are passed to paddle ocr for text extraction inside boxes.

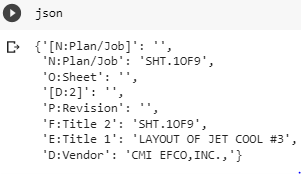


Boxes contains required text

**Post processing**

Using logics the extracted text is to stored in json file as key is Title and value is text which denoted by array in image.

output is shown in below figure.



**Conclusions**

**The given fields are identified and extracted text from them based on colour identifications.**

**Recommendations**

**For the given problem to identify fields in image can be developed generic solution using deep learning techniques.**