

```
In [159]: import PyPDF2 as PDF
import openpyxl as xl
```

```
In [160]: file = open(r'C:\Users\Arsalan Khan\Downloads\ISO SAE 214342021.pdf','rb')
wb = xl.load_workbook('Assignment_2.xlsx')
ws = wb['Sheet2']
#inp = ws.cell(1,1).value
#inp = '[' +inp +']'
```

```
In [161]: pdf = PDF.PdfFileReader(file)
```

```
In [162]: print(pdf.numPages)
```

88

```
In [163]: pdfList = []
for val in range(88):
    pdfList.append(pdf.getPage(val))
```

```
In [164]: type(pdfList)
```

Out[164]: list

```
In [169]: for row in range(1,ws.max_row+1):
    inp = ws.cell(row,1).value
    inp = '[' +inp+']'
    string = ''
    for page in pdfList:
        if page.extractText().find(inp) != -1:
            string = page.extractText()
            break
    val = (string.find(inp))
    out = ''
    val+=12
    while string[val:val+2] != '.\n':# and string[val+1] != '\n' and string[val+2]
        #print(string[val],end='')
        out +=string[val]
        val +=1
    ws.cell(row,2).value = out
```

```
In [170]: wb.save('Assignment_2.xlsx')
```

```
In [ ]:
```