

Using tesseract OCR

Formats that Tesseract accepts

File Input Formats

Tesseract will only take image files for input. These include:

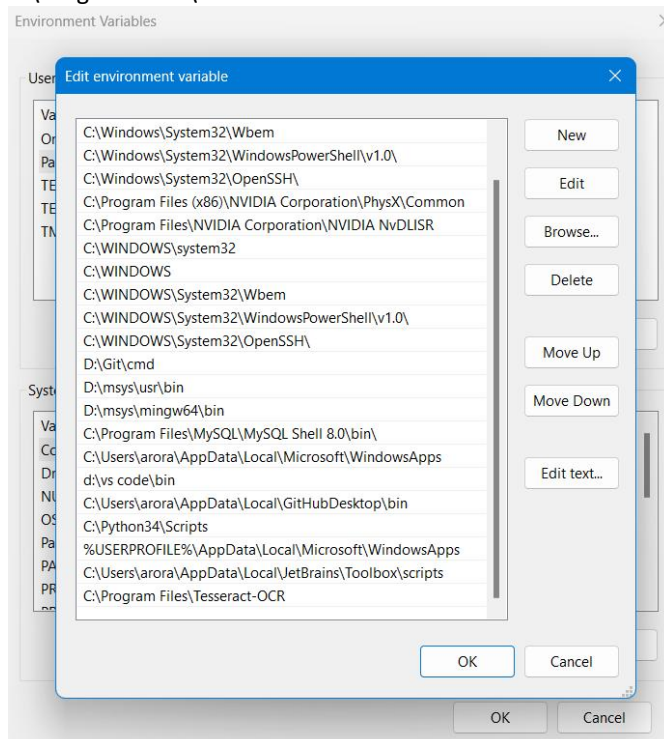
- TIFF (preferred)
- JPG
- PNG

Link: <https://guides.library.illinois.edu/c.php?g=347520&p=4121426>



Steps performed:

1) Installed tesseract ocr. After installation I set the path. Go to system>advance system setting>environment variables>user variable>select path>click edit>add the path of tesseract. For ex: C:\Program Files\Tesseract-OCR



2) Opened command prompt as administrator and gave the path of the folder in which the image file lies

3) Passed the command `tesseract table.jpg -l eng`

Output:

Enrollment Capacity and Utilization Report

Based on 10/31/2021 Audited Registers

Page 1 of 215

Bold indeates co-located campuses.

Target Calculation By School

No.of No. of

Cluste Cluster

Geo Org Bidg Target Target PK -SpecRms Rms

Dist__ 10 Org Name 10 Bldg Name Enroll__Cap_Util_Cap+__Reported+ _Needed+

T Mots PS.15-M Mors PS.15-M 179 2171 7 T z

T MOIS PS.10-m wore PS.19- 176 a7 ot cy 2 2

1 M020 PS.20-M 020 --PS.20-M 35 ost Py 5 3

1 MOM PS.-M was ~PS.34-M 258 Ea) Ee 3 3

1 M0S3 PS.63-M 06s PS.63-M 186 2 «73 Ey 4 2

1 M064 ~PS.64-M Moe --PS.64 (OLD 71)-M 28 mC % 1 2

7 M080 MANHATTANCHARTER SCHOOL & w0ss SHS. 56-M 110 ar 8 ° ° 2

7 -M0S4 _M0B4SPED PS 15-M a es «8

PS.61-M 45 “a

PS.188-M 103 1m

‘M094 SPED -M TOTAL 189 we 68

1 Mit0 PS.110-M mito 330 a «2 Ey 4 3

1 Mia PS.134-M Drrry 7 0 8 8 6 2

1 Miao PS.140-M Dr) ET) | 8 % 6 3

1 iad PS.142-M mrax 30 mT Eo 2 3

7 Mies PS. 1e4-M wta7 77 7 86 % 5 3

1 Miss PS.188-M mies 380 mS 18 FE) 3

1 M262 JHS.292-M M056 290 2m 8 °

1 M315 PS.315-M moet 302 Bt 35 1 3

1 M320 MANHATTANCHARTER -M maz PS.142-M 208 m8 ° 1 2

Analysis: it wasn't able to identify some of the characters efficiently. Other methods were more efficient