

COURSE TITLE: CEF 506 - PYTHON/PERL PROGRAMMING DEVELOPMENT

NAME	MATRICULATION NUMBER
FRU ISIDORE CHE	FE12A078
THEOPHILUS WABA NASALI	FE12A183
NKENG NEWTON	FE12A140
MOBA MELVIS RINGNYU	FE12A107
ENOMBE THIERRY EWANE	FE12A053
ALANGI DERRICK	FE12A113

Title: Programs to get the metadata of pdf and image files

Metadata literally means ‘data about data’. Metadata provide additional information about a certain file, such as *its author, creation date, date modified, file size, possible copyright restrictions or the application used to create the file.*

Hardware Specification

This application was developed and run on a machine with the following hardware requirement. It could be run on a machine with better specification

RAM: 3.00GB

Processor: AMD Phenom™ II N620 Dual-Core Processor

Processor speed: 2.80GHz

Hard drive: 1TB

Mark: Hewlett Packard (HP) ProBook 6455b

Software Specification

The following software used:

Operating System: Ubuntu 14.04

Editor: Sublime text editor 3

Command-line: Ubuntu terminal

FLOWCHARTS

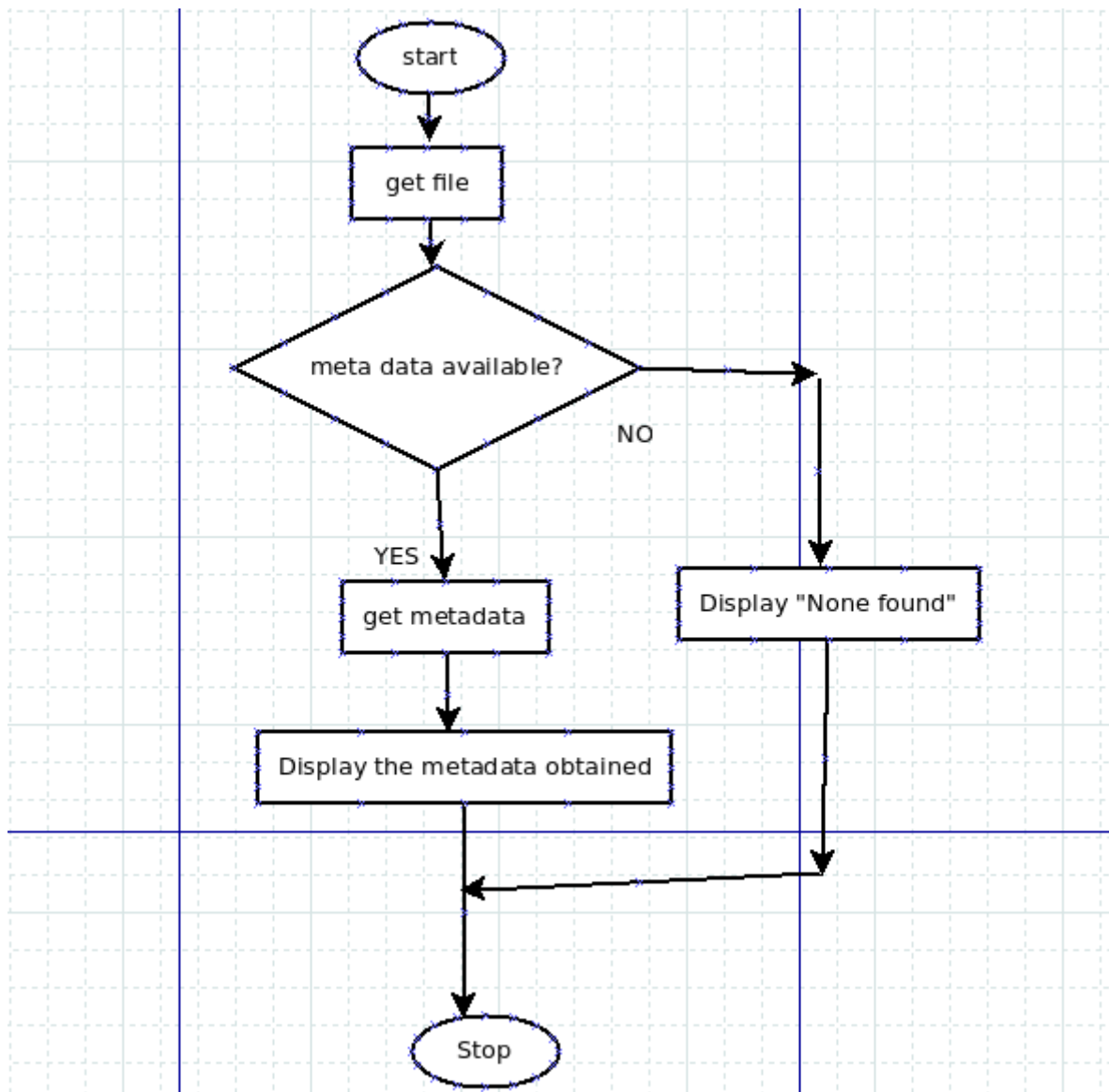


Figure 1: FlowChart of the Metadata of a pdf file

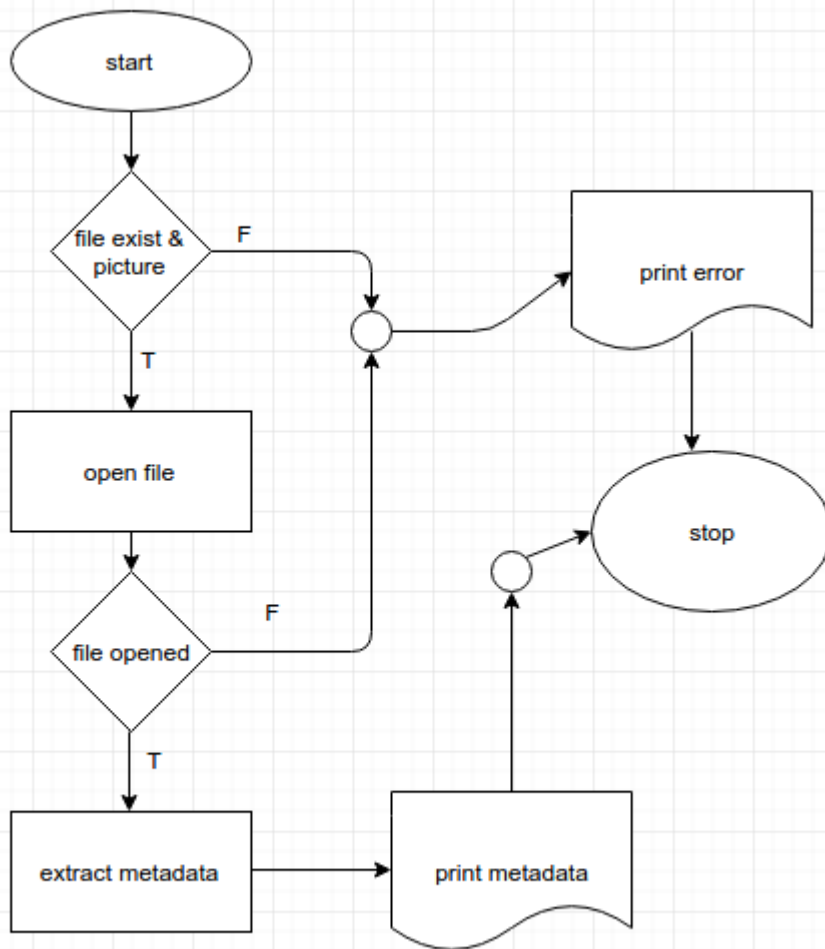


Figure 2: FlowChart metadata of an image file

CODE

```
1  #!/usr/bin/env python2.7
2
3  import pyPdf
4  import optparse
5  from pyPdf import PdfFileReader
6
7  def printMeta(fileName):
8      pdfFile = PdfFileReader(file(fileName, 'rb'))
9      docInfo = pdfFile.getDocumentInfo()
10     #print docInfo.producer
11     print '[*] PDF MetaData For:' + str(fileName)
12     for metaItem in docInfo:
13         print '[+]' + metaItem + ':' + docInfo[metaItem]
14
15  def main():
16      parser = optparse.OptionParser('usage %prog "+\
17      "-F ')
18      parser.add_option('-F', dest='fileName', type='string',\
19      help='specify PDF file name')
20
21      (options, args) = parser.parse_args()
22      fileName = options.fileName
23      if fileName == None:
24          print parser.usage
25          exit(0)
26      else:
27          printMeta(fileName)
28
29  if __name__ == '__main__':
30      main()
```

Figure 3: code to get Metadata of a pdf file

```
meta_data_extractor.py ×
1 from PIL import Image
2 from PIL.ExifTags import TAGS
3
4 def get_exif_data(fname):
5     """Get embedded EXIF data from image file."""
6     ret = {}
7     try:
8         # Open the file
9         img = Image.open(fname)
10        if hasattr( img, '_getexif' ):
11            # extract the EXIF data
12            exifinfo = img._getexif()
13            if exifinfo != None:
14                # Loop through the data extracted and decode
15                for tag, value in exifinfo.iteritems():
16                    decoded = TAGS.get(tag, tag)
17                    ret[decoded] = value
18        # Catch in case of I/O Error
19    except IOError:
20        print 'IOERROR ' + fname
21    return ret
22
23 if __name__ == '__main__':
24     fileName = 'food.jpg'
25     exif = get_exif_data(fileName)
26     print "\n\n"
27     print exif
28     print "\n\n"
29
```

Figure 4: Code to get metadata of an image

OUTPUT

```
ringnyu@ringnyu-HP-ProBook-6455b: ~/Desktop/python ass 2
ringnyu@ringnyu-HP-ProBook-6455b:~/Desktop/python ass 2$ python metadata.py -F '
/home/ringnyu/Desktop/cef 504 Sem2 2015-2016 - 5 Rules-based Systems II.pdf'
[*] PDF MetaData For:/home/ringnyu/Desktop/cef 504 Sem2 2015-2016 - 5 Rules-based Systems II.pdf
[+]/Author:Denis
[+]/Producer:LibreOffice 4.1
[+]/Creator:Impress
[+]/CreationDate:D:20160504172118+01'00'
ringnyu@ringnyu-HP-ProBook-6455b:~/Desktop/python ass 2$ python metadata.py -F '
/media/ringnyu/BACKUP6/SCHOOL/java_tutorial.pdf'
[*] PDF MetaData For:/media/ringnyu/BACKUP6/SCHOOL/java_tutorial.pdf
[+]/CreationDate:D:20140330081206
[+]/Author:ZARA
[+]/Producer:Microsoft® Office Word 2007
[+]/Creator:Microsoft® Office Word 2007
[+]/ModDate:D:20140330081206
[+]/Title:Hibernate Tutorial
ringnyu@ringnyu-HP-ProBook-6455b:~/Desktop/python ass 2$
```

Figure 5: Output of the metadata of a pdf file

```
derick@d3rick:~/Desktop/MetaDataExtractor$ python meta_data_extractor.py

{'YResolution': (72, 1), 'ResolutionUnit': 2, 'FlashPixVersion': '0100', 'ExposureMode': 0, 'Flash': 24, 'SceneCaptureType': 0, 'GPSInfo': {0: '
\x02\x02\x00\x00', 1: u'N', 2: ((4, 1), (8, 1), (5036, 100)), 3: u'E', 4: ((9, 1), (18, 1), (705, 100)), 5: '\x01', 6: (0, 1000), 7: ((5, 1), (5
2, 1), (13, 1)), 11: (22000, 1000), 16: u'M', 17: (338, 1), 27: 'ASCII\x00\x00\x00fused', 29: u'2016:05:16'}, 'YCbCrPositioning': 1, 'XResolutio
n': (72, 1), 'ImageUniqueID': u'f283e481df8034240000000000000000', 'Contrast': 0, 'Saturation': 0, 'ExposureProgram': 2, 'ColorSpace': 1, 'ExifI
mageWidth': 2992, 'DateTimeDigitized': u'2016:05:16 06:52:17', 'DateTimeOriginal': u'2016:05:16 06:52:17', 'ImageWidth': 2992, 'SceneType': '\x0
1', 'Software': u'HDR+ 1.0.118402515r', 'SubjectDistanceRange': 0, 'Make': u'Huawei', 'SensingMethod': 2, 'FNumber': (200, 100), 'CustomRendered
': 1, 'ApertureValue': (200, 100), 'FocalLength': (4670, 1000), 'ComponentsConfiguration': '\x01\x02\x03\x00', 'SubjectDistance': (0, 1), 'Exif0
ffset': 242, 'ExifImageHeight': 4000, 'ISOSpeedRatings': 788, 'Model': u'Nexus 6P', 'DateTime': u'2016:05:16 06:52:17', 'Orientation': 1, 'Expos
ureTime': (29441, 1000000), 'ImageLength': 4000, 'MaxApertureValue': (200, 100), 'Sharpness': 0, 'MeteringMode': 2, 'DigitalZoomRatio': (1, 1),
'ExifVersion': '0210'}

derick@d3rick:~/Desktop/MetaDataExtractor$
```

Figure 6: Output of the metadata of an image file