

# ACQUIRING META DATA FROM PDF'S AND IMAGES

By

ACHO ARNOLD EWIN (FE12A003)

NKWETEYIM DAISY AATSE  
(FE12A145)

FONJI FONKENG ROCARD  
(FE12A069)

EKWOGE ESEH BLAISE  
(FE12A050)

NNOKO DEGRANDO ANYOPEH  
(FE12A147)  
FOKO MABOU PINEL VALDES  
(FE12A063)

# Title

Acquiring meta data from pdf's and images

## Requirements

### Hardware Requirements

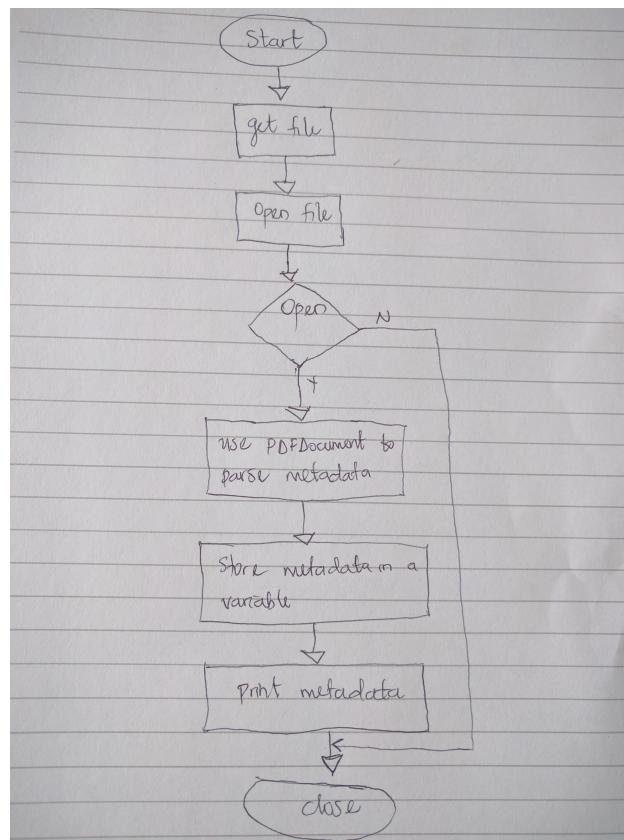
Part	Specification
RAM	16 GB 1600 MHz DDR3
Processor	2.5 GHz Intel Core i7

### Software Requirements

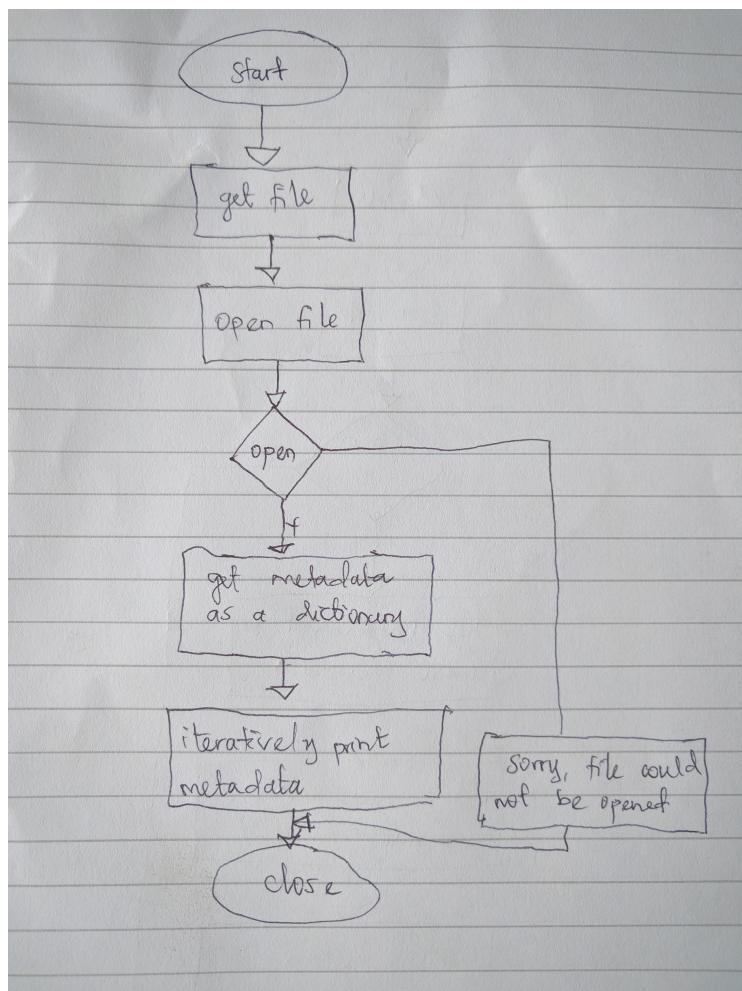
Software Type	Specifications
Operating System	OX X El Capitan
Terminal	interm
Python Version	Python 2.7.10
Zip Version	Zip 3.0

## Flow Chart

### Flow chart for Image Meta Data



## Flow Chart for PDF Meta Data



## Source Code

### Extracting Meta Data from Images

```
from PIL import Image  
from PIL.ExifTags import TAGS
```

```
try:  
    imageDetails = Image.open("IMG_20160516_065215.jpg")._getexif() #  
    Getting the image meta data as a dictionary  
  
    for (k,v) in imageDetails.iteritems():  
        print '%s = %s' % (TAGS.get(k), v) #Printing the meta data  
    except:  
        print 'Sorry the File Could Not Be Opened'
```

# Extracting Meta Data from PDF's

```
from pdfminer.pdfparser import PDFParser  
from pdfminer.pdfdocument import PDFDocument
```

try:

```
fp = open('diveintopython.pdf', 'rb')
```

```
parser = PDFParser(fp)
```

```
doc = PDFDocument(parser)
```

```
print doc.info[0]
```

except:

```
print 'Sorry the File Could Not Be Opened'
```

# Output

## Output from a PDF

```
[najela@Achos-MacBook-Pro:CEF-506-Assignment-2]$ python demo.py
[{"Producer": "XEP PDF Generator \x85 RenderX, Inc.", "Author": "Mark Pilgrim", "Creator": "XEP 3.9 Internal", "Title": "Dive Into Python", "ModDate": "D:20040618005740", "Keywords": "Python, Dive Into Python, tutorial, object-oriented, programming, documentation, book, free", "CreationDate": "D:20040618005740"}]
```

## Output for Image Metadata

```
[najela@Achos-MacBook-Pro:CEF-506-Assignment-2]$ python ./image.py
1. Achos-MacBook-Pro (zsh)
ImageWidth = 2992
ImageLength = 4000
ApertureValue = (200, 100)
DateTimeOriginal = 2016:05:16 06:52:17
DateTimeDigitized = 2016:05:16 06:52:17
MaxApertureValue = (200, 100)
ExifVersion = 0210
ComponentsConfiguration =
Contrast = 0
Flash = 24
FocalLength = (4670, 1000)
SubjectDistanceRange = 0
ExifImageWidth = 2992
Make = Huawei
Model = Nexus 6P
Orientation = 1
YCbCrPositioning = 1
SensingMethod = 2
DigitalZoomRatio = (1, 1)
XResolution = (72, 1)
YResolution = (72, 1)
ExposureTime = (29441, 1000000)
ImageUniqueID = f283e481df803424000000000000000000000000
ExposureProgram = 2
ColorSpace = 1
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), (52, 1), (13, 1)), 11: (22000, 1000), 16: u'M', 17: (338, 1), 27: 'ASCII\x00\x00\x00fused', 29: u'2016:05:16'}
ISOSpeedRatings = 788
ResolutionUnit = 2
MeteringMode = 2
FNumber = (200, 100)
Software = HDR+ 1.0.118402515r
DateTime = 2016:05:16 06:52:17
ShutterSpeedValue = (509, 100)
Saturation = 0
SceneType =
Sharpness = 0
CustomRendered = 1
FlashPixVersion = 0100
SubjectDistance = (0, 1)
SceneCaptureType = 0
ExifImageHeight = 4000
ExposureMode = 0
ExifOffset = 242
[najela@Achos-MacBook-Pro:CEF-506-Assignment-2]$
```