**Curriculum for Applied Learning**

**Operating System**

**CSE2005**

**Under**

**Dr. SELVAKUMAR K**

**Associate Professor**

***“File Arrangement in System based on Unsupervised Clustering”***

***By***

**Shivam Kumar Verma**

**B. Tech**

**in**

**Computer Science and Engineering**

**School of Computer Science & Engineerin**

**AIM**

1. To arrange the files in the system on the basis of file type and groups.
2. To copy file directly from Pen drive to specified folder using python.
3. Unsupervised clustering.

**MOTIVATION**

We always find ourselves in trouble while downloading files or copying files from Pen Drive to our system. We have to go to each folder type and paste our file.

So we think of automating this process by using the concept of unsupervised learning.

**ABSTRACT**

Arrangement of file is always being a headache to us. We always got frustrated while arranging or grouping the files.

Our objective is to copy file from pen drive to the specified group of folders automatically when we insert the pen drive.

Below is the sample code to detect the drive of the inserted pen drive.

import string

from ctypes import windll

import time

import os

def get\_drives():

drives = []

bitmask = windll.kernel32.GetLogicalDrives()

for letter in string:

if bitmask & 1:

drives.append(letter)

bitmask >>= 1

return drives

if \_\_name\_\_ == '\_\_main\_\_':

before = set(get\_drives())

pause = input("Please insert the USB device, then press ENTER")

print ('Please wait...')

time.sleep(5)

after = set(get\_drives())

drives = after - before

delta = len(drives)

if (delta):

for drive in drives:

if os.system("cd " + drive + ":") == 0:

newly\_mounted = drive

print ("There were %d drives added: %s. Newly mounted drive letter is %s" % (delta, drives, newly\_mounted))

else:

print("Sorry, I couldn't find any newly mounted drives.")

Distribution of project:

1. OS walk: to retrieve each folder and files present in pen Drive or any folder
2. Identifying the extension of each file.
3. Clustering of all files in different groups on the basis of file type and file name or tag name.
4. Unsupervised clustering of files.
5. Copying the clustered file in the specified group of folders.

**KEYWORDS**

1. OS PATH, WALK, MOVE etc. (OS library of python)

2. Unsupervised Clustering

**INTRODUCTION**

Generally, people find it difficult to copy files from their portable media to their PCs because files are not sorted according to their file type in windows. This is where our algorithm comes to play. Through the centroid clustering using K-means Algorithm we make clusters of input given to the algorithm.

In the algorithm we first calculate weight of the file on the basis of length of file name matched with all others file in the list and plot it on x axis then on the basis of files extension we plot it on Y axis.

After rendering all the input, we find the clusters as

**Methodology**

1. **Converting file name into weight array**

**Function for converting file name to weight array**

def weightmat(filea,fileb):

c=""

weight=0

for i in filea:

c+=i

if c in fileb:

weight+=1

else:

return weight

return weight

1. Converting extension into weight array

**Function for converting extension into weight array.**

def weightext():

if ext.upper() in audiosformat:

return (0)

elif ext.upper() in videosformat:

return (100)

elif ext.upper() in imagesformat:

return (200)

elif ext.upper() in docsformat:

return (300)

1. Merging two weight array and form weight matrix

costfname=[]

for i in range(len(fname)):

mxwt=0

for j in range(len(fname)):

if i!=j:

currwt=weightmat(fname[i],fname[j])

if mxwt<currwt:

mxwt=currwt

costfname.append(mxwt)

1. Find centroids in the weight matrix and make clusters of file with the found centroid.

kmeans = KMeans(n\_clusters=6).fit(mxt)

**Complete Working code:**

# -\*- coding: utf-8 -\*-

"""

Created on Wed Nov 1 01:41:30 2017

@author: sammacorpy

"""

from ctypes import windll

import time

import os

import shutil

import numpy as np

from sklearn.cluster import KMeans

audiosformat=['3GA', '4MP', '5XB', '5XE', '5XS', '669', '8SVX', 'A2B', 'A2I', 'A2M', 'AA', 'AA3', 'AAC', 'AAX', 'AB','ABC', 'ABM', 'AC3', 'ACD', 'ACD-BAK', 'ACD-ZIP', 'ACM', 'ACP', 'ACT', 'ADG', 'ADT', 'ADTS', 'ADV', 'AFC', 'AGM', 'AGR', 'AIF', 'AIFC', 'AIFF', 'AIMPPL', 'AKP', 'ALC', 'ALL', 'ALS', 'AMF', 'AMR', 'AMS', 'AMS', 'AMXD', 'AMZ', 'ANG', 'AOB', 'APE', 'APL', 'ARIA', 'ARIAX', 'ASD', 'AT3', 'AU', 'AUD', 'AUP', 'AVASTSOUNDS', 'AY', 'B4S', 'BAND', 'BAP', 'BDD', 'BIDULE', 'BNK', 'BRSTM', 'BUN', 'BWF', 'BWG', 'BWW', 'CAF', 'CAFF', 'CDA', 'CDDA', 'CDLX', 'CDO', 'CDR', 'CEL', 'CFA', 'CGRP', 'CIDB', 'CKB', 'CKF', 'CONFORM', 'COPY', 'CPR','CPT', 'CSH', 'CTS', 'CWB', 'CWP', 'CWS', 'CWT', 'DCF', 'DCM', 'DCT', 'DEWF', 'DF2', 'DFC', 'DFF', 'DIG', 'DIG', 'DLS', 'DM', 'DMC', 'DMF', 'DMSA', 'DMSE', 'DRA', 'DRG', 'DS', 'DS2', 'DSF', 'DSM', 'DSS', 'DTM', 'DTS', 'DTSHD', 'DVF', 'DWD', 'EFA', 'EFE', 'EFK', 'EFQ', 'EFS', 'EFV', 'EMD', 'EMP', 'EMX', 'ESPS', 'EXPRESSIONMAP', 'EXS', 'F2R', 'F32', 'F3R', 'F4A', 'F64', 'FDP', 'FEV', 'FLAC', 'FLM', 'FLP', 'FLP', 'FPA', 'FPR', 'FRG', 'FSB', 'FSC', 'FSM', 'FTM', 'FTM', 'FTMX', 'FZF', 'FZV', 'G721', 'G723', 'G726', 'GBPROJ', 'GBS', 'GIG', 'GP5', 'GPBANK', 'GPK', 'GPX', 'GROOVE', 'GSF', 'GSFLIB', 'GSM', 'H0', 'H4B', 'H5B', 'H5E', 'H5S', 'HBE', 'HDP', 'HMA', 'HSB', 'IAA', 'ICS', 'IFF', 'IGP', 'IGR', 'IMP', 'INS', 'INS', 'ISMA', 'ITI', 'ITLS', 'ITS', 'JAM', 'JSPF', 'K26', 'KAR', 'KFN', 'KMP', 'KOZ', 'KOZ', 'KPL', 'KRZ', 'KSC', 'KSF', 'KT3', 'L', 'LA', 'LOF', 'LOGIC', 'LOGICX', 'LSO', 'LWV', 'M3U', 'M3U8', 'M4A', 'M4B', 'M4P', 'M4R', 'M5P', 'MA1', 'MBR', 'MDC', 'MDR', 'MED', 'MGV', 'MID', 'MIDI', 'MINIGSF', 'MINIPSF', 'MINIUSF', 'MKA', 'MMF', 'MMLP', 'MMM', 'MMP', 'MMPZ', 'MO3', 'MOD', 'MOGG', 'MP2', 'MP3', 'MPA', 'MPC', 'MPDP', 'MPGA', 'MPU', 'MSCX', 'MSCZ', 'MSV', 'MT2', 'MTE', 'MTF', 'MTI', 'MTM', 'MTP', 'MTS', 'MUI', 'MUS', 'MUS', 'MUS', 'MUSX', 'MUX', 'MX3', 'MX4', 'MX5', 'MX5TEMPLATE', 'MXL', 'MXMF', 'MYR', 'NARRATIVE', 'NBS', 'NCW', 'NKB', 'NKC', 'NKI','NKM', 'NKS', 'NKX', 'NML', 'NMSV', 'NOTE', 'NPL', 'NRA', 'NRT', 'NSA', 'NTN', 'NVF', 'NWC', 'OBW', 'ODM', 'OFR', 'OGA', 'OGG', 'OKT', 'OMA', 'OMF', 'OMG', 'OMX', 'OPUS', 'OTS', 'OVE', 'OVW', 'OVW', 'PAC', 'PANDORA', 'PBF', 'PCA', 'PCAST', 'PCG', 'PEAK', 'PEK', 'PHO', 'PHY', 'PK', 'PKF', 'PLA', 'PLS', 'PLY', 'PNA', 'PNO', 'PPC', 'PPCX', 'PRG', 'PSF', 'PSF1', 'PSF2', 'PSM', 'PSY', 'PTCOP', 'PTF', 'PTM', 'PTS', 'PTT', 'PTX', 'PTXT', 'PVC', 'QCP', 'R1M', 'RA', 'RAM', 'RAW', 'RAX', 'RBS', 'RBS', 'RCY', 'REX', 'RFL', 'RGRP', 'RIP', 'RMI', 'RMJ', 'RMX', 'RNG', 'RNS', 'ROL', 'RSN', 'RSO', 'RTA', 'RTI', 'RTS', 'RVX', 'RX2', 'S3I', 'S3M', 'S3Z', 'SAF', 'SAP', 'SBG', 'SBI', 'SC2', 'SCS11', 'SD', 'SD', 'SD2', 'SD2F', 'SDAT', 'SDS', 'SDT', 'SEQ', 'SES', 'SESX', 'SF2', 'SFAP0', 'SFK', 'SFL', 'SFPACK', 'SFS', 'SFZ', 'SGP', 'SHN', 'SIB', 'SLP', 'SLX', 'SMA', 'SMF', 'SMP', 'SMP', 'SMPX', 'SND', 'SND', 'SND', 'SNG', 'SNG', 'SNS', 'SONG', 'SOU', 'SPPACK', 'SPRG', 'SPX', 'SSEQ', 'SSEQ', 'SSM', 'SSND', 'STAP', 'STM', 'STX', 'STY', 'SVD', 'SVX', 'SWA', 'SXT', 'SYH', 'SYN', 'SYW', 'SYX', 'TAK', 'TAK', 'TD0', 'TG', 'TOC', 'TRAK', 'TTA', 'TXW', 'U', 'UAX', 'ULT', 'UNI', 'USF', 'USFLIB', 'UST', 'UW', 'UWF', 'VAG', 'VAP', 'VB', 'VC3', 'VDJ', 'VGM', 'VGZ', 'VIP', 'VLC','VMD', 'VMF', 'VMF', 'VMO', 'VOC', 'VOX', 'VOXAL', 'VPL', 'VPM', 'VPW', 'VQF', 'VRF', 'VSQ', 'VSQX', 'VTX', 'VYF', 'W01', 'W64', 'WAV', 'WAVE', 'WAX', 'WEM', 'WFB', 'WFD', 'WFM', 'WFP', 'WMA', 'WOW', 'WPK', 'WPP', 'WPROJ', 'WRK', 'WTPL', 'WTPT', 'WUS', 'WUT', 'WV', 'WVC', 'WVE', 'WWU', 'XA', 'XA', 'XFS', 'XM', 'XMF', 'XMU', 'XRNS', 'XSP', 'XSPF', 'YOOKOO', 'ZPA', 'ZPL', 'ZVD']

videosformat=['AEP', 'PRPROJ', 'DREAM', 'SWF', 'MKV', 'MP4', 'TPD', 'WVM', 'AMC', 'BIK', 'SFD', 'PIV', 'MSWMM', 'WLMP', 'META', 'MPROJ', 'DCR', 'MXF', 'WEBM', 'DIR', 'VID', '3GP', 'SRT', 'VEG', '3GPP', 'AVI', 'VOB', 'WMV', 'FBR', 'GVI', 'VIV', 'SCC', 'DCR', 'REC', 'DPA', 'N3R', 'STR', 'AEC', 'SCM', 'VPJ', 'RMS', 'RMVB', 'WPL', 'M4V', 'PDS', 'TS', 'MVD', 'PHOTOSHOW', 'RDB', 'ZMV', 'IFO', 'BNP', 'DXR', 'MVP', 'MP4INFOVID', 'VP6', 'AMX', 'AVCHD', 'DZM', 'MPEG', 'RCD', 'TREC', 'VRO', 'DZT', 'MSDVD', 'FLV', 'WP3', 'ALE', 'M75', 'MPV2', 'PAC', 'PLAYLIST', 'RCUT', 'SWT', 'CAMPROJ', 'SWI', 'MOV', 'MPG', 'MOB', 'PSH', 'WM', 'WMX', 'XVID', 'BIN', 'CAMREC', 'MANI', 'TRP', '3GP2', 'VCPF', 'CPI', 'DAT', 'EXO', 'MTS', 'OGV', 'ASF', 'DAV', 'HDMOV', 'IVR', 'M2TS', 'VC1', 'VGZ', 'VP3', 'MMV', 'MNV', 'MP4V', 'USM', 'M2P', 'TVS', 'KMV', '60D', 'BU', 'VIDEO', 'AAF','AEPX', 'AET', 'AVV', 'AXV', 'BYU', 'CAMV', 'CREC', 'DB2', 'DMSD', 'EVO', 'FCP', 'GXF', 'IZZY', 'M1V', 'MJPG', 'MVEX', 'PSSD', 'QTM', 'RM', 'SAN', 'SBT', 'SFVIDCAP', 'TIX', 'VEM', 'VEP', 'VSE', 'WMMP', 'WOT', 'WVE', '264', 'SPL', 'MJP', '890', 'DV4', 'M2T', 'DPG', 'DZP', 'ISMV', 'OSP', 'RV', 'SBK', 'VSP', 'F4V', '3P2', 'GFP', 'JMV', 'MTV', 'R3D', 'SMV', 'H264', 'MPEG4', 'ISM', 'NVC', 'TSP', 'XML', 'BSF', 'CLPI', 'DMSM', 'DMX', 'DNC', 'DVDMEDIA', 'FLI', 'HDV', 'INT', 'JSS', 'M15', 'MJ2', 'NSV', 'OGM', 'QTL', 'RSX', 'RVL', 'SMK', 'SQZ', 'TP', 'WVX', 'XESC', '3G2', 'DVR', 'G64', 'MVP', 'LRV', 'BVR', 'DASH', 'DDAT', 'IRCP', 'IVF', 'MVE', 'PSB', 'RMP', 'ROQ', 'TDT', 'STX', '3MM', 'AETX', 'CINE', 'GL', 'WCP', 'ZM3', 'CED', 'GIFV', 'MP2V', 'MPE', 'WMD', 'BDMV', 'DIVX', 'LREC', 'LSX', 'SFERA', 'ARF', 'QT', 'MOI', 'PMF', 'AJP', 'PAR', 'YUV','AMV', 'AQT', 'ARCUT', 'ASX', 'AVB', 'AVD', 'AVP', 'AXM', 'BDT3', 'BMC', 'CMMTPL', 'CPVC', 'CVC', 'D2V','D3V', 'DV', 'EYETV', 'F4F', 'F4M', 'FLC', 'FLH', 'FPDX', 'G2M', 'GOM', 'GTS', 'HKM', 'IMOVIELIBRARY', 'IMOVIEMOBILE', 'INP', 'ISMC', 'IZZ', 'JTS', 'JTV', 'K3G', 'KDENLIVE', 'LVIX', 'M21', 'M21', 'MK3D', 'MOD', 'MP21', 'MPEG2', 'MPGINDEX', 'MPLS', 'MPV', 'MQV', 'MSE', 'MSH', 'MXV', 'MYS', 'NCOR', 'NUT', 'NUV', 'ORV', 'OTRKEY', 'PLPROJ', 'PPJ', 'PREL', 'PRTL', 'PXV', 'QTZ', 'RCREC', 'RUM', 'RVID', 'SBZ', 'SCREENFLOW', 'SDV', 'SEDPRJ', 'SEQ', 'SIV', 'SSM', 'TDA3MT', 'TIVO', 'TOD', 'TPR', 'TTXT', 'TVSHOW', 'USF', 'VBC', 'VCR', 'VCV', 'VDO', 'VDR', 'VFZ', 'VIVO', 'VLAB', 'VR', 'WFSP', 'WTV', 'WXP', 'XEL', 'XFL', 'XLMV', 'Y4M', 'ZEG', 'ZM1', 'ZM2', 'BDM', 'BMK', 'IMOVIEPROJ', 'THP', 'XMV', 'VTT', 'DVX', 'RMD', 'SSF', 'TP0', 'W32', 'AWLIVE', 'F4P', 'IRF', 'FCPROJECT', 'SVI', 'VIEWLET', 'DVR-MS', 'MOVIE', 'MPL', 'CLK', 'DMSM3D', 'GCS', 'M1PG', 'M2V', 'PROQC', 'SEC', 'SNAGPROJ', 'STL', 'VS4', 'NFV', 'SCM', '787', 'AVS', 'BIX', 'BOX', 'CMMP', 'CMREC', 'DCK', 'DMSS', 'DV-AVI', 'IMOVIEPROJECT', 'ISMCLIP', 'IVA', 'M2A', 'MGV', 'MOFF', 'MP21', 'MPG4', 'MPL', 'NTP', 'PGI', 'PRO', 'RMV', 'SCN', 'TVLAYER', 'WSVE', 'MODD', 'FFM', 'FTC', 'MVC', 'XEJ', 'CIP', 'EDL', 'EVO', 'IVS', 'M4E', 'MPEG1', 'SMI', 'VP7', 'AM', 'MOOV', 'RCPROJECT', 'TDX', 'YOG', 'MPG2', 'SMIL', '3GPP2', 'AVS', 'DCE', 'VF', 'AECAP', 'AEGRAPHIC', 'ANIM', 'ANX', 'BDT2', 'BS4', 'CMPROJ', 'CMV', 'CST', 'CX3', 'DLX', 'DMB', 'DMSD3D', 'FBR', 'FBZ', 'FFD', 'FLIC', 'FLX', 'JDR', 'KTN', 'M4U', 'MVB', 'MVY', 'OGX', 'PJS', 'PNS', 'PRO4DVD', 'PRO5DVD', 'QTCH', 'QTINDEX', 'RP', 'RTS', 'SMI', 'SML', 'TID', 'TVRECORDING', 'VIX', 'WGI', 'EYE', 'SEC', 'DIF', 'EZT', 'GVP', 'PVA', 'PVR', 'AVC', 'FCARCH', 'GRASP', 'MPSUB', 'THEATER', 'VFT', 'VFW', 'VMLT', 'MJPEG', 'VMLF', 'AVM', 'CEL', 'DSY', 'FVT', 'LSF', 'MPF', 'MT2S', 'PMV', 'RMD', 'RTS', 'V264', 'VDX']

docsformat=['PDF','PPTX','PPT','FODT', 'TXT', 'BIB', 'LST', 'DOCX', 'DOC', 'SUB', 'FBL', 'FCF', 'FDR', 'GV', 'NOW', 'SAVE', 'SMF', 'LST', 'MAN', 'STY', 'UPD', 'RTF', 'SAM', 'APKG', 'APT', 'DIZ', 'ODM', 'LOG', 'NFO', 'TEXT', 'DROPBOX', 'FPT', 'SAF', 'GPD', 'AIM', 'DOC', 'FDX', 'LATEX', 'UTF8', 'LTX', 'EIO', '1ST', 'DOCXML', 'KNT', 'MD5TXT', 'SIG', 'ERR', 'ASC', 'ME', 'WPD', 'FOUNTAIN', 'OPEICO', 'RTX', 'STRINGS', 'JARVIS', 'MBOX', 'OTT', 'VNT', 'ANS', 'LP2', 'RUN', 'TEX', 'MSG', 'ABW', 'AWW', 'FADEINTEMPLATE', 'LIS', 'LUF', 'PWD', 'RTFD', 'TEXTCLIPPING','WP', 'WTT', 'WPS', 'PAGES', 'RIS', 'WPS', 'BIB', 'BAD', 'EML', 'TLB', 'TMD', 'WPT', 'ODT', 'WRI', 'EIT', 'FRT', 'HHT', 'LUE', 'PRT', 'QDL', 'BIBTEX', 'DVI', 'EMLX', 'HS', 'IPSPOT', 'KES', 'KLG', 'RAD', 'README', 'PSW', 'KON', 'TAB', 'TEMPLATE', 'BOC', 'DSC', 'PWI', 'TAB', 'DOCM', 'EPP', 'GSD', 'PVM', 'XWP', 'RPT','SCC', 'ZRTF', 'LXFML', 'COD', 'P7S', 'RST', 'ADOC', 'ASC', 'ASCII', 'ASE', 'BDP', 'BDR', 'BEAN', 'BML', 'BNA', 'CHARSET', 'DFTI', 'DGS', 'DWD', 'DX', 'ETF', 'EUC', 'FAQ', 'FDT', 'FDXT', 'FLR', 'FWD', 'JP1', 'KLG', 'LBT', 'MELLEL', 'MNT', 'MW', 'MWD', 'MWP', 'OPENBSD', 'PAGES-TEF', 'PJT', 'PU', 'RVF', 'RZK', 'RZN', 'SAFETEXT', 'SCRIV', 'SCRIVX', 'SCT', 'SLAGZ', 'STORY', 'SUBLIME-PROJECT', 'SUBLIME-WORKSPACE', 'SXG', 'TPC','U3I', 'XYW', 'SSA', 'SXW', 'BTD', 'HBK', 'JNP', 'PRT', 'SCM', 'SLA', 'ACT', 'NDOC', 'OFL', 'EMULECOLLECTION', 'FDF', 'HWP', 'HZ', 'LYT', 'NOTES', 'TDF', 'XWP', 'IDX', 'ERR', 'WPL', 'XDL', 'DNE', 'DXB', 'ETX', 'JIS', 'RTD', 'SAM', 'SGM', 'SMS', 'TMV', 'UTXT', 'WP6', 'WPT', 'SE', 'CHORD', 'DOCZ', 'JOE', 'TRELBY', 'WP7', 'WPA', 'XBDOC', 'XY', 'LWP', 'CNM', 'ODO', 'DCA', 'EMF', 'NJX', 'CRWL', 'ATY', 'AWP', 'BBS', 'BRX', 'CALCA', 'CHART', 'CWS', 'CYI', 'DXP', 'FDS', 'FGS', 'GPN', 'GTHR', 'KWD', 'LNT', 'LYX', 'MELL', 'MIN', 'NGLOSS', 'NWM', 'NWP', 'PLANTUML', 'PMO', 'PVJ', 'PWDP', 'PWDPL', 'PWR', 'QPF', 'SCW', 'SDM', 'SDW', 'SESSION', 'SKCARD', 'STW', 'TDF', 'TM', 'UNAUTH', 'UNX', 'UOF', 'VCT', 'WBK', 'WEBDOC', 'WP4', 'WP5', 'WTX', 'XBPATE', 'XDL', 'XY3', 'XYP', 'ZABW', 'PFX', 'NB', 'WPD', 'ZW', 'HWP', 'WPD', 'BZABW', 'AWT', 'NWCTXT', 'OCR', 'ORT', 'PDPCMD', 'TVJ', 'UOT', 'XWP', 'GMD', 'DOX', 'FFT', 'FWDN', 'IIL', 'IPF', 'JRTF', 'LTR', 'MCW','ODIF', 'RFT', 'SDOC', 'THP', 'VW', 'WN', 'WPW', 'WSD']

imagesformat=['3FR', 'ARI', 'ARW', 'BAY', 'CR2', 'CRW', 'CS1', 'CXI', 'DCR', 'DNG', 'EIP', 'ERF', 'FFF', 'IIQ', 'J6I', 'K25', 'KDC', 'MEF', 'MFW', 'MOS', 'MRW', 'NEF', 'NRW', 'ORF', 'PEF', 'RAF', 'RAW', 'RW2', 'RWL', 'RWZ', 'SR2', 'SRF', 'SRW', 'X3F', 'PSD', 'JPG', 'PNG', 'JPEG', 'TGA', 'GIF', 'PISKEL', 'XCF', 'DDS', 'AWD', 'EXR', 'HDR', 'I3D', 'IWI', 'JBIG2', 'LIP', 'PICNC', 'PNS', 'PSPBRUSH', 'SPRITE2', 'TBN', 'WEBP', 'PCX', 'SPR', 'CT', 'PDN', 'AVATAR', 'PIC', 'PPF', 'TN3', 'ARR', 'PXD', 'DJVU', 'PAT', 'PWP', 'SAI', 'XPM', 'BMP', 'JXR', 'OTA', 'SUMO', 'UFO', 'PM', 'ITC2', 'TIFF', 'JPC', 'MSP', 'PI2', 'PXR', 'SPRITE', 'TN', 'JPS', 'APM', 'OZJ', 'WBZ', '8PBS', 'HEIF', 'JPE', 'PGM', 'RLI', 'SVM', 'VPE', 'VRIMG', 'TIF', 'ICN', 'PSP', 'FPX', 'MPF', 'PSB', 'SPP', 'TG4', 'INFO', '2BP', 'DJV', 'TPF', 'PPM', 'PSPIMAGE', 'SIG', 'MPO', '9PNG', 'BPG', 'CAN', 'DIB', 'ECW', 'FLIF', 'GRO', 'ICON', 'INT', 'JFI', 'LRPREVIEW', 'NCD', 'PAP', 'PDD', 'PFI', 'PGF', 'PMG', 'POV', 'PSE', 'PYXEL', 'RCL', 'RIF', 'TFC', 'TM2', 'WBC', 'ALBM', 'CPT', 'FIL', 'THM', 'THUMB', 'FAC', 'J2K', 'JPF', 'PICTCLIPPING', 'PJPG', 'DGT', 'PBM', 'PX', 'CLIP', 'FITS', 'HPI', 'JNG', 'KDK', 'PZS', 'QTIF', 'RSR', 'SLD', 'V', 'TIF', '001', 'DPX', 'KODAK', 'BMQ', 'CDG', 'DIC', 'GP4', 'ICA', 'ORA', 'OZT', 'PRW', 'SPA', 'SUP', 'WDP', 'ZIF', 'JPX', 'SID', 'ABM', 'RPF', 'SDR', 'RGB', 'MNG', 'CPG', 'PIC', 'PVR', 'ITHMB', 'THM', 'ACCOUNTPICTURE-MS', 'AFX', 'AGP', 'APD', 'ARW', 'AVB', 'BLKRT', 'BM2', 'BMC', 'BTI', 'CAL', 'CD5', 'CDC', 'CE', 'CID', 'CIT', 'COLZ', 'DCM', 'DMI', 'FACE', 'FBM', 'FPOS', 'GBR', 'GCDP', 'GIH', 'HDRP', 'HEIC', 'HF', 'J2C', 'JBF', 'JBR', 'JIA', 'JIF', 'JPG2', 'LB', 'LBM', 'LIF', 'LZP', 'MBM', 'MCS', 'MIC', 'MIX', 'MNR', 'MYL', 'NEO', 'OC3', 'OC4', 'OC5', 'OCI', 'OPLC', 'OTB', 'OZB', 'PC1', 'PC2', 'PE4', 'PIXADEX', 'PNM', 'PSDX', 'PTEX', 'PXM', 'PZA', 'PZP', 'QMG', 'QTI', 'S2MV', 'SAR', 'SGD', 'SIG', 'SIM', 'SKITCH', 'SKM', 'SMP', 'SPE', 'SPU', 'SR', 'T2B', 'TJP', 'TNY', 'TUB', 'USERTILE-MS', 'VRPHOTO', 'WB1', 'WB2', 'WBMP', 'WBP', 'WIC', 'WPB', 'XWD', 'CIN', 'DCX', 'APNG', 'PCD', 'JP2', '8CA', 'PICT', 'SFF', 'VFF', '360', 'ACORN', 'KIC', 'RGF', 'BLZ', 'DM3', 'DTW', 'GFIE', 'GIM', 'GMSPR', 'GPD', 'HDP', 'JTF', 'NOL', 'RCU', 'RRI', 'RS', 'SBP', 'SUN', 'VSS', 'WI', 'PANO', '411', 'MET', 'MSK', 'PTG', 'WBD', 'TEX', 'AGIF', 'APX', 'ART', 'AWD', 'BRK', 'CPD', 'CSF', 'EPP', 'IPX', 'JB2', 'JBIG', 'JFIF', 'KDI', 'MRB', 'MXI', 'RIC', 'RIX', 'SFC', 'SOB', 'TARGA', 'TPS', 'UGA', 'JIFF', 'MAX', 'DT2', 'MAT', 'MBM', 'MIFF', 'MRXS', 'PI1', 'PJPEG', 'PM3', 'PTK', 'QIF', 'VNA', 'SFW', 'BMX', 'FPPX', 'OMF', 'RSB', 'SKYPEEMOTICONSET', 'SVA', 'WMP', 'CAM', 'MAC', '73I', '8CI', '8XI', 'APS', 'BMZ', 'BRN', 'CALS', 'CMR', 'CPC', 'CPS', 'CPX', 'DC2', 'DDT', 'FAX', 'FPG', 'GGR', 'GMBCK', 'GRY', 'ICPR', 'ILBM', 'IMG', 'IPICK', 'JAS', 'JPD', 'JPG-LARGE', 'JWL', 'KPG', 'LJP', 'NCR', 'NCT', 'OTI', 'PAC', 'PC3', 'PE4', 'PFR', 'PIXELA', 'PNC', 'PNI', 'PNT', 'POP', 'PP4', 'PP5', 'PTS', 'PXICON', 'RAS', 'RIFF', 'RLE', 'RTL', 'RVG', 'SCT', 'SHG', 'SID', 'SPH', 'SPJ', 'TSR', 'U', 'UGOIRA', 'UPF', 'VDA', 'VIC', 'VICAR', 'VIFF', 'WB0', 'WBM', 'WVL', 'Y', 'YSP', 'ZVI', 'YUV', 'DVL', 'PSF','BSS', 'DICOM', 'JBG', 'SGI', 'VST', 'C4', 'AIC', 'AIS', 'BMF', 'BRT', 'BW', 'CIMG', 'CPBITMAP', 'ODI', 'PIX', 'PTX', 'PTX', 'SCN', 'WPE', 'XBM', 'IVR', 'J', 'MIP', 'PIC', 'STE', 'FRM', 'INK', 'ACR', 'ADC', 'ARTWORK', 'CUT', 'DDB', 'DRZ', 'FAL', 'G3', 'GFB', 'GROB', 'HR', 'HRF', 'IC1', 'IC2', 'IC3', 'ICB', 'IMJ', 'IPHOTOPROJECT', 'IVUE', 'JBMP', 'KFX', 'NLM', 'PAL', 'PI2', 'PI3', 'PI4', 'PI5', 'PI6', 'PIX', 'PNTG', 'POV', 'RGB', 'RGBA', 'SCG', 'SCI', 'SCP', 'SCU', 'SEP', 'SPC', 'SPIFF', 'SUNIFF', 'TAAC', 'TB0', 'TN1', 'TN2', 'TPI', 'TRIF', 'URT', 'ABC', 'AC5', 'AC6', 'AF2', 'AF3', 'AFDESIGN', 'AI', 'ART', 'ARTB', 'ASY', 'AWG', 'CAG', 'CCX', 'CDD', 'CDDZ', 'CDLX', 'CDMM', 'CDMT', 'CDMTZ', 'CDMZ', 'CDR', 'CDS', 'CDSX', 'CDT', 'CDTX', 'CDX', 'CDX', 'CGM', 'CIL', 'CLARIFY', 'CMX', 'CNV', 'COR', 'CSY', 'CV5', 'CVG', 'CVI', 'CVS', 'CVX', 'CWT', 'CXF', 'DCS', 'DDRW', 'DED', 'DESIGN', 'DHS', 'DIA', 'DPP', 'DPR', 'DPX', 'DRAWING', 'DRAWIT', 'DRW', 'DRW', 'DSF', 'DXB', 'EGC', 'EMF', 'EMZ', 'EP', 'EPS', 'EPSF', 'ESC', 'EZDRAW', 'FH10', 'FH11', 'FH3', 'FH4', 'FH5', 'FH6', 'FH7', 'FH8', 'FH9', 'FHD', 'FIF', 'FIG', 'FMV', 'FS', 'FT10', 'FT11', 'FT7', 'FT8', 'FT9', 'FTN', 'FXG', 'GDRAW', 'GEM', 'GKS', 'GLOX', 'GLS', 'GRAFFLE', 'GSD', 'GSTENCIL', 'GTEMPLATE', 'HGL', 'HPG', 'HPGL', 'HPL', 'IDEA', 'IGT', 'IGX', 'IMD', 'INK', 'INK', 'JSL', 'LMK', 'MGC', 'MGCB', 'MGMF', 'MGMT', 'MGMX', 'MGS', 'MGTX', 'MMAT', 'MP', 'MVG', 'NAP', 'ODG', 'OTG', 'OVP', 'OVR', 'PAT', 'PCS', 'PD', 'PEN', 'PFD', 'PFV', 'PL', 'PLT', 'PLT', 'PMG', 'POBJ', 'PS', 'PSID', 'PWS', 'RDL', 'SCV', 'SDA', 'SK1', 'SK2', 'SKETCH', 'SLDDRT', 'SMF', 'SNAGITSTAMPS', 'SNAGSTYLES', 'SSK', 'STD', 'STN', 'SVF', 'SVG', 'SVGZ', 'SXD', 'TLC', 'TNE', 'TPL', 'UFR', 'VBR', 'VEC', 'VML', 'VSD', 'VSDM', 'VSDX', 'VST', 'VSTM', 'VSTX', 'WMF', 'WMZ', 'WPG', 'WPI', 'XAR', 'XMIND', 'XMMAP', 'XPR', 'YAL', 'ZGM', '3D', '3D2', '3D4', '3DA', '3DC', '3DC', '3DF', '3DL', '3DM', '3DMF', '3DMK', '3DON', '3DP', '3DS', '3DV', '3DW', '3DX', '3DXML', '3MF', 'A2C', 'A3D', 'A8S', 'AC', 'ACT', 'ALBUM', 'AMF', 'AN8', 'ANIM', 'ANIM', 'ANIM', 'ANIMSET', 'ANIMSET\_INGAME', 'ANM', 'AOF', 'AOI', 'ASAT', 'ATF', 'ATL', 'ATM', 'B3D', 'BIO', 'BIP', 'BLD', 'BLEND', 'BR3', 'BR4', 'BR5','BR6', 'BR7', 'BRG', 'BRO', 'BSK', 'BTO', 'BVH', 'C3Z', 'C4D', 'CAF', 'CAL', 'CAL', 'CAS', 'CCB', 'CCP','CFG', 'CG', 'CG3', 'CGA', 'CGFX', 'CHR', 'CHR', 'CHRPARAMS', 'CM2', 'CMF', 'CMOD', 'CMZ', 'CPY', 'CR2', 'CRF', 'CRZ', 'CSD', 'CSF', 'CSM', 'CSO', 'D3D', 'DAE', 'DAZ', 'DBC', 'DBL', 'DBM', 'DBS', 'DDD', 'DES', 'DFF', 'DFS', 'DIF', 'DMC', 'DRF', 'DS', 'DSA', 'DSB', 'DSD', 'DSE', 'DSF', 'DSI', 'DSI', 'DSO', 'DSV', 'DUF', 'DWF', 'E57', 'EGG', 'EXP', 'F3D', 'FACEFX', 'FACEFX\_INGAME', 'FBX', 'FC2', 'FCP', 'FCZ', 'FG', 'FIG', 'FLT', 'FNC', 'FP', 'FP3', 'FPE', 'FPF', 'FPJ', 'FRY', 'FSH', 'FSQ', 'FUN', 'FX', 'FXA', 'FXL', 'FXM','FXS', 'FXT', 'GEO', 'GLF', 'GLM', 'GMF', 'GMMOD', 'GMT', 'GRN', 'HD2', 'HDZ', 'HIP', 'HIPNC', 'HLSL', 'HR2', 'HRZ', 'HXN', 'IFC', 'IGES', 'IGI', 'IGM', 'IGS', 'IK', 'IRR', 'IRRMESH', 'IV', 'IVE', 'J3O', 'JAS','KFM', 'KMC', 'KMCOBJ', 'KTZ', 'LDM', 'LLM', 'LND', 'LP', 'LPS', 'LT2', 'LTZ', 'LWO', 'LWS', 'LXF', 'LXO', 'M3', 'M3D', 'M3D', 'MA', 'MAKERBOT', 'MAT', 'MAX', 'MAXC', 'MB', 'MC5', 'MC6', 'MCZ', 'MD5ANIM', 'MD5CAMERA', 'MD5MESH', 'MDD', 'MDL', 'MDX', 'MEB', 'MESH', 'MESH', 'MGF', 'MIX', 'MNM', 'MOT', 'MP', 'MPJ', 'MQO', 'MRML', 'MS3D', 'MSH', 'MTL', 'MTX', 'MTZ', 'MU', 'MUD', 'MXM', 'MXS', 'N2', 'N3D', 'NFF', 'NIF', 'NM', 'NSBTA', 'OBJ', 'OBP', 'OBZ', 'OCT', 'OFF', 'OGF', 'OL', 'P21', 'P2Z', 'P3D', 'P3L', 'P3M', 'P3R', 'P4D', 'P5D', 'PAR', 'PAT', 'PGAL', 'PHY', 'PIGM', 'PIGS', 'PKG', 'PKG', 'PL0', 'PL1', 'PL2', 'PLY', 'PMD', 'PMD', 'PMX', 'PP2', 'PPZ', 'PRC', 'PREFAB', 'PRIMITIVES', 'PRIMITIVES\_PROCESSED', 'PRM', 'PRO', 'PRO', 'PRV' 'PSA', 'PSK', 'PZ2', 'PZ3', 'PZZ', 'QC', 'RAD', 'RAD', 'RAY', 'RCS', 'RDS', 'RFT', 'RIG', 'S', 'S3G', 'SC4MODEL', 'SDB', 'SESSION', 'SGN', 'SH3D', 'SH3F', 'SHP', 'SI', 'SKL', 'SKP', 'SM', 'SMD', 'SRF', 'STC','STEP', 'STO', 'STP', 'T3D', 'T3D', 'TCN', 'TDDD', 'TGO', 'THING', 'THL', 'TMD', 'TME', 'TMO', 'TPS', 'TRI', 'TRI', 'TRUCK', 'TS1', 'TVM', 'U3D', 'UMS', 'V3D', 'V3O', 'VAC', 'VEG', 'VERT', 'VISUAL', 'VISUAL\_PROCESSED', 'VMD', 'VMO', 'VOB', 'VOX', 'VP', 'VPD', 'VRL', 'VS', 'VSH', 'VSO', 'VTX', 'VUE', 'VVD', 'W3D', 'WFT', 'WRL', 'WRP', 'WRZ', 'X', 'X3D', 'X3G', 'XAF', 'XAF', 'XMF', 'XMM', 'XOF', 'XPR', 'XRF', 'XSF', 'XSI', 'XV0', 'YAODL', 'YDL', 'Z3D', 'ZT']

def get\_drives():

drives = []

bitmask = windll.kernel32.GetLogicalDrives()

for letter in ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j', 'k', 'l', 'm', 'n', 'o', 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x', 'y', 'z']:

if bitmask & 1:

drives.append(letter)

bitmask >>= 1

return drives

fname=[]

fext=[]

def weightmat(filea,fileb):

c=""

weight=0

for i in filea:

c+=i

if c in fileb:

weight+=1

else:

return weight

return weight

def grouping(files):

for file in files:

filedata=file.split('.')

name=filedata[:-1]

name=".".join(name)

ext=filedata[-1]

fname.append(name)

if ext.upper() in audiosformat:

music.append(file)

fext.append(0)

elif ext.upper() in videosformat:

videos.append(file)

fext.append(100)

elif ext.upper() in imagesformat:

images.append(file)

fext.append(200)

elif ext.upper() in docsformat:

docs.append(file)

fext.append(300)

if \_\_name\_\_ == '\_\_main\_\_':

before = set(get\_drives())

pause = input("Please insert the USB device, then press ENTER")

print ('Please wait...')

time.sleep(5)

after = set(get\_drives())

drives = after - before

delta = len(drives)

if (delta):

for drive in drives:

if os.system("cd " + drive + ":") == 0:

newly\_mounted = drive

print ("There were %d drives added: %s. Newly mounted drive letter is %s" % (delta, drives, newly\_mounted))

else:

print("Sorry, I couldn't find any newly mounted drives.")

grp=list(os.walk(newly\_mounted+':'))

files=grp[0][2]

folder=grp[0][1]

music=[]

videos=[]

docs=[]

images=[]

grouping(files)

costfname=[]

for i in range(len(fname)):

mxwt=0

for j in range(len(fname)):

if i!=j:

currwt=weightmat(fname[i],fname[j])

if mxwt<currwt:

mxwt=currwt

costfname.append(mxwt)

source=newly\_mounted+':'

#print(docs,'\n',videos,'\n',music,'\n',images)

try:

os.makedirs(source+'/music')

except:

print("Folder already exist")

try:

os.makedirs(source+'/videos')

except:

print("Folder already exist")

try:

os.makedirs(source+'/images')

except:

print("Folder already exist")

try:

os.makedirs(source+'/docs')

except:

print("Folder already exist")

dest1 = source+'/music'

dest2=source+'/images'

dest3=source+'/videos'

dest4=source+'/docs'

for f in images:

shutil.move(source+f, dest2)

for f in videos:

shutil.move(source+f, dest3)

for f in music:

shutil.move(source+f, dest1)

for f in docs:

shutil.move(source+f, dest4)

mxt=[]

clust=int(input("enter number of clusters u want"))

for i in range(len(fname)):

mxt.append([costfname[i],fext[i]])

kmeans = KMeans(n\_clusters=clust).fit(mxt)

labels = kmeans.labels\_

print(labels)

l=[[] for i in range(clust)]

for i in range(len(kmeans.labels\_)):

for j in range(clust):

if kmeans.labels\_[i]==j:

l[j].append(files[i])

for i in l:

print(i)

**Result:**

We got clusters of file as:

Cluster 1:

['03 - Woh Lamhe (Zeher) [DJMaza.Info].mp3', '04 - Saari Raat Teri Yaad (Footpath) [DJMaza.Info].mp3', '05 - Kaho Na Kaho (Murder) [DJMaza.Info].mp3', '08 - Hale Dil (Muder 2) [DJMaza.Info].mp3', '10 - Dil Ibaadat (Tum Mile) [DJMaza.Info].mp3', '15 - Kya (Crook) [DJMaza.Info].mp3', '16 - Woh Ajnabee (The Train) [DJMaza.Info].mp3', '17 - Judai (Jannat) [DJMaza.Info].mp3', '18 - Haan Tu Hain (Jannat) [DJMaza.Info].mp3', '20 - Teri Yaadon Mein (The Killer) [DJMaza.Info].mp3', '21 - Bharat Mata Ki Jai (Shanghai) [DJMaza.Info].mp3', '22 - Kaise Ye Raaz Hai (Raaz TMC) [DJMaza.Info].mp3', '25 - Babu Rao Mast Hai (OUATIM) [DJMaza.Info].mp3', '28 - Dil Deewana (Jawani Diwani) [DJMaza.Info].mp3', '29 - Dillagi Mein Jo Beet Jaaye (ABA) [DJMaza.Info].mp3', '30 - Deewana Kar Raha Hai (Raaz 3) [DJMaza.Info].mp3', '31 - Ishq Sufiyana (The Dirty Picture) [DJMaza.Info].mp3']

Cluster2:

['Screenshot (22).png', 'Screenshot (35).png', 'Screenshot (36).png', 'Screenshot (37).png', 'Screenshot (38).png', 'Screenshot (54).png', 'Screenshot (55).png', 'Screenshot (56).png', 'Screenshot (57).png', 'Screenshot (58).png', 'Screenshot (59).png', 'Screenshot (60).png', 'Screenshot (61).png', 'Screenshot (62).png', 'Screenshot (63).png', 'Screenshot (64).png', 'Screenshot (65).png', 'Screenshot (66).png', 'Screenshot (67).png', 'Screenshot (68).png', 'Screenshot (69).png', 'Screenshot (70).png', 'Screenshot (71).png', 'Screenshot (72).png', 'Screenshot (73).png', 'Screenshot (74).png', 'Screenshot (75).png', 'Screenshot (76).png', 'Screenshot (77).png', 'Screenshot (78).png', 'Screenshot (79).png', 'Screenshot (80).png', 'Screenshot (81).png']

Cluster3:

['ECE1005.slno.10.File.1.ppt', 'ECE1005.slno.10.File.2.ppt', 'ECE1005.slno.10.File.3.pptx', 'ECE1005.slno.11.File.1.pdf', 'ECE1005.slno.11.File.1.ppt', 'ECE1005.slno.11.File.2.pptx', 'ECE1005.slno.11.File.3.ppt']

Cluster4:

['ScreenCaptureProject1.mkv']

Cluster5:

['CSE4019.docx', 'CSE4019\_daily.docx']

Cluster6:

['ECE1005.slno.7.File.1.pdf', 'ECE1005.slno.9.File.1.pdf']

**Conclusion:**

1. We are able to find the clusters on the basis of file weight and extension weight
2. But we are not 100% sure that they are interrelated
3. We fail to distinguish if the match of two distinct file has same length.
4. Can’t decide the number of cluster required in this algorithm.

**Areas to be worked on in this project:**

Finding a better weight calculating function

Researching the relation between number of frequency of letter occurred in file name and file name.

**References:**

[2.3. Clustering — scikit-learn 0.19.1 documentation](http://scikit-learn.org/stable/modules/clustering.html#k-means) scikit-learn.org

<http://fileinfo.com>

[k-means clustering algorithm - Data Clustering Algorithms](https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm) sites.google.com

[Introduction to K-means Clustering](https://www.datascience.com/blog/k-means-clustering) www.datascience.com