Ingeniería de software

**Comparador de PDF**

Leonardo Miranda Rico

Jorge Emilio Gutiérrez Cortes

Luis Ángel Ventura Rivera

Línea horizontal

# Código:

from customtkinter import \*

from CTkMessagebox import \*

ventana = CTk()

ventana.geometry("500x300")

ventana.title("PdFindDir")

def read\_pdfinder(vpdfname, vdictname):

import re

import os

from pdfminer.high\_level import extract\_text

txtlinea =""

valclean=[]

var\_final\_file=[]

folder\_path = r"D:\jorge\Desktop"

if not os.path.exists(folder\_path):

os.makedirs(folder\_path)

pdftext = extract\_text(vpdfname)

archivo = extract\_text(vdictname)

if archivo:

archivo = archivo.upper()

if pdftext:

pdftext = pdftext.upper()

strippdf = pdftext.split('\n')

stripdic = archivo.strip()

dic = stripdic.splitlines()

lineas\_sin\_vacios = [linea for linea in dic if len(linea.strip()) > 0]

lineas\_sin\_vacios = [linea.rstrip() for linea in dic if len(linea.strip()) > 0]

for linea in strippdf:

if linea != "":

txtlinea= txtlinea + linea

else:

txtlinea =txtlinea + " "

val =txtlinea.split()

txtlinea = ""

for valor in val:

aux = re.sub(r"[^a-zA-Z0-9ÑñÁáÉéÍíÓóÚú]","",valor)

aux = aux.strip()

if aux != "":

valclean.append(re.sub(r"[^a-zA-Z0-9ÑñÁáÉéÍíÓóÚú]","",valor))

txtlinea = txtlinea + valor + " "

for linea\_arch in lineas\_sin\_vacios:

linea\_arch = linea\_arch.strip()

if linea\_arch.count(" ") >= 1:

if txtlinea.count(linea\_arch)>=1:

for i in range(txtlinea.count(linea\_arch)):

var\_final\_file.append(linea\_arch)

print (linea\_arch)

for valorf in valclean:

if linea\_arch.count(" ") >= 1:

"HOLA"

else:

if linea\_arch==valorf:

var\_final\_file.append(linea\_arch)

final\_file = open(os.path.join(folder\_path, "resultados.txt"), 'w')

final\_file.writelines("\n".join(var\_final\_file).strip())

final\_file.close()

def msgerror():

CTkMessagebox(title= "Error", message="Formato no valido", icon="cancel")

def msgnofile():

CTkMessagebox(title= "Error", message="Debes seleccionar 2 archivos", icon="cancel")

def buscarpdf():

global v\_file\_entrada\_pdf

v\_file\_entrada\_pdf = filedialog.askopenfilename()

if (v\_file\_entrada\_pdf[v\_file\_entrada\_pdf.find(".pdf"):len(v\_file\_entrada\_pdf)] == ".pdf") or (v\_file\_entrada\_pdf ==""):

print (v\_file\_entrada\_pdf)

else:

msgerror()

print (v\_file\_entrada\_pdf)

def buscardic():

global v\_file\_entrada\_pdf2

v\_file\_entrada\_pdf2 = filedialog.askopenfilename()

if (v\_file\_entrada\_pdf2[v\_file\_entrada\_pdf2.find(".pdf"):len(v\_file\_entrada\_pdf2)] == ".pdf") or (v\_file\_entrada\_pdf2 ==""):

print (v\_file\_entrada\_pdf2)

else:

msgerror()

def execdict():

try:

print(v\_file\_entrada\_pdf)

print(v\_file\_entrada\_pdf2)

except:

msgnofile()

if v\_file\_entrada\_pdf=="" or v\_file\_entrada\_pdf2=="":

msgnofile()

else:

print(v\_file\_entrada\_pdf)

print(v\_file\_entrada\_pdf2)

read\_pdfinder(v\_file\_entrada\_pdf, v\_file\_entrada\_pdf2)

btnpdf = CTkButton(master=ventana,text="Selecciona pdf",corner\_radius=32, fg\_color='#ff00ff', font=('console',35), command=buscarpdf)

btnpdf.place(relx=0.5,rely=0.5,anchor="center")

btntxt = CTkButton(master=ventana,text="Selecciona diccionario",corner\_radius=32, fg\_color='#ff00ff', font=('console',35), command=buscardic)

btntxt.place(relx=0.5,rely=0.7,anchor="center")

btnexe = CTkButton(master=ventana,text="Ejecutar",corner\_radius=32, fg\_color='#ff00ff', font=('console',35), command=execdict)

btnexe.place(relx=0.5,rely=0.9,anchor="center")

ventana.mainloop()