import os

import re

from langdetect import detect

# Conjunto de expresiones regulares para habilidades blandas en inglés para IT

soft\_skills\_regex\_en\_it = {

'communication': r'\b(communicat(e|ion)|presentation|persuasiv(e|o)|collaborat(e|ion|ions)|multidisciplinary)\b',

'teamwork': r'\b(teamwork|collaborat(e|ion|ions)|collaborate|multidisciplinary)\b',

'problem solving': r'\b(problem solv(e|ing)|analytic(al|s))\b',

'creativity': r'\b(creativ(e|ity)|innovat(e|ion|ive))\b',

'adaptability': r'\b(adaptabilit(y|ies)|flexibilit(y|ies)|learn(ing|s) new technolog(y|ies))\b',

'attention to detail': r'\b(attention to detail|detail|defect(s|ions))\b',

'leadership': r'\b(leadership|stakehold(er|ers) manag(e|ement))\b',

'analytical skills': r'\b(analytic(al|s) think(ing|s)|data analys(is|es))\b',

'project management': r'\b(project manag(e|ement)|scrum|agile)\b',

'customer service': r'\b(customer servic(e|e)|client interact(ion|ions))\b',

'time management': r'\b(effective time manag(e|ement))\b',

'autonomous work': r'\b(ability to work autonomously)\b',

'fast-paced environment': r'\b(fast-paced environment)\b'

}

# Conjunto de expresiones regulares para habilidades blandas en español para IT

soft\_skills\_regex\_es\_it = {

'comunicación': r'\b(comunicaci(ó|o)n|presentaci(ó|o)n|persuasiv(o|a)|colaboraci(ó|o)n|multidisciplinari(a|o|as|os)|comunicar resultad(os|as) de manera efectiva)\b',

'trabajo en equipo': r'\b(trabajo en equip(o|a)|colaboraci(ó|o)n|colaborar|multidisciplinari(a|o|as|os)|trabajar en equip(o|a))\b',

'resolución de problemas': r'\b(resoluci(ó|o)n de problem(a|as)|anal(í|i)tico)\b',

'creatividad': r'\b(creatividad|creativ(o|a|os|as)|innovador(a|as|es)|innovaci(ó|o)n)\b',

'adaptabilidad': r'\b(adaptabilidad|flexibilidad|aprend(e|izaj)e de nueva(s)? tecnolog(í|i)a(s)?|trabajar bajo presión)\b',

'atención al detalle': r'\b(atenci(ó|o)n al detalle|detalle|defecto(s)?)\b',

'liderazgo': r'\b(liderazgo|gesti(ó|o)n de stakehold(er|ers))\b',

'habilidades analíticas': r'\b(pensamient(o|a) anal(í|i)tico|análisis de dat(o|os))\b',

'gestión de proyectos': r'\b(gesti(ó|o)n de proyect(o|os)|scrum|agile)\b',

'atención al cliente': r'\b(atenci(ó|o)n al client(e|es)|interacci(ó|o)n con client(e|es))\b',

'ágil': r'\b(ágil)\b',

'colaborativo': r'\b(colaborativo)\b',

'resiliencia': r'\b(resiliencia)\b',

'empatía': r'\b(empat(í|i)a)\b',

'gestión efectiva del tiempo': r'\b(gestión efectiva del tiempo)\b',

'trabajar de manera autónoma': r'\b(habilidad para trabajar de manera autónoma)\b',

'entorno acelerado': r'\b(ambiente de trabajo acelerado)\b'

}

def label\_soft\_skills\_it(text, language):

labels = []

# Etiquetado basado en expresiones regulares

if language == 'en':

soft\_skills\_regex = soft\_skills\_regex\_en\_it

elif language == 'es':

soft\_skills\_regex = soft\_skills\_regex\_es\_it

else:

raise ValueError("Idioma no admitido")

for skill, regex\_pattern in soft\_skills\_regex.items():

if re.search(regex\_pattern, text, flags=re.IGNORECASE):

labels.append(skill)

return labels

# Ejemplo de uso con la postulación proporcionada

text\_es = "Como Data Scientist en nuestro equipo, participarás en la identificación de patrones y tendencias a partir de grandes conjuntos de datos. Se valora la capacidad de presentar resultados de manera clara y persuasiva."

# Identificar habilidades blandas

language = detect(text\_es)

skills = label\_soft\_skills\_it(text\_es, language)

# Imprimir resultados

print(f"Idioma: {language}")

print(f"Habilidades blandas identificadas: {skills}")

# Carpeta que contiene archivos de postulaciones

folder\_path = "Applications"

postulation\_data = []

for i in range(1, 61):

file\_number = str(i).zfill(2)

file\_path = os.path.join(folder\_path, f"App\_{file\_number}.txt")

with open(file\_path, 'r', encoding='utf-8') as file:

text = file.read()

# Identificar habilidades blandas

language = detect(text)

skills = label\_soft\_skills\_it(text, language)

# Almacenar datos de postulación y habilidades blandas

postulation\_data.append({

'file\_number': file\_number,

'language': language,

'text': text,

'soft\_skills': skills

})

# Imprimir resultados o realizar otras acciones según sea necesario

for data in postulation\_data:

print(f"\nPostulación {data['file\_number']}:")

print(f"Idioma: {data['language']}")

print(f"Habilidades blandas identificadas: {data['soft\_skills']}")