Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра інформаційних систем і технологій

Лабораторна робота №1

З дисципліни «Обробка та аналіз текстових даних на Python»

Варіант 13

Виконав студент гр. ІС-11

Бондарчук О. А.

Перевірила ст. в. каф. ІСТ

Тимофєєва Ю. С.

Київ 2024

Завдання: Знайти всі дати та перевести їх у тексті в один формат.  
Початковий текст:

Slowly and solemnly he was borne into Briony LODgE 145a and laid out in the  
principal room, while I (094)-44-33 still observed the proceedings from my post  
by the window 12/15/1892. The lamps had been lit, but the blinds had not been  
drawn, so that I could see Holmes sh\_holmes221@mail.uk as he lay upon the couch.... I do not  
know whether he was seized with compunction at that moment for the part  
he was playing, but I know that I never felt more heartily ashamed of  
myself in my life 34-34-32 than when I saw the beautiful creature against whom  
I wston23@mymail.com was conspiring, or the grace and kindliness with which she waited  
upon the injured man. And yet it would be the blackest treachery to 1895-11-17  
Holmes to draw back now from the part which he had intrusted to me AveNUe 123.  
I hardened my heart Wtason Street 45b, and took the smoke-rocket from under my ulster.  
After all,, 344-21-01 I thought, we are not injuring her. We are but preventing  
her from injuring another adlerrr@mail.com.

Лістинг програми

import re  
from datetime import datetime  
  
def read\_file\_to\_string(fileName: str):  
 try:  
 with open(fileName, mode='r') as file:  
 content = file.read()  
 return content  
 except FileNotFoundError:  
 print(f"Error: File '{fileName}' not found.")  
 return None  
 finally:  
 file.close()  
  
  
filename: str = "text3.txt"  
  
file\_content = read\_file\_to\_string(filename)  
  
text = file\_content[100:302]  
print("Part of the text, received from file:\n")  
print("'" + text + "'\n")  
  
print("Text properties:")  
print("length: ", len(text))  
print("Count of 't' letters: ", text.count('t'))  
print("Index of 's' letter: ", text.find('s'))  
print("Index where 'seed' substring starts: ", text.index("ceed"))  
print("\nCapitalized text: ", text.capitalize())  
print("\nText in upper case: ", text.upper())  
print("\nText in lower case: ", text.lower())  
print("\nTitle text: ", text.title())  
print("\nJoined text: ", "-".join(text))  
print("Other methods:\n")  
print("Text is alpha: ", text.isalpha())  
print("Text is digit: ", text.isdigit())  
print("Text[1:2] is alphanumeric: ", text[1:3].isalnum())  
print("Text is lower: ", text.islower())  
print("Text is upper: ", text.isupper())  
print("Text is title: ", text.istitle())  
print("Text is space: ", text.isspace())  
print("Text starts with '(': ", text.startswith("("))  
print("Text ends with ' ': ", text.endswith(" "))  
  
print("\nText split by ' ': ", text.split(" "))  
print("\nText replace ' ': ", text.replace(" ", "\_"))  
  
date\_pattern = re.compile(r'\b((0?[1-9]|1[0-2])/(0?[1-9]|[12]\d|3[01])/\d{4}|\d{4}-\d{2}-\d{2})\b')  
  
# Знаходимо всі дати у тексті  
dates = date\_pattern.findall(file\_content)  
print("\nDates in the text:", [date[0] for date in dates])  
  
formatted\_dates = []  
for date in dates:  
 # Використовуємо datetime для переведення у зазначений формат  
 date\_obj = datetime.strptime(date[0], "%m/%d/%Y" if '/' in date[0] else "%Y-%m-%d")  
 formatted\_date = date\_obj.strftime("%Y-%m-%d")  
 formatted\_dates.append((date[0], formatted\_date))  
  
# Замінюємо всі дати у тексті на переведені у формат "%Y-%m-%d"  
for old\_date, new\_date in formatted\_dates:  
 file\_content = file\_content.replace(old\_date, new\_date)  
  
print("\nText with replaced dates\n\n" + file\_content)

Результати виконання програми:

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис

Зображення, що містить текст, знімок екрана, Шрифт

Автоматично згенерований опис  
  
Посилання на Github:  
<https://github.com/SashaBondarchuk/text-data-analysis-labs/tree/main>