**Міністерство освіти і науки України**

**Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського"**

**Факультет інформатики та обчислювальної техніки**

**Кафедра інформатики та програмної інженерії**

**Звіт**

з лабораторної роботи №1 з дисципліни

«Обробка та аналіз текстових даних на Python»

„**Робота з текстовими даними в Python**”

**Виконав(ла)**

(шифр, прізвище, ім'я, по батькові)

*ІП-11 Головня Олександр Ростиславович*

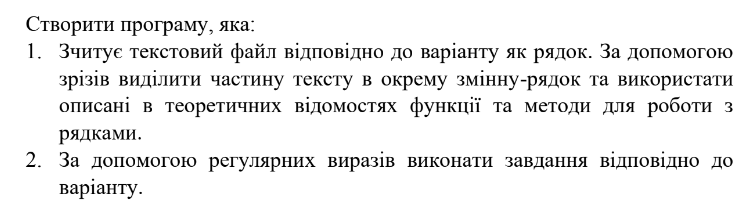
**Перевірив**

(прізвище, ім'я, по батькові)

Київ 2024

**Мета роботи:** Ознайомитись з представленням тексту Python в та

регулярними виразами.



**Варіант 2. Знайти та вивести всі адреси електронної пошти. Замінити у тексті символи у цих адресах на символ «-».**

**Результати виконання коду:**

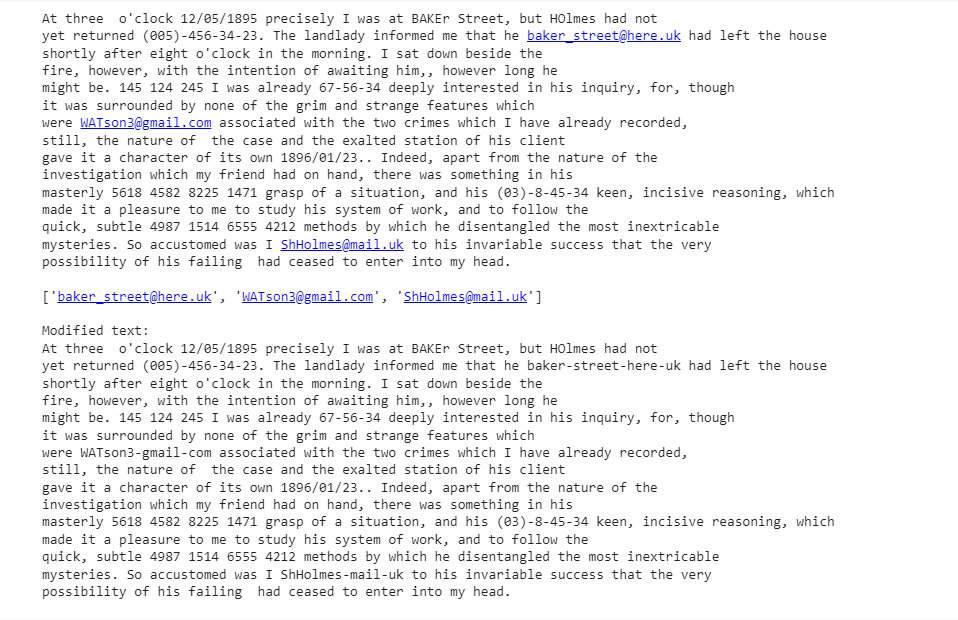


Рисунок 1. – Демонстрація виводу тексту, знаходження скриньок, заміна символів

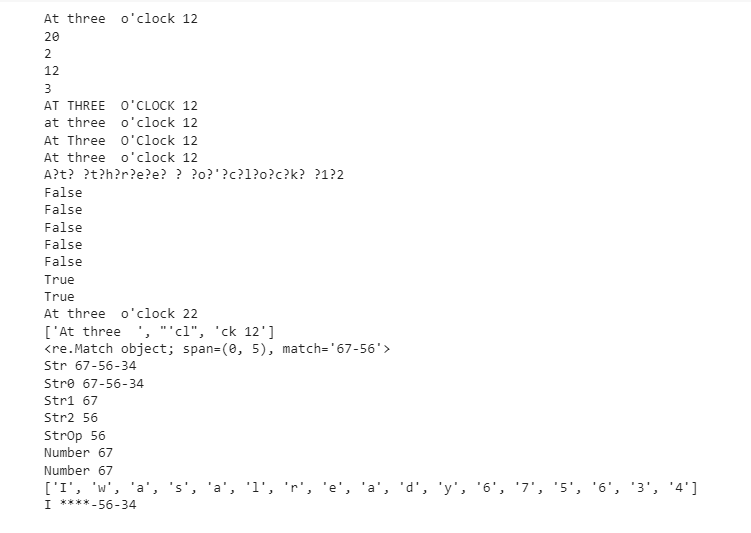


Рисунок 2. – Демонстрація описаних в теоритичних відомостях функції та методи для роботи з рядками

**Код програми:**

import re

# Зчитування тексту з файлу

with open('text1.txt', 'r', encoding='utf-8') as file:

    text = file.read()

print(text)

print()

def find\_emails(text):

    # Регулярний вираз для пошуку електронних адрес

    pattern = r'\b[A-Za-z0-9.\_%+-]+@[A-Za-z0-9.-]+\.[A-Z|a-z]{2,}\b'

    # Знаходимо всі входження

    emails = re.findall(pattern, text)

    return emails

emails\_from\_text = find\_emails(text)

print(emails\_from\_text)

def replace\_non\_letter\_chars(text):

    # Регулярний вираз для знаходження електронних адрес

    pattern = r'\b[A-Za-z0-9.\_%+-]+@[A-Za-z0-9.-]+\.[A-Z|a-z]{2,}\b'

    # Функція для заміни символів у знайдених електронних адресах на "-"

    def replace\_chars(match):

        return re.sub(r'[^A-Za-z0-9]', '-', match.group(0))

    # Заміна символів у знайдених електронних адресах

    modified\_text = re.sub(pattern, replace\_chars, text)

    return modified\_text

modified\_text = replace\_non\_letter\_chars(text)

print("\n"+"Modified text:" +"\n"+ modified\_text)

substring = text[0:20]

print(substring)

print(len(substring))

print(substring.count("t"))

print(substring.find("c"))

print(substring.index("three"))

print(substring.upper())

print(substring.lower())

print(substring.title())

print(substring.capitalize())

print("?".join(substring))

print(substring.isalnum())

print(substring.isalpha())

print(substring.isdigit())

print(substring.istitle())

print(substring.isspace())

print(substring.endswith('2'))

print(substring.startswith('A'))

print(substring.replace("12", "22"))

print(substring.split("o"))

import re

# Пошук за шаблоном

n = re.compile(r'(\d\d-)?\d\d')

substr= n.search('67-56-34')

print(substr)

n = re.compile(r'(\d\d)-(\d\d)-(\d\d)')

substr = n.search('I was already 67-56-34')

print('Str ' + substr.group())

print('Str0 ' + substr.group(0) )

print('Str1 ' +substr.group(1) )

print('Str2 ' +substr.group(2) )

n = re.compile(r'(\d\d)?\d\d')

substr = n.search('I was already 6-56-34')

print('StrOp ' + substr.group())

n = re.compile(r'(\d){2}')

substr = n.search('I was already 67-56-34')

print('Number ' + substr.group())

n = re.compile(r'(\d){1,4}')

substr = n.search('I was already 67-56-34')

print('Number ' + substr.group())

n = re.compile(r'[a-zA-Z0-9]')

substr = n.findall('I was already 67-56-34')

print(substr)

n = re.compile(r'was (\w)\w\*')

# Заміна шаблону у рядку

substr = n.sub(r'\*\*\*\*', "I was already67-56-34")

print(substr)

**Висновок:** Під час розробки лабораторної роботи, я ознайомився з представленням тексту Python в та регулярними виразами. Використав та показав їх роботу та властивості на практиці.