

Національний технічний університет України «КПІ ім. Ігоря Сікорського»  
Факультет Інформатики та Обчислювальної Техніки  
Кафедра інформаційних систем та технологій

**Лабораторна робота № 1**

з дисципліни «Обробка та аналіз текстових даних на Python»

На тему:

«Робота з текстовими даними в Python»

Варіант №3

Виконала:

студентка групи ІС-12.

Мельникова К.О.

Перевірила:

Тимофєєва Ю. С.

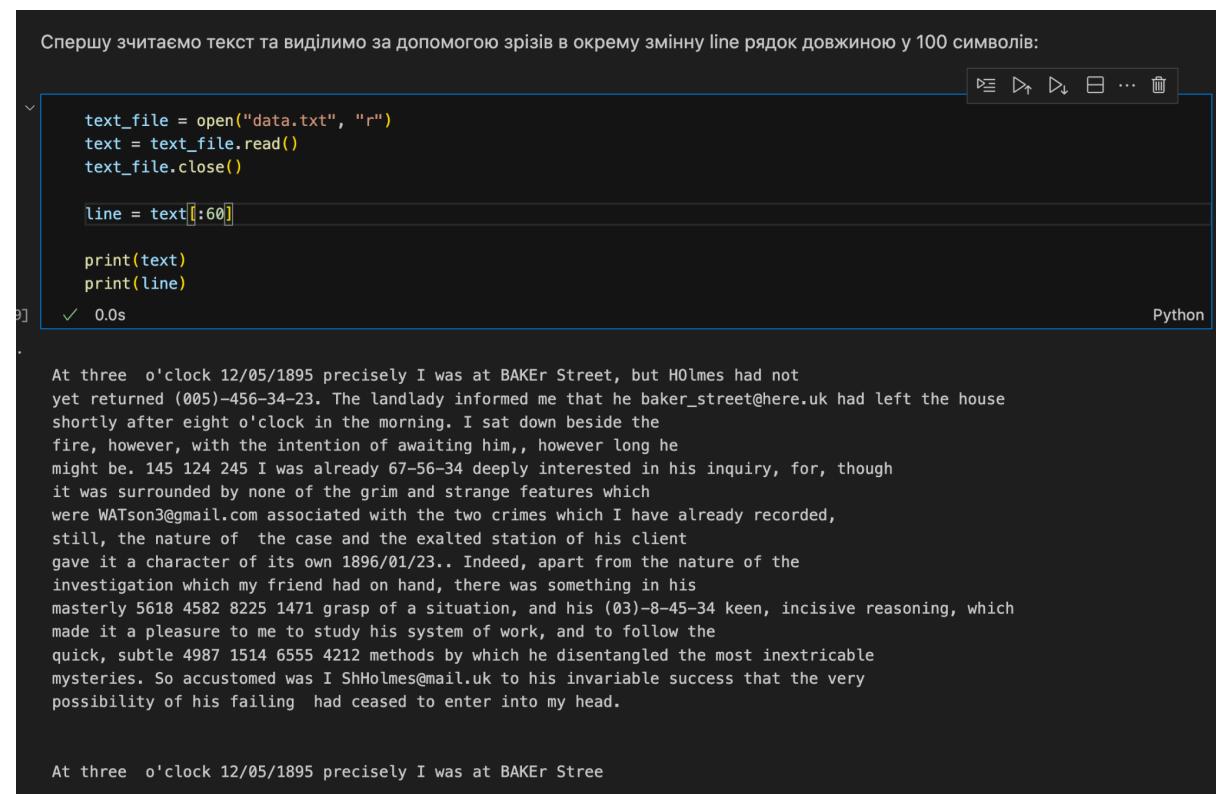
**Мета роботи:** Ознайомитись з представленням тексту Python в та регулярними виразами.

## Завдання до лабораторної роботи

Створити програму, яка:

*1. Зчитує текстовий файл відповідно до варіанту як рядок. За допомогою зрізів виділити частину тексту в окрему змінну-рядок та використати описані в теоретичних відомостях функції та методи для роботи з рядками.*

Спершу зчитуємо текст та виділимо за допомогою зрізів в окрему змінну line рядок довжиною у 60 символів:



Спершу зчитуємо текст та виділимо за допомогою зрізів в окрему змінну line рядок довжиною у 100 символів:

```
text_file = open("data.txt", "r")
text = text_file.read()
text_file.close()

line = text[:60]

print(text)
print(line)
```

✓ 0.0s Python

At three o'clock 12/05/1895 precisely I was at BAKER Street, but HOLMES had not yet returned (005)-456-34-23. The landlady informed me that he baker\_street@here.uk had left the house shortly after eight o'clock in the morning. I sat down beside the fire, however, with the intention of awaiting him,, however long he might be. 145 124 245 I was already 67-56-34 deeply interested in his inquiry, for, though it was surrounded by none of the grim and strange features which were WATson3@gmail.com associated with the two crimes which I have already recorded, still, the nature of the case and the exalted station of his client gave it a character of its own 1896/01/23.. Indeed, apart from the nature of the investigation which my friend had on hand, there was something in his masterly 5618 4582 8225 1471 grasp of a situation, and his (03)-8-45-34 keen, incisive reasoning, which made it a pleasure to me to study his system of work, and to follow the quick, subtle 4987 1514 6555 4212 methods by which he disentangled the most inextricable mysteries. So accustomed was I SHolmes@mail.uk to his invariable success that the very possibility of his failing had ceased to enter into my head.

At three o'clock 12/05/1895 precisely I was at BAKER Stree

Тепер використаємо функції для обробки тексту перераховані у звіті:

```
# Зрізи
print("Перший символ: " + line[0])
print("Елементи з початку до четвертого: " + line[:4])
print("Кожен другий елемент: " + line[::2])
print("Останній елемент: " + line[-1])

print("Кількість літер t: " + str(line.count("t"))) #кількість літер t
print("Індекс літери s: " + str(line.find("s"))) #індекс літери s
print("Індекс, з якого починається three: " + str(line.index("three"))) #індекс, з якого починається three
print("Upper case: " + line.upper())
print("Lower case: " + line.lower())
print("Всі перші літери слів upper case: " + line.title())
print("Перша літера рядка upper case: " + line.capitalize())

print("Метод Join: " + "?".join(line))

print("Чи містить рядок лише букви або цифри: " + str(line.isalnum()))
print("Чи містить рядок лише букви: " + str(line.isalpha()))
print("Чи містить рядок лише цифри: " + str(line.isdigit()))
print("Чи має рядок першу літеру кожного слова у верхньому регістрі: " + str(line.istitle()))
print("Чи є третій символ рядка пробілом: " + str(line[2].isspace()))
print("Чи закінчується рядок на заданий символ '!': " + str(line.endswith('!')))
print("Чи починається рядок з заданого символу 'A' " + str(line.startswith('A')))

print("Замінімо слово 'three' на 'two': ", line.replace("three", "two"))

print("Розіб'ємо рядок на слові 'precisely': ", text.split("precisely"))
```

```
Перший символ: A
Елементи з початку до четвертого: At t
Кожен другий елемент: A he osok1/519 rcsl a tBKrSre
Останній елемент: t
Кількість літер t: 5
Індекс літери s: 34
Індекс, з якого починається three: 3
Upper case: AT THREE O'CLOCK 12/05/1895 PRECISELY I WAS AT BAKER STREET
Lower case: at three o'clock 12/05/1895 precisely i was at baker street
Всі перші літери слів upper case: At Three O'Clock 12/05/1895 Precisely I Was At Baker Street
Перша літера рядка upper case: At three o'clock 12/05/1895 precisely i was at baker street
Метод Join: A?t? ?t?h?r?e?e? ? ?o?'?c?l?o?c?k? ?1?2?/?0?5?/?1?8?9?5? ?p?r?e?c?i?s?e?l?y? ?I? ?w?a?s? ?a?t? ?B?A?K?E?r? ?S?t?
Чи містить рядок лише букви або цифри: False
Чи містить рядок лише букви: False
Чи містить рядок лише цифри: False
Чи має рядок першу літеру кожного слова у верхньому регістрі: False
Чи є третій символ рядка пробілом: True
Чи закінчується рядок на заданий символ '!': False
Чи починається рядок з заданого символу 'A': True
Замінімо слово 'three' на 'two': At two o'clock 12/05/1895 precisely I was at BAKEr Street
Розіб'ємо рядок на слові 'precisely': ["At three o'clock 12/05/1895 ", ' I was at BAKEr Street']
```

2. За допомогою регулярних виразів виконати завдання відповідно до варіанту.

Використаємо регулярні вирази для пошуку дат та зведення їх до одного формату:

```
import re
from datetime import datetime

n = re.compile(r'(\d{1,4})[-./](\d{1,2})[-./](\d{1,4})')

dates = n.findall(text)

formatted_dates = []
for date in dates:
    day, month, year = map(int, date)
    if(month>12):
        day, month = month, day
    if(day > 31):
        day, year = year, day
    if(day>31 or month>12):
        continue
    formatted_dates.append(datetime(year, month, day).strftime("%Y-%m-%d"))

# Виводимо перетворені дати
print("Знайдені та переведені дати:")
for date in formatted_dates:
    print(date)
```

[79] ✓ 0.0s Python

... Знайдені та переведені дати:  
1895-05-12  
1896-01-23