



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

Лабораторна робота №1  
з дисципліни  
Обробка та аналіз текстових даних на Python

Виконав

студент групи ІП-11:  
Сідак К. І.

Перевірила:

ст. викладач  
Тимофєєва Ю. С.

Київ 2023

```
In [1]: import re
```

## Зчитування текстового файлу як рядок

```
In [2]: with open('text2.txt') as file:  
        text = file.read()  
        text
```

```
Out[2]: 'It was a 18:15 when we left Baker Street 221b, and it still\nwanted ten  
minutes to the hour 6-50pm when we found ourselves in SERPENTine\nAvenue  
@mail.com. It was already dusk,, and the lamps were just being lighted a  
s\nwe paced up and down in front of BrionY Lodge, waiting for the coming  
\nof its (343)-45-23 occupant. The house 345-A was just such as I had pi  
ctured it from\nSherlock Holmes's succinct description, but the localit  
y appeared to\nbe less private that I expected 456-45-34. On the cont  
rary, for a small Street 23 \nin a quiet neighborhood here1232@emm.uk, i  
t was remarkably animated. There was a\ngroup of shabbily-dressed men  
smoking and laughing in a corner, a\nscissors-grinder +308-45-32-21 with  
his wheel, two guardsmen who were flirting with a\nnurse-girl, and sever  
al well-dressed young men who were lounging up and\ndown with cigars in  
their mouths j.watson@gmail.com!!!'
```

## Виділення частини тексту (перше речення) в окрему змінну-рядок за допомогою зрізу

```
In [3]: text_slice = text[:152]  
        text_slice
```

```
Out[3]: 'It was a 18:15 when we left Baker Street 221b, and it still\nwanted ten  
minutes to the hour 6-50pm when we found ourselves in SERPENTine\nAvenue  
@mail.com.'
```

## Довжина рядка

```
In [4]: len(text_slice)
```

```
Out[4]: 152
```

## Кількість символів '1' у рядку

```
In [5]: text_slice.count('1')
```

```
Out[5]: 3
```

## Індекс першого символу 'a' в рядку

```
In [6]: text_slice.find('a')
```

```
Out[6]: 4
```

## Індекс, з якого починається підрядок when

```
In [7]: text_slice.index('when')
```

```
Out[7]: 15
```

## Весь рядок у верхньому регістрі

```
In [8]: text_slice.upper()
```

```
Out[8]: 'IT WAS A 18:15 WHEN WE LEFT BAKER STREET 221B, AND IT STILL\nWANTED TEN\nMINUTES TO THE HOUR 6-50PM WHEN WE FOUND OURSELVES IN SERPENTINE\nAVENUE\n@MAIL.COM.'
```

## Весь рядок у нижньому регістрі

```
In [9]: text_slice.lower()
```

```
Out[9]: 'it was a 18:15 when we left baker street 221b, and it still\nwanted ten\nminutes to the hour 6-50pm when we found ourselves in serpentine\navenue\n@mail.com.'
```

## Весь рядок у тайтл кейсі (кожне слово з великої літери)

```
In [10]: text_slice.title()
```

```
Out[10]: 'It Was A 18:15 When We Left Baker Street 221B, And It Still\nWanted Ten\nMinutes To The Hour 6-50Pm When We Found Ourselves In Serpentine\nAvenue\n@Mail.Com.'
```

## Рядок, де перша літера велика, а решта рядку в нижньому регістрі

```
In [11]: text_slice.capitalize()
```

```
Out[11]: 'It was a 18:15 when we left baker street 221b, and it still\nwanted ten\nminutes to the hour 6-50pm when we found ourselves in serpentine\navenue\n@mail.com.'
```

## Об'єднання усіх символів рядку, використовуючи двокрапку в якості роздільника

```
In [12]: ':'.join(text_slice)
```

```
Out[12]: 'I:t: :w:a:s: :a: :1:8:::1:5: :w:h:e:n: :w:e: :l:e:f:t: :B:a:k:e:r: :S:\nT:r:e:e:t: :2:2:1:b:, : :a:n:d: :i:t: :s:t:i:l:l:\n:w:a:n:t:e:d: :t:e:n:\n:m:i:n:u:t:e:s: :t:o: :t:h:e: :h:o:u:r: :6:-:5:0:p:m: :w:h:e:n: :w:e: :'
```

```
f:o:u:n:d: :o:u:r:s:e:l:v:e:s: :i:n: :S:E:R:P:E:N:t:i:n:e:\n:A:v:e:n:u:
e:@:m:a:i:l:..c:o:m:.'
```

## Перевірка чи містить рядок лише літери та цифри

```
In [13]: text_slice.isalnum()
```

```
Out[13]: False
```

## Перевірка чи містить рядок лише літери

```
In [14]: text_slice.isalpha()
```

```
Out[14]: False
```

## Перевірка чи містить рядок лише цифри

```
In [15]: text_slice.isdigit()
```

```
Out[15]: False
```

## Перевірка чи рядок у тайтл кейсі

```
In [16]: text_slice.istitle()
```

```
Out[16]: False
```

## Перевірка чи містить рядок лише пробіли

```
In [17]: text_slice.isspace()
```

```
Out[17]: False
```

## Перевірка чи закінчується рядок крапкою

```
In [18]: text_slice.endswith('.')
```

```
Out[18]: True
```

## Перевірка чи починається рядок з 'It'

```
In [19]: text_slice.startswith('It')
```

```
Out[19]: True
```

## Заміна всіх рядків 'mail' на 'yahoo' у даному рядку

```
In [20]: text_slice.replace('mail', 'yahoo')
```

```
Out[20]: 'It was a 18:15 when we left Baker SStreet 221b, and it still\nwanted ten minutes to the hour 6-50pm when we found ourselves in SERPENTine\nAvenue @yahoo.com.'
```

## Розбиття рядку на символах \n \r \t \f та пробілах

```
In [21]: text_slice.split()
```

```
Out[21]: ['It',  
          'was',  
          'a',  
          '18:15',  
          'when',  
          'we',  
          'left',  
          'Baker',  
          'SStreet',  
          '221b,',  
          'and',  
          'it',  
          'still',  
          'wanted',  
          'ten',  
          'minutes',  
          'to',  
          'the',  
          'hour',  
          '6-50pm',  
          'when',  
          'we',  
          'found',  
          'ourselves',  
          'in',  
          'SERPENTine',  
          'Avenue@mail.com.']
```

## Створення регулярного виразу для пошуку усіх номерів телефону в тексті

```
In [22]: phones_pattern = re.compile(r'(?:(\d{3})|(\d{3}))-\d{2}-\d{2}(?:-\d{2})?'  
phone_numbers = phones_pattern.findall(text)  
phone_numbers
```

```
Out[22]: ['(343)-45-23', '456-45-34', '308-45-32-21']
```

## Заміна у тексті зірочками всіх цифр номерів телефонів, крім останньої

```
In [23]: for number in phone_numbers:
          masked_number = re.sub(r'\d', '*', number[:-1]) + number[-1]
          text = text.replace(number, masked_number)
```

text

```
Out[23]: 'It was a 18:15 when we left Baker Street 221b, and it still\nwanted ten
minutes to the hour 6-50pm when we found ourselves in SERPentine\nAvenue
@mail.com. It was already dusk,, and the lamps were just being lighted a
s\nwe paced up and down in front of BrionY Lodge, waiting for the coming
\nof its (***)-**-*3 occupant. The house 345-A was just such as I had pi
ctured it from\nSherlock Holmes's succinct description, but the localit
y appeared to\nbe less private that I expected ***-**-*4. On the cont
rary, for a small Street 23 \nin a quiet neighborhood here1232@emm.uk, i
t was remarkably animated. There was a\ngroup of shabbily-dressed men
smoking and laughing in a corner, a\nscissors-grinder +***-**-**-*1 with
his wheel, two guardsmen who were flirting with a\nnurse-girl, and sever
al well-dressed young men who were lounging up and\ndown with cigars in
their mouths j.watson@gmail.com!!!'
```