МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М.Є. Жуковського “ХАІ”

кафедра 706

Лабораторна робота № 3

«Обработка текстовых данных. Работа со словарем»

ХАІ.706.733л(1).15О.020303.147л02.Ф

Виконав студент гр.733л(1) Бунін В.В.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_(№ групи) (П.І.Б)

(підпис, дата)

Перевірив:ст. преподаватель Лучшева О.В.

(П.І.Б)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис, дата)

2016

**Постановка задачи**

Для нормализованного текста, представленного в формате txt, выполнить обработку на языке Python согласно варианту №12:[Лексика по теме "Профессии"](http://study-english.info/vocabulary-professions.php):

1. Создать словарь, состоящий из **двадцати** и более слов (способ создания студент выбирает на свое усмотрение). При составлении словаря можно воспользоваться электронным ресурсом <http://study-english.info/vocabularies.php>.
2. Выполнить обработку исходного текста (см. задание 1) на согласно варианту №12: Найти частоту встречаемости личных местоимений единственного числа объектного падежа (***me, him, her, it***).

Вывести все слова текста в алфавитном порядке. При разработке интерфейса программы выполнить вывод на экран и в файл (с соответствующими пояснениями) обрабатываемого текста и результатов вычислений (основных и промежуточных). При написании программы обработки текста использовать функции пользователя (по желанию). Иметь возможность добавления и удаления слов.

**Алгоритм решения задачи**

* открыть файл для чтения со словами для словаря;
* открыть файл для чтения с текстом;
* открыть файл для записи результатов;
* прочесть данные из файла и вывести текст на экран;
* занести текст из файла в список
* посчитать количество местоимений me, him, her, it в списке выше
* посчитать количество переведенных слов
* рассортировать список выше по алфавиту
* добавить ввод слова для удаления, при вводе проверить слово на присутствие в словаре и удалить при присутствии
* добавить слово для добавления
* добавить значение слова для добавления выше
* вывести результаты на экран и в файл

При написании программы были разработаны следующие функции:

- функция f\_load()предназначена для чтения содержимого файла; входных данных нет, выходные данные – содержимое файла;

- функция f\_load\_t() предназначена для чтения содержимого файла; входных данных нет, выходные данные – содержимое файла;

- функция f\_process (StorageForTextFromReading) – предназначена для обработки содержимого файла; определения использованых для перевода слов, слов, символовбез пробелов и знаком препинания, определения первых слов в предложении и самых длинных среди них; записи результатов в отдельный файл. Входные данные - (StorageForTextFromReading) содержимое файла; выходных данных нет, выводит на экран и в файл количества предложений, слов, количества символов без пробелов и знаков препинания, определения первых слов в предложении и самых длинных среди них.

**Обрабатываемый текст:**

Peter Sikorsky is a very busy man He is 45 years old and he has thirteen jobs He is a postman a policeman a fireman a taxi driver a bus-driver a boatman an ambulance man an accountant a petrol attendant a barman and an undertaker Also he and his wife Kate have a shop and a small hotel Peter lives and works on the island of Gigha in the west of Scotland Only 120 people live on Gigha but in summer 150 tourists come by boat every day Every weekday Peter gets up at 6 00 and makes breakfast for the hotel guests At 8 00 he drives the island's children to school At 9 00 he collects the post from the boat and delivers it to all the houses on the island He also delivers the beer to the island's only pub Then he helps Kate in the shop He says: Kate likes being busy too We never have holidays and we don't like watching television In the evenings Kate makes supper and I do the accounts At 10 00 we have a glass wine and then we go to bed Perhaps our life isn't very exciting but we like it

Обрабатываемый текст находящийся в файле .txt можно увидеть на Рис.1

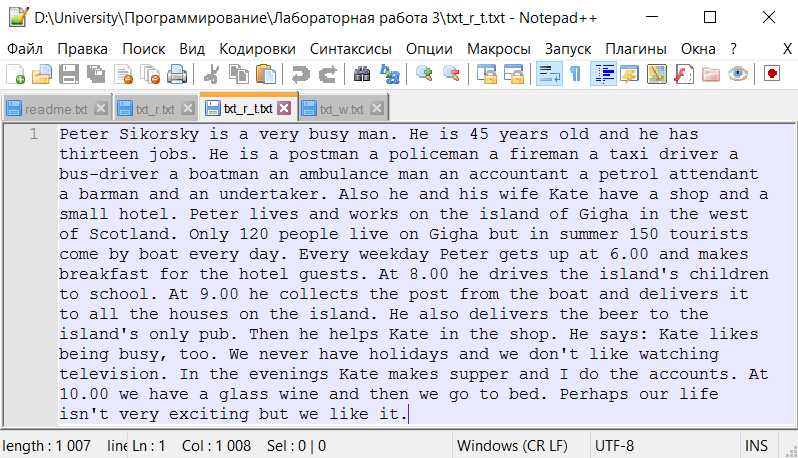


Рис. 1 – Файл формата .txt для чтения, содержимое

# **Листинг программы**

print(**"Servus!\n"**)  
  
  
**def** f\_load():  
 *#Файл для чтения внизу* fileForReading=open(**'txt\_r.txt'**,**'r'**,encoding=**"utf8"**)  
 *#Информацию из открытого файла перекидуем в переменную, указаную ниже* StorageForTextFromReading=fileForReading.read()  
 StorageForTextFromReading=StorageForTextFromReading.replace(**". "**, **"."**)  
 fileForReading.close()  
 **return**(StorageForTextFromReading)  
  
**def** f\_load\_t():  
 *#Файл для чтения внизу* fileForReading=open(**'txt\_r\_t.txt'**,**'r'**,encoding=**"utf8"**)  
 *#Информацию из открытого файла перекидуем в переменную, указаную ниже* StorageForTextFromReading\_t=fileForReading.read()  
 StorageForTextFromReading\_t=StorageForTextFromReading\_t.replace(**". "**, **"."**)  
 fileForReading.close()  
 **return**(StorageForTextFromReading\_t)  
  
**def** f\_process(StorageForTextFromReading,StorageForTextFromReading\_t):  
 fileForWriting=open(**'txt\_w.txt'**,**"w"**,encoding=**"utf8"**)  
 fileForWriting.write(**"Слова, использованые в переводе: \n\n"**)  
 LIST\_1=[]  
 LIST\_2=[]  
 **for** slovo **in** StorageForTextFromReading.split(**","**):  
 **if** slovo==**"."**:**break** LIST\_1.append(slovo)  
 **for** slovo **in** StorageForTextFromReading.split(**"+"**):  
 LIST\_2.append(slovo)  
 **del** LIST\_2[0]  
  
 SLOVAR=dict(zip(LIST\_1,LIST\_2))  
 StorageForTextFromReading\_t=StorageForTextFromReading\_t.replace(**","**, **" "**)  
 StorageForTextFromReading\_t=StorageForTextFromReading\_t.replace(**"."**, **" "**)  
 NEW\_SLOVAR=[]  
 **for** slovo **in** StorageForTextFromReading\_t.split(**" "**):  
 NEW\_SLOVAR.append(slovo)  
 i=0ko=0  
 **for** i **in** NEW\_SLOVAR:   
 **if** i **in** SLOVAR:  
 ko+=1  
 i=i.rstrip(**'.'**)   
 print(i, **'-'**, SLOVAR[i])  
 a1=str(i)  
 a2=str(SLOVAR[i])  
 fileForWriting.write(a1)  
 fileForWriting.write(**":"**)  
 fileForWriting.write(a2)  
 fileForWriting.write(**"\n"**)  
 a10=**" "**.join(NEW\_SLOVAR)  
 print(**"\nОбрабатываемый текст:"**)  
 print(a10)  
 fileForWriting.write(**"\nОбрабатываемый текст:"**)  
 fileForWriting.write(a10)  
 me=0  
 him=0  
 her=0  
 it=0  
 print()  
 **for** i **in** NEW\_SLOVAR:  
 **if** i == **"me"**:  
 me+=1  
 print(**"\n\nКоличество местоимений me: "**,me)  
 fileForWriting.write(**"\nКоличество местоимений me: "**)  
 a3=str(me)  
 fileForWriting.write(a3)  
 **for** i **in** NEW\_SLOVAR:  
 **if** i == **"him"**:  
 him+=1  
 print(**"Количество местоимений him: "**,him)  
 a4=str(him)  
 fileForWriting.write(**"\nКоличество местоимений him: "**)  
 fileForWriting.write(a4)  
 **for** i **in** NEW\_SLOVAR:  
 **if** i == **"her"**:  
 her+=1  
 print(**"Количество местоимений her: "**,her)  
 a5=str(her)  
 fileForWriting.write(**"\nКоличество местоимений her: "**)  
 fileForWriting.write(a5)  
 **for** i **in** NEW\_SLOVAR:  
 **if** i == **"it"**:  
 it+=1  
 print(**"Количество местоимений it: "**,it)  
 a6=str(it)  
 fileForWriting.write(**"\nКоличество местоимений it: "**)  
 fileForWriting.write(a6)  
 print(**"\nКоличество переведенных слов: "**,ko)  
 a7=str(ko)  
 fileForWriting.write(**"\n\nКоличество переведенных слов: "**)  
 fileForWriting.write(a7)  
 NEW\_SLOVAR.sort()  
 print(NEW\_SLOVAR)  
 fileForWriting.write(**"\n\nСлова текста в алфавитном порядке:\n\n"**)  
 a8=**" ,"**.join(NEW\_SLOVAR)  
 fileForWriting.write(a8)  
 kek=str(input(**"Введите слово для удаления (Оставить пустым при нежелании): "**))  
 print(kek)  
 **for** i **in** list(SLOVAR):  
 **if** kek==i:  
 SLOVAR.pop(kek)  
 print(**"\_\_\_\_\_\_\_"**,SLOVAR)  
 VALUE=str(input(**"Введите слово для добавления (англ): "**))  
 KEY=str(input(**"Введите перевод введенного слова (рус): "**))  
*## ПОЧЕМУ-ТО ВЫДАЕТ ИМЕННО ИМЯ ПЕРЕМЕННОЙ В ПЕРЕВОД, А НЕ ЕЁ СОДЕРЖАНИЕ* SLOVAR[KEY] = VALUE  
 print(**"\n"**,SLOVAR)  
   
a=f\_load()  
b=f\_load\_t()  
f\_process(a,b)  
  
  
print(**"\nServus!"**)

**Результат работы**

Результат в окне PyCharm после запуска программы можно увидеть на Рис. 2

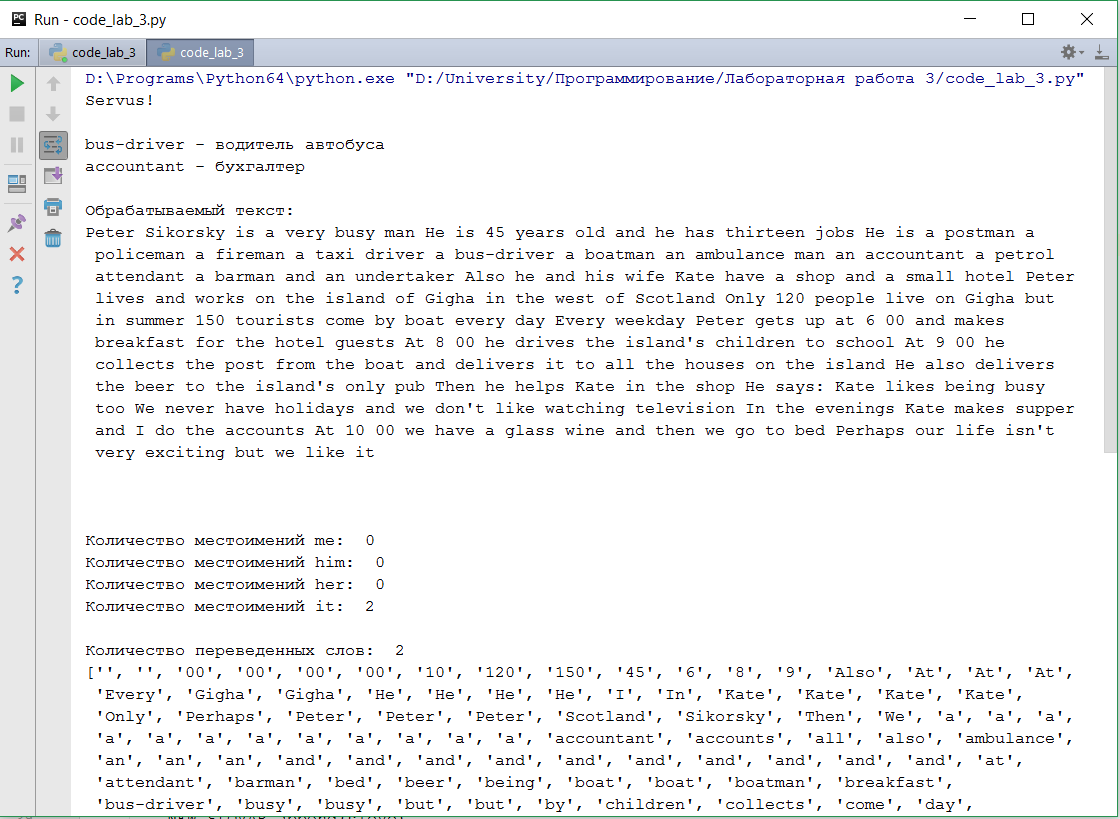


Рис. 2 – Окно Shell при запуске программы

Файл, созданный программой после запуска можно увидеть на Рис. 3

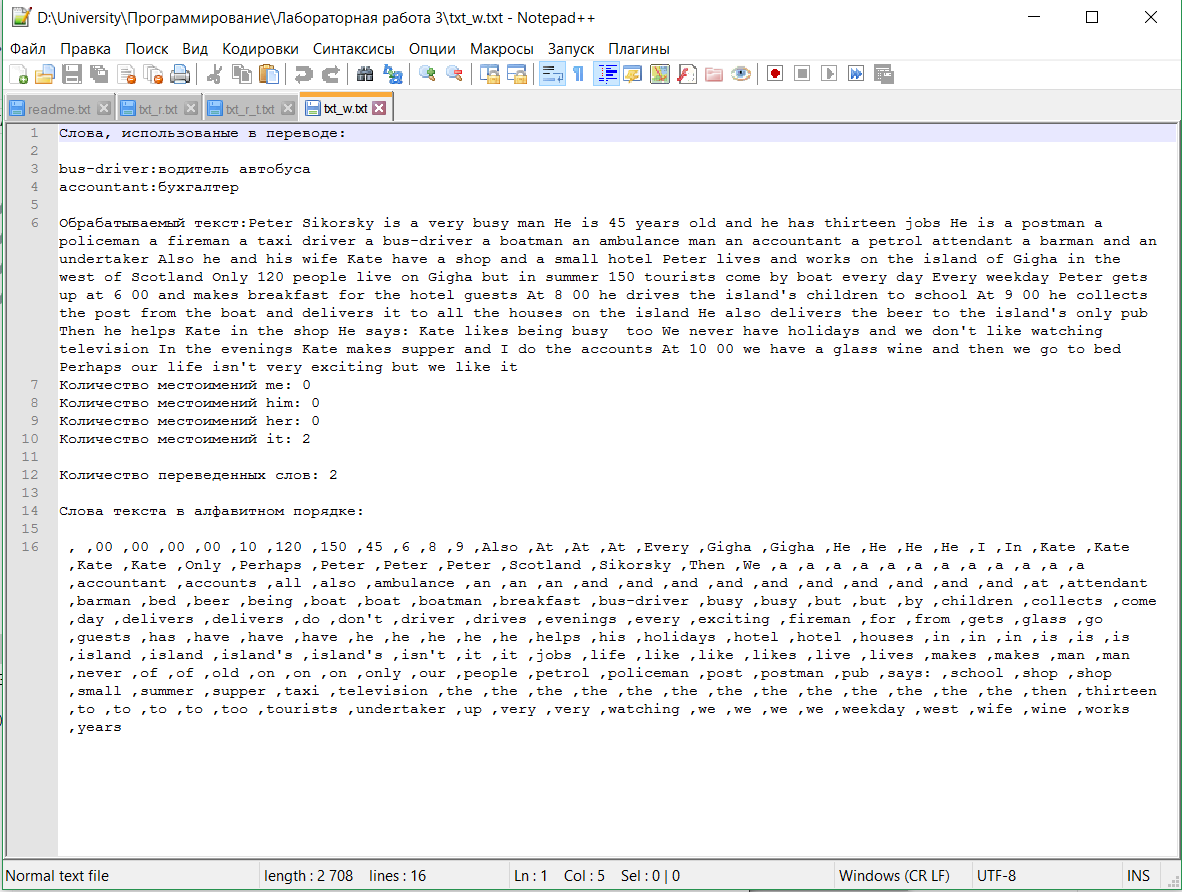


Рис. 3 – Файл формата .txt для записи созданный после запуска программы