МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

ІНСТИТУТ КОМП’ЮТЕРНИХ НАУК ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

### *Кафедра “Системи автоматизованого проектування”*



Звіт

до лабораторної роботи №8

на тему: ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮ­ВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ.СТРУКТУРНЕ ПРОГРАМУВАННЯ МОВОЮ PYTHON.

ОСНОВИ ПРОГРАМУВАННЯ НА МОВІ PYTHON(частина 2) ”

з дисципліни “Комп’ютерна лінгвістика”

Виконала:

студентка групи ПРЛм-11

Павлів І.О.

Прийняв:

викладач

Дупак Б.П.

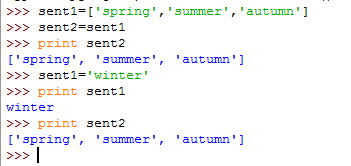
Львів-2015

**Мета роботи:**

* Вивчення основ програмування на мові *Python*.
* Вивчення основ структурного програмування мовою *Python*.
* Повторення та закріплення знань отриманих при виконанні попередніх лабораторних робіт.
* Покращення загальних навичок у програмуванні.

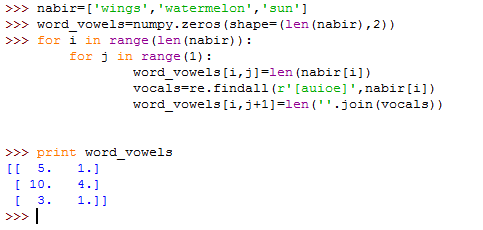
**Варіант №13**

1. Створити список слів і зберегти їх в змінній sent1. Здійснити операцію присвоювання sent2 = sent1. Змінити один з елементів в sent1 і перевірити чи змінився sent2. Результат письмово пояснити.



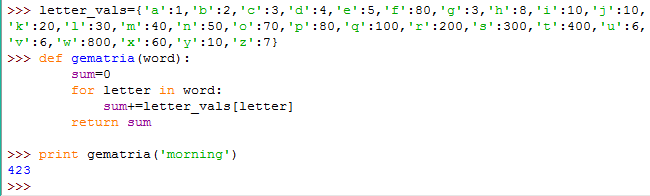
Через те, що ми спершу прирівняли sent1 і sent2, а потім провели зміни, тому sent2 не змінився.

1. Написати програму для створення двовимірного масиву word\_vowels елементами якого є набори. Програма повинна обробити список слів і додати результати обробки до word\_vowels[l][v] де l – довжина слова, v – кількість голосних у слові.

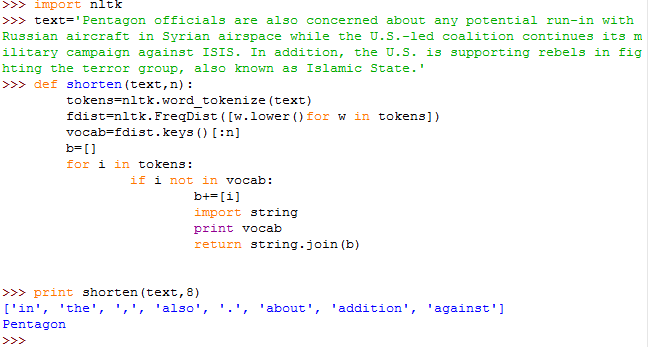


|  |  |  |
| --- | --- | --- |
| |  |  | | --- | --- | |  | **>>> letter\_vals = {a':1, 'b':2, 'c':3, 'd':4, 'e':5, 'f':80, 'g':3, 'h':8,'i':10, 'j':10, 'k':20, 'l':30, 'm':40, 'n':50, 'o':70, 'p':80, 'q':100, 'r':200, 's':300, 't':400, 'u':6, 'v':6, 'w':800, 'x':60, 'y':10, 'z':7}** | |

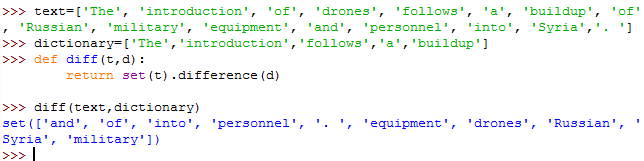
1. Гематрія – метод виявлення прихованого змісту слів на основі порівняння чисел, які відповідають словам. Слова з однаковими числами мають однаковий зміст. Число слова визначається сумуванням чисел, як відповідають його літерам. Написати функцію gematria() для сумування числових значень літер в слові згідно наступних значень letter\_vals:



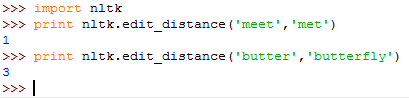
1. Написати функцію shorten(text, n) обробки тексту , для вилучення n найбільш частотних слів в тексті. Яким чином змінилась читабельність тексту, після вилучення цих слів?



15. Написати функцію, яка приймає текст і словник, як аргументи і повертає набір слів, які є у тексті але відсутні у словнику. Аргументи повинні бути представлені, як списки стрічок. Чи може функція мати один рядок при використанні set.difference()?



17. В NLTK реалізовано алгоритм Левінштейна для порівняння стрічок. Спробуйте скористатись цим модулем nltk.edit\_dist(). Яким чином в цьому модулі використовується динамічне програмування? Який підхід використовується знизу-вверх чи зверху-вниз? Пояснити письмово.



Відстань Левенштайна обраховує відстань між двома стрічками, використовуючи підхід знизу-вверх.

**Висновок:**

на цій лабораторній роботі я ознайомилася з явищем гематрії, та навчилася обраховувати відстань Левенштайна у мові програмування Python.