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

Свіляк О.Ю.

Прийняв:

викладач

Дупак Б.П.

Львів-2015

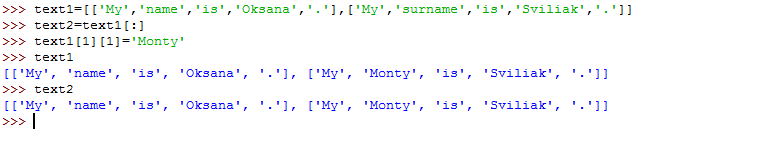
**Мета роботи:**

* 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.

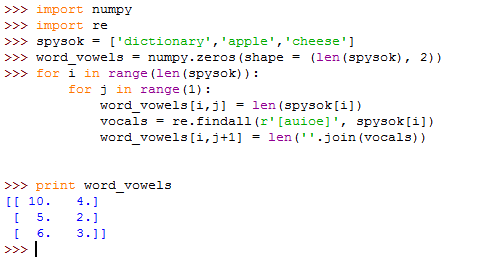
**Тексти програм на мові Python**

**Варіант 15**

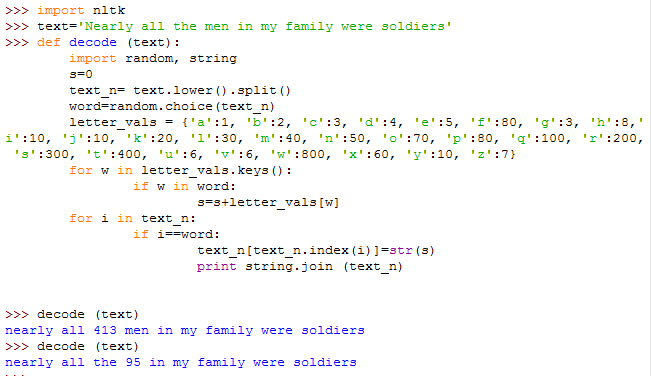
3. Створити список списків слів [ [‘’,’’,’’… ], [[‘’,’’,’’… ], [[‘’,’’,’’…],…] (наприклад текст складається з речень, які складаються з стрічок). Здійснити операцію присвоювання text2 = text1[:], та здійснити операцію присвоювання нового значення одному зі слів (text1[1][1] = 'Monty'). Перевірити як ці операції вплинули на text2. Результат письмово пояснити.

****

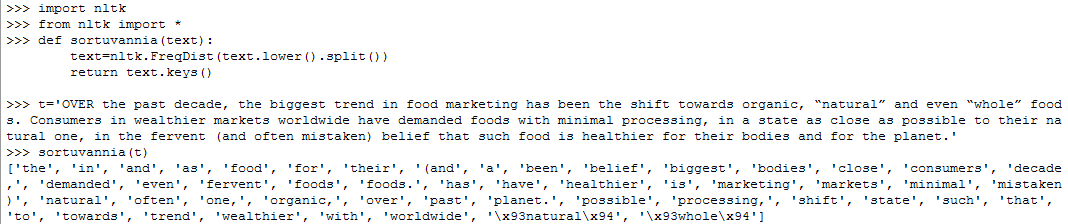
6. Написати програму для створення двовимірного масиву word\_vowels елементами якого є набори. Програма повинна обробити список слів і додати результати обробки до word\_vowels[l][v] де l – довжина слова, v – кількість голосних у слові.



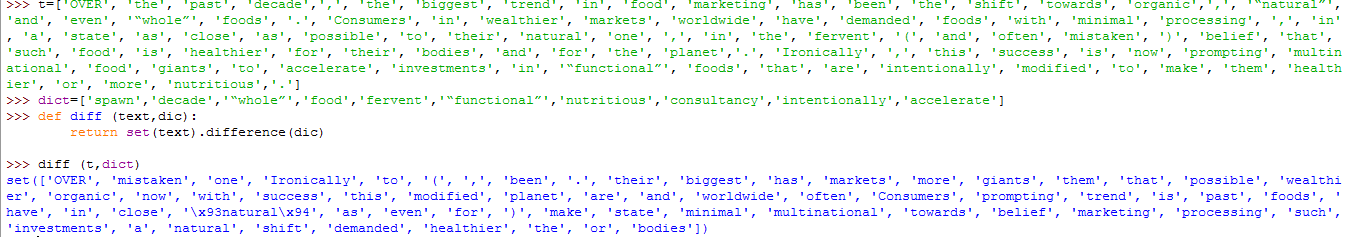
11. Гематрія – метод виявлення прихованого змісту слів на основі порівняння чисел, які відповідають словам. Слова з однаковими числами мають однаковий зміст. Число слова визначається сумуванням чисел, як відповідають його літерам. Написати функцію decode() для обробки тексту, яка випадковим чином замінює слова на їх Гематрія-еквіваленти. Чи вдалося виявити "прихований зміст" тексту? (Використовувати letter\_vals з попередньої задачі).



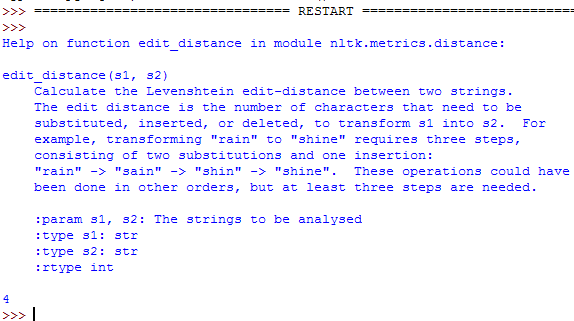
14. Написати функцію, яка обробляє список слів (з дублюванням слів) і повертає список слів (без дублювання) відсортований в порядку спадання їх частоти.



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 Левенштайна 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.