МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

Кафедра САПР



ЗВІТ

до лабораторної роботи № 2

на тему:

ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ

ОСНОВИ ПРОГРАМУВАННЯ НА МОВІ PYTHON (частина 2)

з дисципліни “Комп’ютерна лінгвістика”

Виконала:

Студентка групи ПРЛм-12

Щур Н. М.

Перевірив:

Старший викладач кафедри САПР

Дупак Б. П.

Львів 2015

**МЕТА РОБОТА**

Вивчення основ програмування на мові *Python* та ознайомлення з контрольними структурами та класом FreqDist.

**КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ**

***Засоби керування логікою.***

Основна особливість програмування – це спосібність програми приймати рішення від імені людини, виконуючи інструкції коли справджуються певні умови або послідовно обробляти текстові дані до тих пір поки не задоволена певна умова. Таким засобом керування поведінкою програми є контрольні структури.

***Умовні вирази***

Python підтримує широкий набір операторів для встановлення взаємозв’язків між змінними (значеннями).

***Засоби здійснення простого керування.***

Більшість мов програмування дозволяють виконання окремих блоків програми, коли використовуються умовні вирази або *if* оператори. В наступній програмі ми створили змінну *word*, яка містить значення *‘cat’* типу стрічка. *If*- оператор перевіряє умову чи довжина слова <5, чи ні. Якщо умова виразу справджується, то виконується тіло *if* оператора і виконується оператор print .

***Клас FreqDist для простих статистичних досліджень***

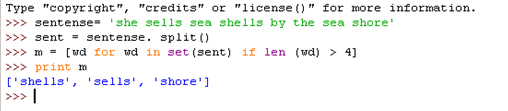
Для автоматичного визначення слів, які є найбільш інформативними для текстів певного жанру або певної тематики спочатку інтуітивно виникає думка побудувати частотний список або частотний розподіл. Частотний розподіл вказує на частоту з якою в тексті зустрічається кожне зі слів. Такий частотний список називають розподілом тому, що він вказує яким чином загальна кількість слів розподіляється між словниковими статями (оригінальні слова) в тексті. Враховуючи що побудова частотних розподілів часто необхідна при обробці природної мови в NLTK реалізовано окремий клас FreqDist в модулі nltk.probability .

**Тексти програм на мові PYTHON**

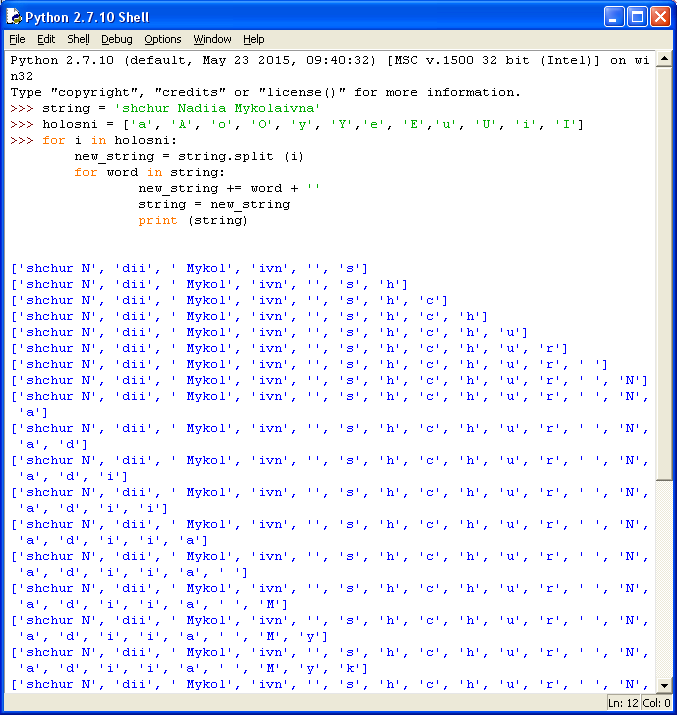
**Варіант 11**

**Завдання 2, 4, 8, 9, 11, 24**

**Завдання 2.** Створіть змінну *sentence* і присвойте їй значення *‘she sells sea shells by the sea shore’* та напишіть фрагмент програми для виведення на екран всіх сліва, довжина яких більша ніж 4 символи.

  
Рис. 1 Завдання 2

**Завдання 4.** Напишіть програму, яка видаляє всі голосні зі стрічки, яка відповідає імені, по батькові та прізвищу студента. Програма повинна здійснювати наступну послідовність дій: створення початкової стрічки; створення стрічки, у якій буде зберігатися результат; *for* цикл для обробки стрічки символ за символом і запису неголосних символів в результуючу стрічку.



**Завдання 8.** Виконати наступні приклади і пояснити різницю між ними w.isupper() not w.islower()

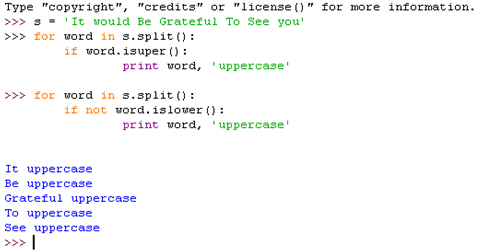
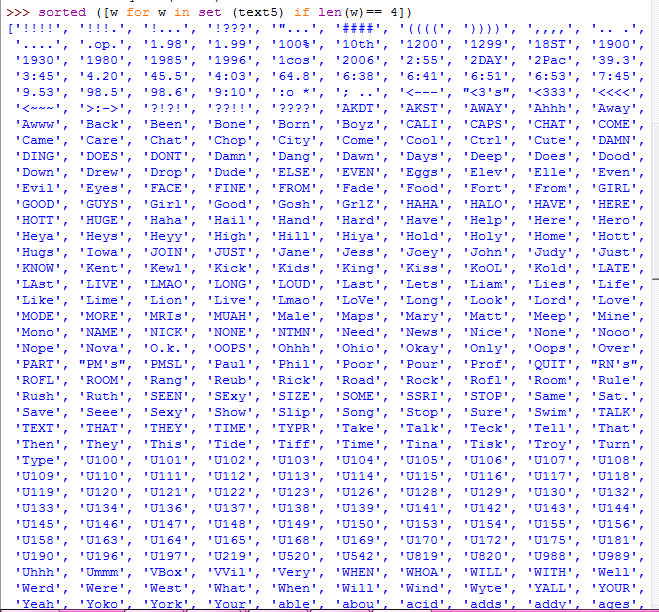
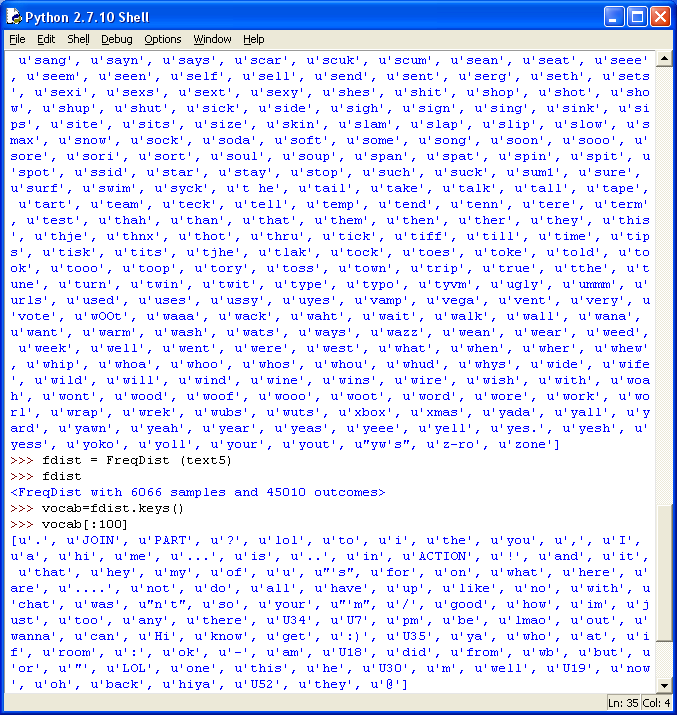


Рис. 3 Завдання 8

**Завдання 9.** Знайдіть в тексті № 5 всі слова довжина яких дорівнює 4 і побудуйте для них частотний розподіл.

  
Рис. 4 Завдання 9



**Завдання 11.** Напишіть вираз для знаходження в тексті №6 всіх слів які відповідають наступним вимогам: закінчуються на ize; містять літеру z; містять послідовність літер pt; написані з великої літери . Результат представити, як список слів.



Рис. 5 Завдання 11

**Завдання 24.** Побудуйте колокації для текстів №2 та №9. Результати порівняйте.

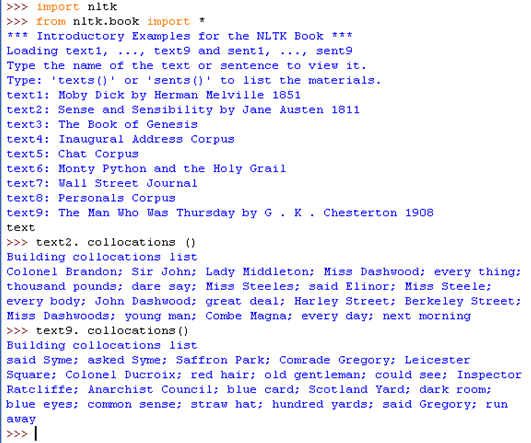


Рис. 6 Завдання 24

Висновок: Вивчили основи програмування на мові *Python* та ознайомились з контрольними структурами та класом FreqDist.