МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

Кафедра САПР

ЗВІТ

ДО ЛАБОРАТОРНОЇ РОБОТИ №8

З ДИСЦИПЛІНИ «КОМП’ЮТЕРНА ЛІНГВІСТИКА»

***ВВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ.***

***СТРУКТУРНЕ ПРОГРАМУВАННЯ МОВОЮ PYTHON (частина2)***

Виконала:

Студентка групи ПРЛм-12

Василишин В. В.

Перевірив:

Викладач кафедри САПР

Дупак Б. П.

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.

**Теоретичні відомості**

Мета розробки модулів програми - виділити логічно пов’язані визначення і функції для спрощення їх повторного використання та розуміння. Модуль в Python не що інше, як окремий \*.py файл. Наприклад, для роботи з форматом певного корпусу, функції читання та запису цього формату повинні бути зібрані разом. Константи, які використовуються (розділювачі полів, чи розширення імен файлів EXTN = ".inf") повинні бути спільні. Якщо формат буде змінено, то ці зміни потрібно буде внести тільки в один файл. Аналогічно, модуль може містити текст програми для створення і маніпулювання певними структурами даних, наприклад синтаксичними деревами, або текст програми для виконання певної задачі, наприклад друк статистичних даних корпусу.

Перед створенням модулів Python корисно розглянути приклади. Можна знайти скомпільовані модулі NLTK на жорсткому диску використовуючи змінну \_\_file\_\_.

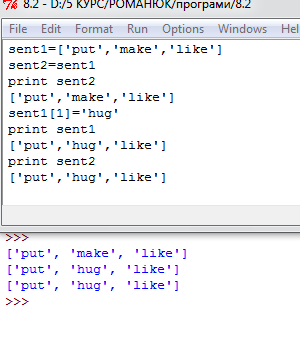
Деякі програми розробляються для вирішення багатьох різнотипних задач. Наприклад, програма повинна завантажити дані, здійснити їх певну обробку чи аналіз а результати візуалізувати. Якщо вже є готові стабільні модулі для завантаження даних та візуалізаці, то розробка такої програми полягає тільки в створенні тієї її частини, яка здійснює обробку чи аналіз даних та виклику потрібних функцій з існуючих модулів.

Найбільш відома стратегія розробки алгоритмів це divide-and-conquer (Поділяй і володарюй). Якщо потрібно вирішити задачу (проблему) розміром n, то можна її поділити на дві проблеми розмірами n/2, вирішити ці проблеми і поєднати ці результати для рішення початкової задачі (проблеми). Наприклад, припустимо, що потрібно відсортувати слова записані на окремих картках. Таке сортування можна виконати віддавши одну половину карток одній людині а іншу другій, які будуть сортувати тільки свої картки (звичайно вони можуть поступити аналогічно). Дві частини вже відсортованих карток не складно об’єднати в одне ціле.

ТЕКСТИ ПРОГРАМ НА МОВІ PYTHON

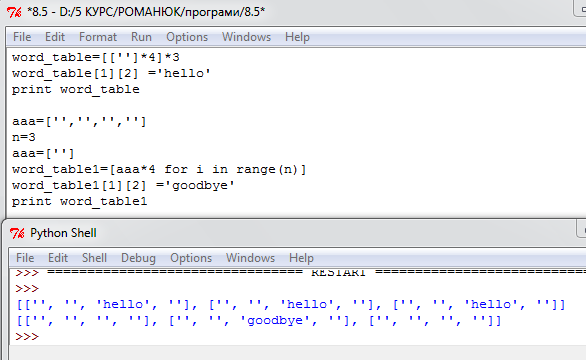
ВАРІАНТ №2

2. 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.

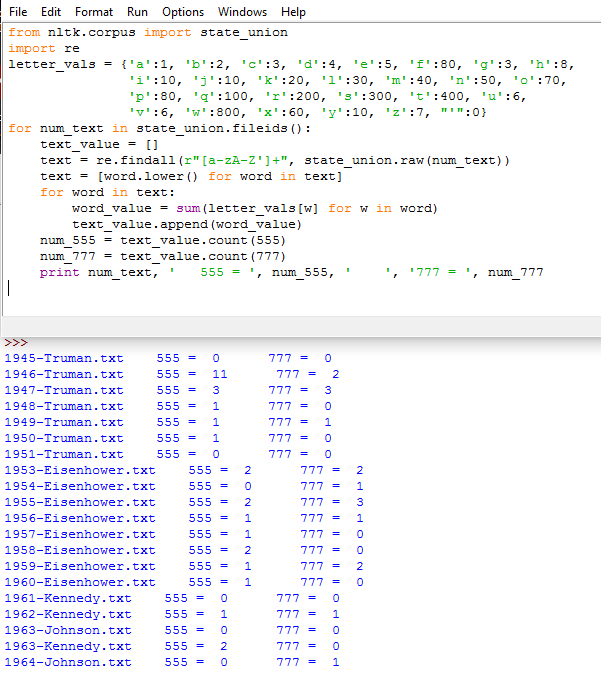


При зміні першого списку слів, другий залишається все рівно таким самим, адже присвоєння відбулося ще до внесення зміни.

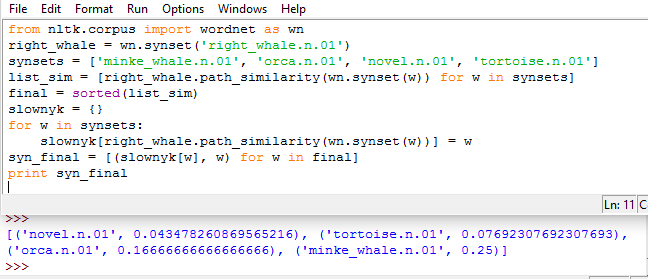
5. Створити список списків пустих стрічок використовуючи множення списків, наприклад word\_table = [[''] \* n] \* m. Що станеться, якщо встановити одне зі значень такого списку, наприклад word\_table[1][2] = "hello"? Пояснити письмово отримані результати. Створіть аналогічний список використовуючи функцію range() і продемонструйте що такий список позбавлений такого недоліку при присвоєнні значення одному з його елементів.



10. 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. Здійснити аналіз корпусу (наприклад nltk.corpus.state\_union). Для кожного з текстів визначити скільки слів мають номер 555 та 777. (Використовувати letter\_vals з попередньої задачі).

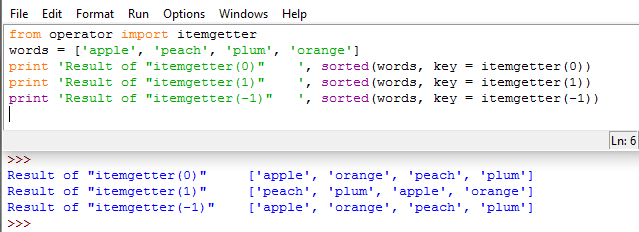


13. Написати list comprehension для сортування списку синсетів WordNet за близькістю до заданого синсету. Наприклад, дані синсети minke\_whale.n.01, orca.n.01, novel.n.01, та tortoise.n.01, потрібно їх відсортувати згідно їх path\_distance() від right\_whale.n.01.

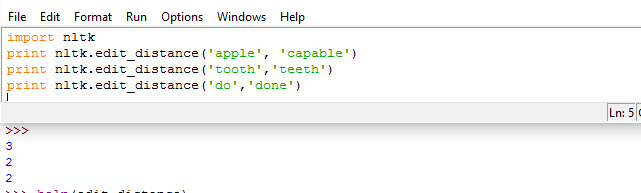


16. Імпортувати функцію itemgetter() модуля operator зі стандартної бібліотеки Python ( from operator import itemgetter). Створити список words , який містить декілька слів. Спробувати виконати: sorted(words, key=itemgetter(1)), та sorted(words, key=itemgetter(-1)). Пояснити письмово роботу функції itemgetter().

Функції itemgetter() повертає позицію, за якою треба здійснити сортування. -1 — сортування за останньою буквою, 1 — сортування за другою буквою, 0 — сортування за першою буквою.



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.



Даний модуль використовує динамічне програмування, адже завдання містить підзавдання, які перетинаються. Використовується підхід знизу вверх, адже робота починається з визначення довжини стрічок, а потім визначає відстань між ними.

**Висновки:** під час виконання даної лабораторної роботи були вивчені основи програмування мови Python, а саме основи структурного програмування.