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

Cтефанко Х.В.

Прийняв:

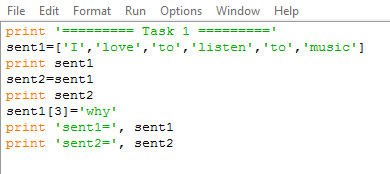
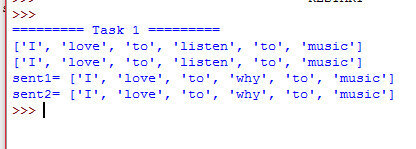
Дупак Б.П.

Львів-2015

.

**Варіант 16**

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.

sent2 змінився, оскільки відбувається присвоєння посилання на значення змінної sent1 новій змінній sent2 .

1. Завантажити функцію Python deepcopy() ( from copy import deepcopy). Прочитати документацію про цю функцію. Перевірити яким чином ця функція створює нові копії об’єктів.

Операція присвоювання не копіює об'єкт, а лише створює посилання на об'єкт. Для змінюваних колекцій, чи для колекцій, що містять змінювані елементи, часто необхідна така копія, щоб її можна було змінити, не змінюючи оригінал. Даний модуль надає загальні (поверхнева і глибока) операції копіювання.

copy.copy (x) - повертає поверхневу копію x.

copy.deepcopy (x) - повертає повну копію x.

Виняток copy.error - виникає, якщо об'єкт неможливо скопіювати.

Різниця між поверхневим і глибоким копіюванням істотна лише для складових об'єктів, що містять змінні об'єкти (наприклад, список списків, або словник, в якості значень якого - списки або словники):

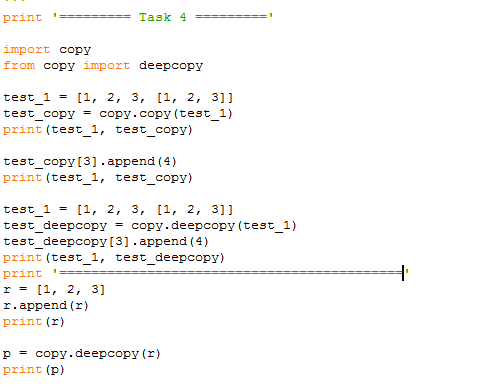
Поверхнева копія створює новий складений об'єкт, і потім (в міру можливості) вставляє в нього посилання на об'єкти, що знаходяться в оригіналі.

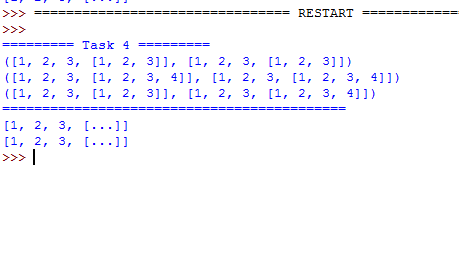
Глибока копія створює новий складений об'єкт, і потім рекурсивно вставляє в нього копії об'єктів, що знаходяться в оригіналі.

Для операції глибокого копіювання часто виникають дві проблеми, яких немає у операції поверхневого копіювання:

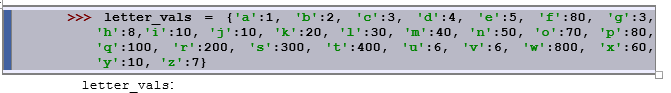
Рекурсивні об'єкти (складені об'єкти, які явно чи неявно містять посилання на себе) можуть стати причиною рекурсивного циклу;

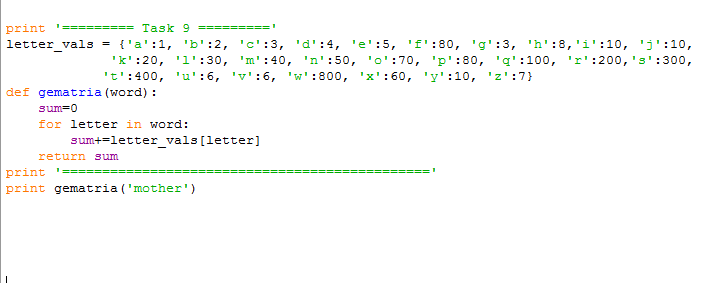
Оскільки глибока копія копіює все, вона може скопіювати занадто багато, наприклад, адміністративні структури даних, які повинні бути розділені навіть між копіями.

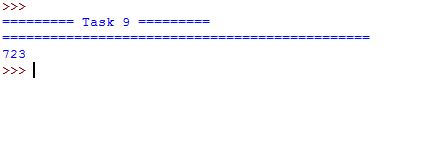




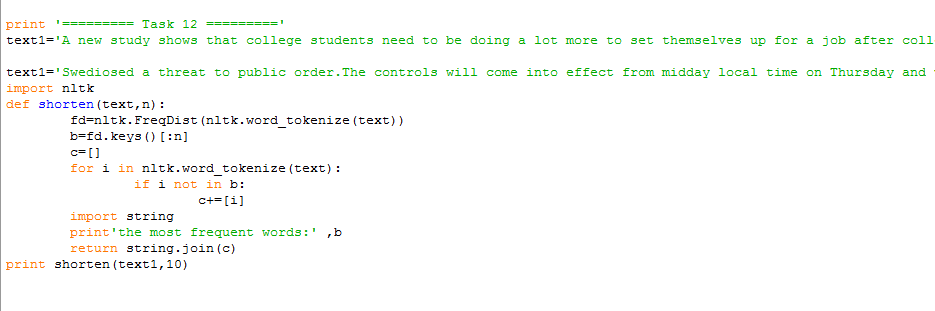
1. Гематрія – метод виявлення прихованого змісту слів на основі порівняння чисел, які відповідають словам. Слова з однаковими числами мають однаковий зміст. Число слова визначається сумуванням чисел, як відповідають його літерам. Написати функцію gematria() для сумування числових значень літер в слові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

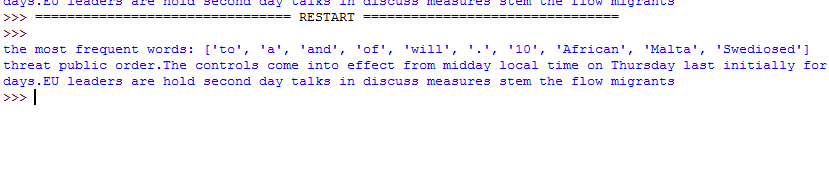


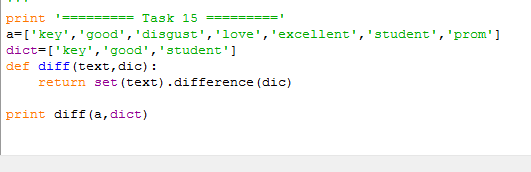


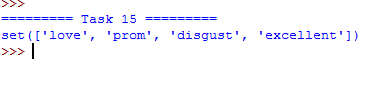


12. Написати функцію shorten(text, n) обробки тексту , для вилучення n найбільш частотних слів в тексті. Яким чином змінилась читабельність тексту, після вилучення цих слів?



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()?





1. 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.

>>> import nltk

>>> help(nltk.edit\_distance)

Help on function edit\_distance in module nltk.metrics.distance:

edit\_distance(s1, s2)

Calculate the Levenshtein edit-distance between two strings.

The edit distance is the number of characters that need to be

substituted, inserted, or deleted, to transform s1 into s2. For

example, transforming "rain" to "shine" requires three steps,

consisting of two substitutions and one insertion:

"rain" -> "sain" -> "shin" -> "shine". These operations could have

been done in other orders, but at least three steps are needed.

:param s1, s2: The strings to be analysed

:type s1: str

:type s2: str

:rtype int

>>> nltk.edit\_distance('curry','criminal')

7

>>> nltk.edit\_distance('bartender','breathe')

6

>>> nltk.edit\_distance('cool','cold')

2

Даний модуль використовує динамічне програмування, оскільки задача містить підзадачі, що перетинаються. Замість повторного рішення цих підзадач, рішення зберігаються у таблицю. Використовується підхід знизу-вверх, оскільки робота починається з визначення довжин стрічок, а тоді йде до визначення відстані між стрічками.