Міністерство освіти та науки України

Національний університет “Львівська політехніка”

інститут комп’ютерних наук та інформаційних технологій

Кафедра “Системи автоматизованого проектування”

Звіт

до лабораторної роботи № 1

з курсу «Комп’ютерна лінгвістика»

на тему:

«Вивчення бібліотеки прикладних програм NTLK, для опрацювання текстів природною мовою. Основи програмування на мові Python(частина 1)»

Виконала:

Студентка групи ПРЛм-11  
Михайлів Р.Б.

Прийняв:

Дупак Б.П.

Львів 2015

**Мета роботи**: вивчення основ програмування на мові Python.

**Короткі теоретичні відомості**

Python - це проста і потужна об’єктно-орієнтована мова програмування високого рівня з чудовими можливостями для обробки лінгвістичних даних.

Natural Language Toolk (NLTK) – набір Python бібліотек, які призначені для аналізу текстів природною мовою. NLTK дозволяє здійснювати символьний та статистичний аналіз текстів, створювати графічні звіти, містить детальну документацію і використовується у проектах з лінгвістики, штучного інтелекту, машинного навчання, автоматизації документообігу. Його можна застосовувати як навчальний комплекс, готовий аналітичний інструмент або платформу для створення прикладних систем опрацювання текстів.

Python – інтерпретаційна мова, що дозволяє зекономити час, що витрачається на компіляцію. Інтерпретатор можна використовувати інтерактивно, що дозволяє експериментувати з можливостями мови і створювати фрагменти програм або тестувати окремі функції. Інтерпретатор даної мови програмування можна використовувати як калькулятор, виконуючи різні математичні операції.

Текст або частини тексту у програмах на Python представляються за допомогою стрічок (string) і повинен відділятися від решти програми лапками (одинарними, подвійними або потрійними).

Змінна, яка називається стрічкою (має тип змінної - стрічка) є дуже важливою при опрацюванні текстів. Усі позиції в стрічці є пронумеровані починаючи з нуля. Для доступу до окремої позиції у стрічці ми записуємо цю позицію у квадратних дужках. Така операція називається індексуванням. Позиція записана в квадратних дужках називається індексом. Цей номер додається до фізичної адреси змінної, яка знаходиться в пам’яті. Для визначення довжини стрічки і відповідно максимального індекса можна порахувати символи і відняти 1, або використати функцію len().

У випадку, коли ми хочемо отримати доступ до більш ніж одного символа ми повинні у квадратних дужках вказати індекс першого символа і останнього розділивши їх двокрапкою. Цей процес називається slicing (зріз).

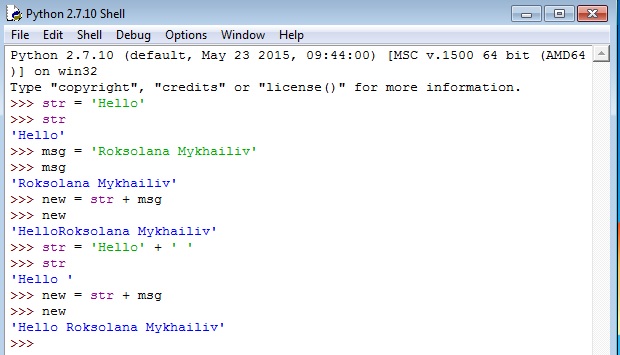
Lists (списки) – тип даних для опису послідовності значень. У Python списки представляються як послідовність, записана через кому і у квадратних дужках. Списки багато в чому подібні до стрічок, але елементами списків на відміну від стрічок можуть бути не тільки окремі символи, але і цілі стрічки та інші списки. Доступ до значень списків отримуємо аналогічно до стрічок. Значення у списках можно змінювати на відміну від стрічок, де значення змінити неможливо. Вміст списків можна міняти в будь-який момент часу і тому списки підтримують набір операцій або методів.

Часто при роботі з послідовностями виникає необхідність послідовно працювати з її елементами. Ми можемо це робити використовуючи for цикл.

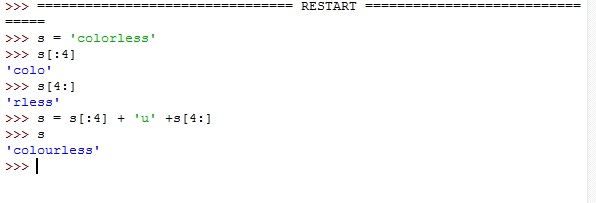
**Тексти програм на мові Python**

Варіант 20

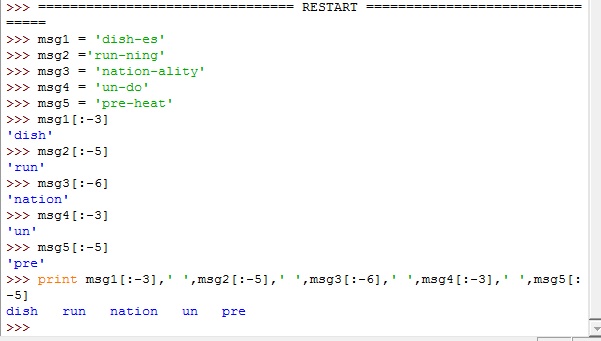
* 1. 3.4 Визначити нову стрічку *hello*. Здійснити операцію *hello+ msg.* Змінити стрічку *hello* додавши в її кінці символ пробілу і знову виконати операцію *hello+ msg.*

**

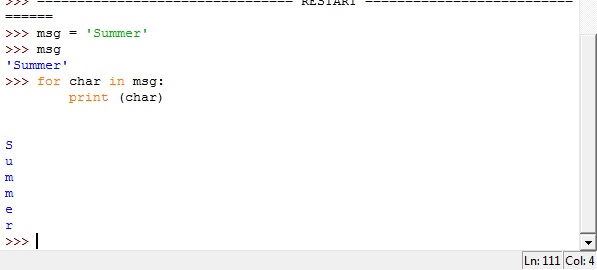
* 1. 3.6 Визначити стрічку s=’colorless’. Використовуючи зрізи та операцію поєднання змінити стрічку до вигляду ‘colourless’.



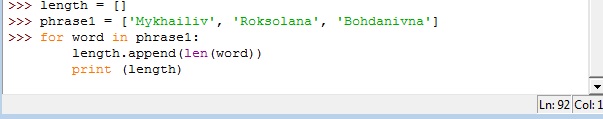
* 1. 3.7 Використовуючи зрізи видаліть афікси у наступних словоформах: *dish-es, run-ning*, *nation-ality, un-do, pre-heat*.



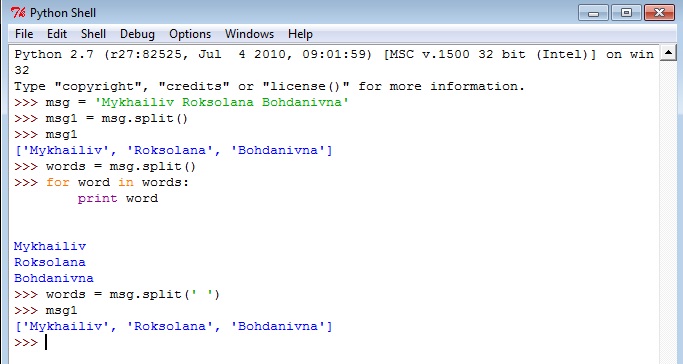
* 1. 3.14 Напишіть for цикл, який виведе на екран символи стрічки *msg* по одному на рядок.



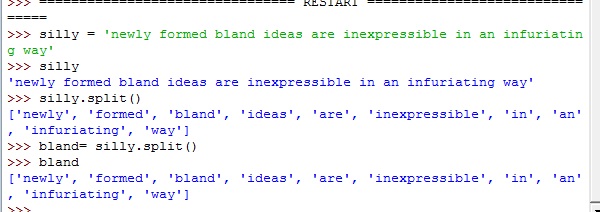
* 1. 3.18 Напишіть for цикл, який обробить *phrase1* визначивши довжину кожного елементу і результати збереже в новому списку *lengths*. (Створіть пустий список *lengths = [].* Далі використовуйте метод *append()* в тілі циклу для додавання довжин до списку).



* 1. 3.19 Перетворіть стрічку *msg* на список стрічок, кожна з яких відповідає одному слову використовуючи *split()* оператор без символа розділення та наступними символами розділення: подвійні лапки, табуляція, послідовність пробілів, послідовність табуляцій та пробілів.

****

* 1. 3.20 Визначіть змінну *silly*, яка буде містити стрічку ’newly formed bland ideas are inexpressible in an infuriating way’ і напишіть програму перетворення стрічки на список стрічок (кожне слово окрема стрічка). Результати збережіть в змінній *bland.*

****

**Висновок**

На даній лабораторній роботі ознайомилася із мовою програмування Python, яка є потужною об’єктно-орієнтованою мовою програмування високого рівня з чудовими можливостями для обробки лінгвістичних даних. Ознайомилась з типами даних: стрічка, список, кортеж. Навчилась застосовувати операцію поєднання, використовувати зрізи, перетворювати стрічку у список і навпаки, сортувати елементи за абеткою, створювати файл test.py і використовувати його для виконання завдань. Виконала приклади, які були подані у теоретичних відомостях та вправи згідно варіанта індивідуального завдання.