МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

Кафедра САПР



ЗВІТ

до лабораторної роботи № 1

на тему:

ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ

ОСНОВИ ПРОГРАМУВАННЯ НА МОВІ PYTHON (частина 1)

з дисципліни “Комп’ютерна лінгвістика”

Виконала:

Студентка групи ПРЛм-12

Щур Н. М.

Перевірив:

Старший викладач кафедри САПР

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 та містить детальну документацію і використовується в проектах з лінгвістики, штучного інтелекту, машинного навчання, автоматизації документообігу.

**1.1 Використання інтерпретатора Python, як калькулятора.**

Спробуємо використати Python, як калькулятор. Після того як ми натискаємо Enter виконуються дії, інтерпретатор видає результат і чекає вводу наступного оператора. Операція множення виконана вірно з додержанням пріоритету виконання арифметичних дій.

**1.2 Стрічки і змінні.**

**1.2.1 Представлення тексту.**

Текст або частини тексту в програмах на Python представляються за допомогою стрічок (string) і повинен відділятися від решти програми лапками (одинарними, подвійними або потрійними) Якщо стрічка містить одинарні лапки необхідно використовувати лівий слеш перед апострофом для того, щоб символ апострофа не розглядався як символ завершення стрічки або використовувати подвійні лапки. Якщо цього не зробити то отримаємо помилку.

**1.2.2 Змінні. Операції присвоювання.**

Для того щоб використовувати, в подальшому, значення, які ми отримали вище, нам потрібно їх зберегти розмістивши в певному місці в пам’яті комп’ютера. Місце де ми збережемо ці значення називається змінна і ми їй присвоюємо якесь ім’я. Значення цій змінній присвоюються операцією присвоювання. Оператор присвоювання це знак дорівнює.

**1.2.3 Друк і перегляд стрічок.**

Для того щоб вивести вміст змінної на екран ми набираємо ім’я цієї змінної і натискаємо Enter. Для відображення значення змінної на екрані наша програма має містити команду print.

**1.3 Доступ до окремих символів стрічки та виділення під стрічок.**

**1.3.1 Доступ до окремих символів стрічки.**

Змінна яка називається стрічкою (має тип змінної - стрічка) є дуже важливою при опрацюванні текстів. Всі позиції в стрічці є пронумеровані починаючи з нуля. Для доступу до окремої позиції в стрічці ми записуємо цю позицію в квадратних дужках. Така операція називається індексуванням. Позиція записана в квадратних дужках називається індексом. Коротко, функція це іменований фрагмент програми, який виконує певні дії при виклику і виконанні функції. Функція len() є вбудованою функцією її виклик здійснюється за назвою і параметром є назва стрічки.

**1.3.2 Доступ до підстрічок.**

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

**1.4.1 Списки.**

**Lists (списки)** – тип даних для опису послідовності значень. В 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 і стрічки і інші списки.

**1.4.2 Обробка послідовностей.**

Часто при роботі з послідовностями виникає необхідність послідовно працювати з її елементами. Ми можемо це робити використовуючи for цикл. Це є контрольна структура Python.

**1.4.3 Кортежі.**

**Tuples(кортежі)** – подібний до списків тип даних але подібно до стрічок кортежі також не можемо змінювати.

**1.4.4 Форматування стрічок.**

Команда print має наступний синтаксис: format % values. Format це частина яка містить специфікатори формату такі як %s %d (%s – специфікатор вказує Python що відповідна змінні є стрічкою або може бути конвертована в стрічку, %d - специфікатор вказує Python що відповідна змінна може бути конвертована в десяткове число). Values це частина де записаний кортеж значень для який визначені специфікатори формату.

**Тексти програм на мові PYTHON**

**Варіант 11**

**Завдання 3, 9, 10, 13, 17, 23,24**

**Завдання 3.** Здійснити арифметичні операції зі стрічкою *msg.*

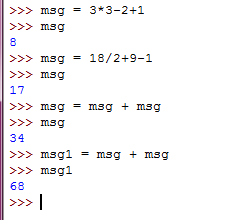


Рис. 1 Завдання 3.

**Завдання 9.** Організуйте доступ до елементів стрічки з визначеним кроком. Результати поясніть.

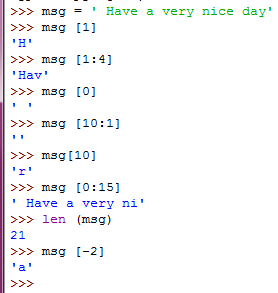


Рис. 2. Завдання 9.

Всі позиції в стрічці є пронумеровані починаючи з нуля. Для доступу до окремої позиції в стрічці я записувала цю позицію в 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.

**Завдання 10**. Поясніть результат виконання *msg[::-1].*

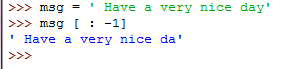


Рис.3 Завдання 10.

Такий від’ємний зріз [ : -1] означає, що значення стрічки *msg* в моєму випадку буде наступним: від початку значення і до кінця, але мінус 1 індекс, що в результаті і показав Python.

**Завдання 13.** Представити прізвище, ім’я та по батькові як список стрічок. Використовуючи метод .reverse() та зріз [::-1] змінити стрічку. Результати пояснити.

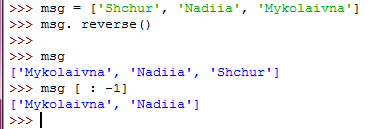


Рис. 4. Завдання 13.

При застосування функції *reverse ()* значення виводиться на екран в оберненому порядку, що в результаті і показала програма. При застосуванні функції зрізу [ : -1] я отримала всі значення змінної *msg,*окрім, як було зазначено вище, останньої.

**Завдання 17.** Створіть файл test.py, який містить стрічку *msg* .Використайте наступні оператори і поясніть отримані результати.

>>> from test import msg

>>> msg

На першому етапі я створила файл test.py, який містив стрічку *msg.*

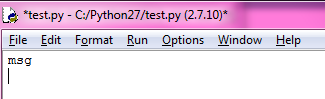


Рис. 5.1 Завдання 17

Оскільки компілятор не працює при такому «пустому» вводі,:

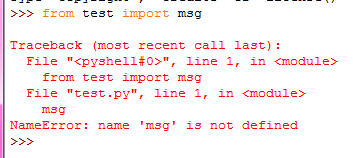


Рис. 5.2 Завдання 17

я присвоїла змінній *msg* значення:

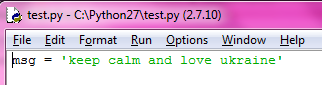


Рис. 5.3 Завдання 17

Після такої дії, результат був очевидним:

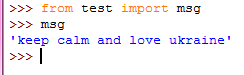


Рис. 5.4. Завдання 17

**Завдання 23.** Напишіть програму, яка надрукує слова із стрічки *silly* за абеткою.

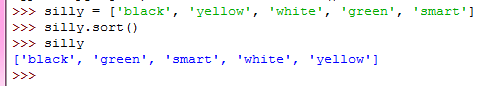


Рис. 6. Завдання 23

**Завдання 24**. Використайте функцію index() наступним чином ’inexpressible’.index(’e’). Що станеться якщо виконати ’inexpressible’.index(’re’)

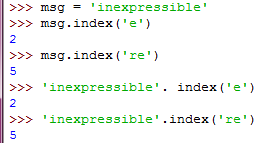


Рис. 7. Завдання 24

**Висновок:** У цій лабораторній роботі я вивчила основи програмування на Python. Навчилася виконувати операції на стрічками та списками. Виконала завдання лабораторної роботи.