МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

ІНСТИТУТ КОМП’ЮТЕРНИХ НАУК ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

### *Кафедра “Системи автоматизованого проектування”*



Звіт

до лабораторної роботи №9

на тему: ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮ­ВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ.АВТОМАТИЧНИЙ МОРФОЛОГІЧНИЙ АНАЛІЗ.

ОСНОВИ ПРОГРАМУВАННЯ НА МОВІ PYTHON(частина 2) ”

з дисципліни “Комп’ютерна лінгвістика”

Виконала:

студентка групи ПРЛм-11

Свіляк О.Ю.

Прийняв:

викладач

Дупак Б.П.

Львів-2015

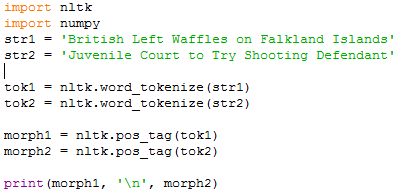
**Мета роботи:**

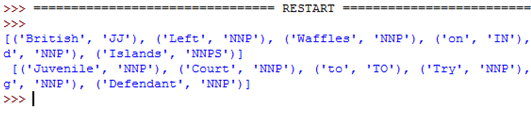
* Вивчення основ програмування на мові *Python*.
* Ознайомлення з автоматичним морфологічним аналізом в NLTK.

**Тексти програм на мові Python**

**Варіант 15**

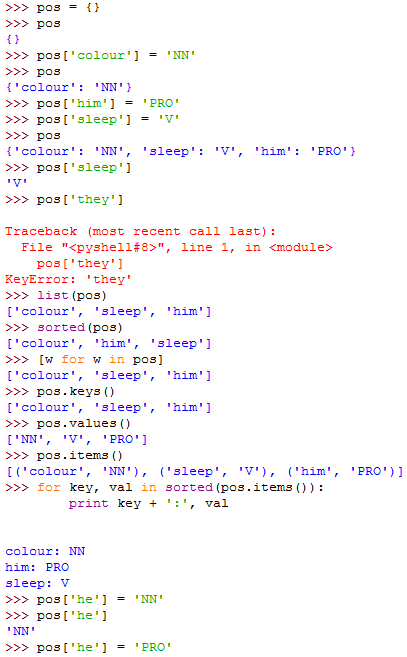
15. Знайдіть заголовки газетних публікацій подібні до наступних «British Left Waffles on Falkland Islands», «Juvenile Court to Try Shooting Defendant». Промаркуйте ці заголовки і опишіть, яким чином знання про набори тегів корпуса Brown дозволяють розв’язати неднозначності.



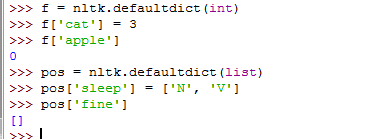


Програма починає аналіз першого слова, яке вважає прикметником, наступне слово вважається іменником, так як після прикметників зазвичай є іменники. Всі слова вважаються власними назвами .

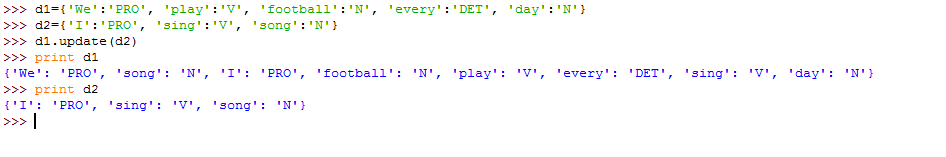
3. Опрацювати всі приклади з методичних вказівок по роботі зі словниками. Що станеться, якщо доступитися до неіснуючого запису звичайного словника та словника ?



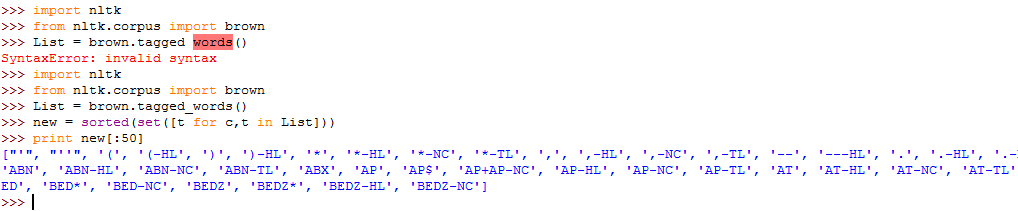




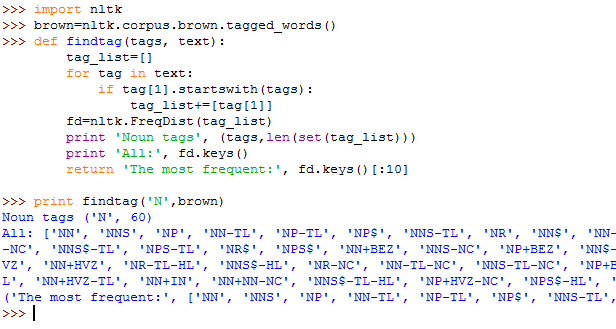
5. Створити два словники. Що станеться зі словниками після виконання команди d1.update(d2) ?



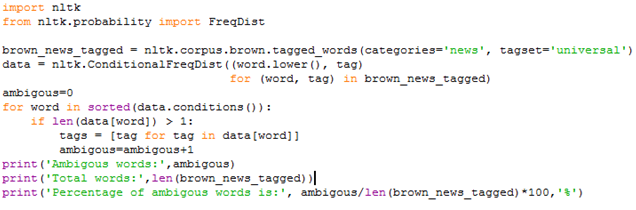
7. Використовуючи sorted() та set() отримайте відсортований список всіх тегів корпуса Brown без їх дублювання.



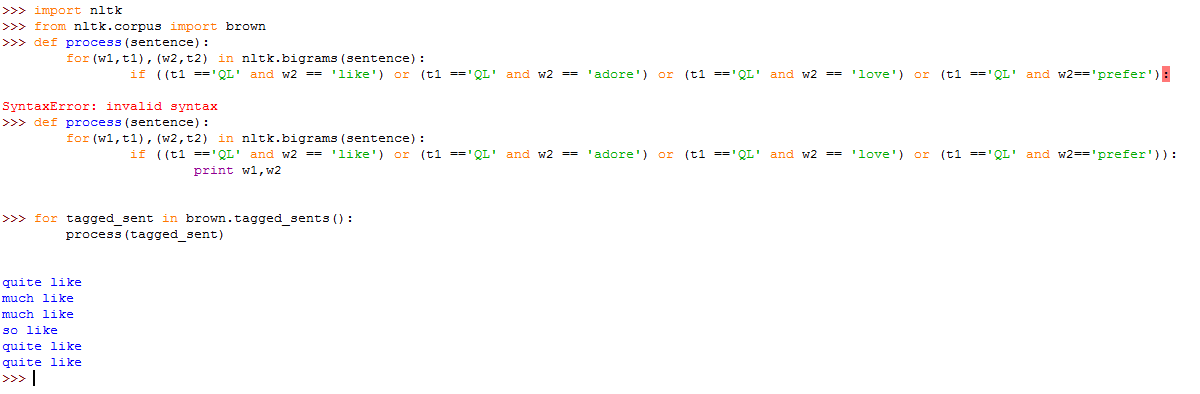
11. Напишіть програму, яка обробить Brown Corpus і допоможе відповісти на наступне запитання: які теги для маркування іменників найчастіше використовуються і що вони означають.



14. Напишіть програму для збору статистичних даних по розмічених корпусах і відповіді на наступне запитання: який відсоток слів корпусу Brown мають неоднозначності.



20. Напишіть програму для знаходження всіх слів, які маркуються як QL в корпусі Brown і зустрічаються перед наступними дієсловами adore, love, like, prefer.



21. Написати програму побудови словника, записами якого будуть набори словників. Використовуючи створений словник, збережіть у ньому набори можливих тегів, які зустрічаються після заданого слова з певним тегом, наприклад wordi → tagi → tagi+1.

