**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”**

**іНСТИТУТ КОМП’ютерних НАУК та ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

### Кафедра “Системи автоматизованого проектування”



Звіт

до лабораторної роботи № 3

на тему: «ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ. ДОСТУП ТА РОБОТА З КОРПУСАМИ ТЕКСТІВ.”

з дисципліни “Комп’ютерна лінгвістика”

Виконала:

Студентка групи ПРЛм-11

Турчиняк О.І.

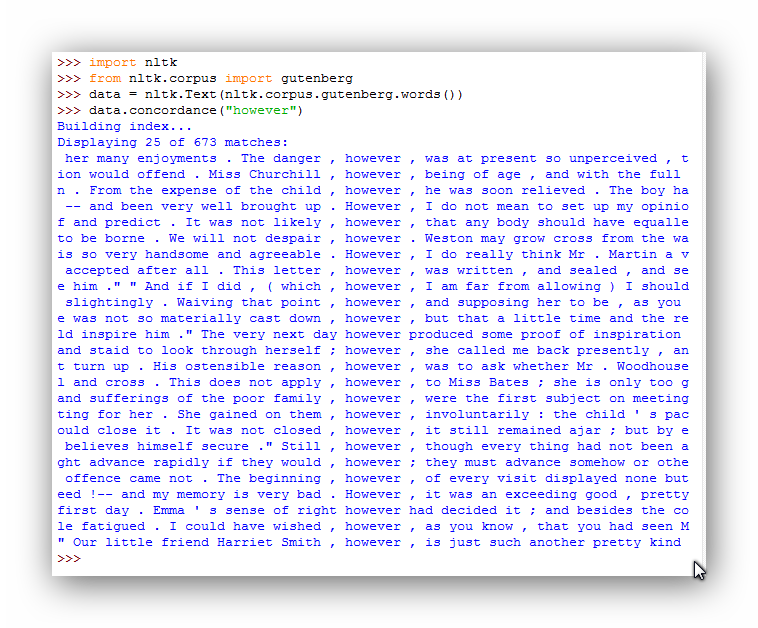
Прийняв:

Викладач

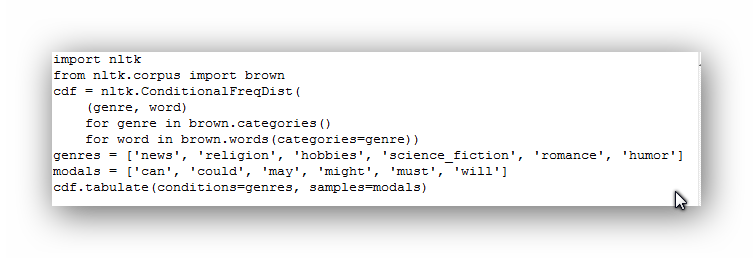
Дупак Б.П.

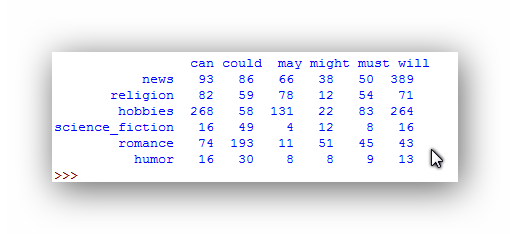
Львів 2015

4. Використовуючи конкорданси поясніть відмінності у вживанні слова however на початку речення ("in whatever way", "to whatever extent", або "nevertheless”).



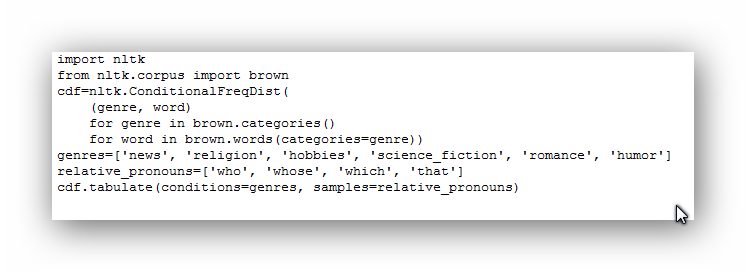
6. Проаналізуйте таблицю частот модальних дієслів для різних жанрів. Спробуйте її пояснити. Знайдіть інші класи слів вживання яких також відрізняються в різних жанрах.

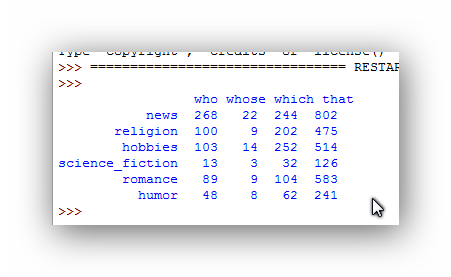




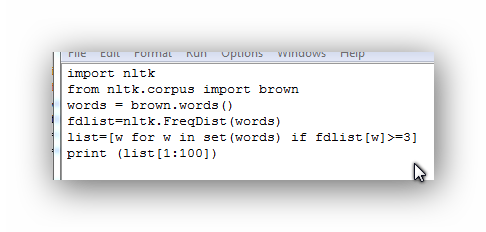
Проаналізувавши таблицю частот модальних дієслів для різних жанрів, можна зробити висновок, що: у жанрі news слово will зустрічається найчастіше. Can – найчастотніше слово у жанрах religion та hobbies. У жанрах science\_fiction, romancе, humour – could.

Вживання відносних займенників у різних жанрах.



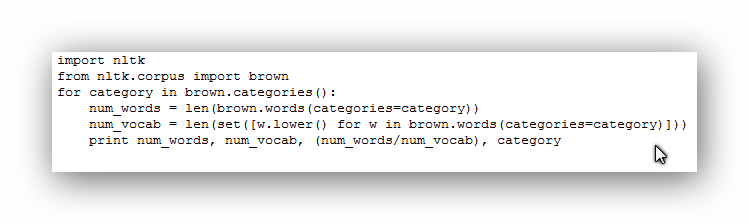
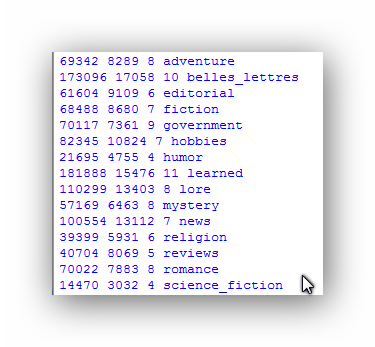


7. Напишіть програму для знаходження всіх слів в корпусі Brown, які зустрічаються не менш ніж три рази.

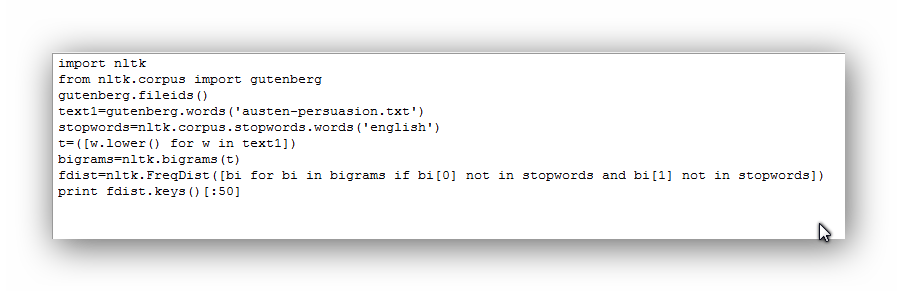
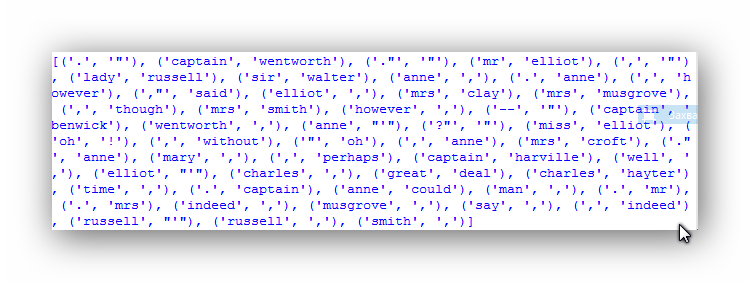




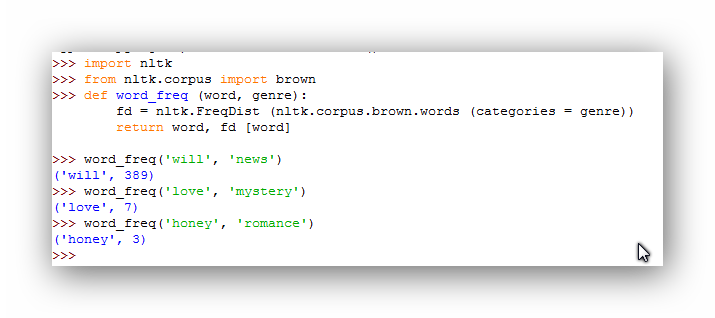
8. Напишіть програму генерації таблиці відношень кількість слів/кількість оригінальних слів для всіх жанрів корпуса Brown.

10. Напишіть програму яка виводить на екран 50 найчастотніших біграмів тексту, за виключенням біграмів до складу яких входять незначущі слова.

12. Напишіть функцію word\_freq(), яка приймає слово і назву частини корпуса Brown як аргументи і визначає частоту слова в заданій частині корпуса.



Висновок: я ознайомилась основами програмування на мові Python.Освоїла методів доступу до корпусів текстів. Та опанувала клас ConditionalFreqDist.