МІНІСТЕРСТВО ОСВІТИ, НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”**



Звіт

до лабораторної роботи №11 на тему:

«ВИВЧЕННЯ БІБЛІОТЕКИ ПРИКЛАДНИХ ПРОГРАМ NLTK, ДЛЯ ОПРАЦЮВАННЯ ТЕКСТІВ ПРИРОДНОЮ МОВОЮ.

АВТОМАТИЧНИЙ СИНТАКСИЧНИЙ АНАЛІЗ (частина2)»

Виконала:

ст. групи ПРЛм-12

ІКНІ

Василишин В. В.

Прийняв:

Асистент кафедри САПР

## Дупак Б. П.

## Львів – 2015

**МЕТА РОБОТА**

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

КОРОТКІ ТЕОРЕТИЧНІ ВІДОМОСТІ

Мета граматики – дати явний опис природної мови. Щоб описати мову потрібно визначитись що вважати природною мовою та вивчити основні підходи до її представлення.

В лабораторній роботі розглядається формальне представлення породжуючої граматики, згідно якої мова представляється, як множина всіх граматично вірних речень, а граматика це формальна система , яка може бути використана для генерації елементів цієї множини.

## Контекстно-вільна граматика

### Проста граматика

Розглянемо просту (найпростішу) контекстно-вільну граматику. Згідно означення, першим символом зліва в першому правилі граматики є спеціальний початковий символ S, і всі дерева повинні мати цей символ , як корінь. В NLTK, контекстно-вільна граматика визначається в модулі nltk.grammar .

| **Символ** | **Значення** | **Приклад** |
| --- | --- | --- |
| S | sentence | the man walked |
| NP | noun phrase | a dog |
| VP | verb phrase | saw a park |
| PP | prepositional phrase | with a telescope |
| Det | determiner | the |
| N | noun | dog |
| V | verb | walked |
| P | preposition | in |
|  |  |  |

Правило подібне до VP -> V NP | V NP PP містить оператор диз’юнкції в правій частині |, і це є скорочений запис двох правил VP -> V NP та VP -> V NP PP.

**Рекурсія в синтаксичних структурах**

Граматику називають рекурсивною, якщо категорії з лівої частини її правил також зустрічаються і їх правих частинах, як показано в наступному прикладі. Правило Nom -> Adj Nom містить пряму рекурсію категорії Nom, тоді як не пряма рекурсія S виникає з комбінації двох правил S -> NP VP та VP -> V S.

**Синтаксичний аналіз на основі контексно-вільної граматики.**

Синтаксичний аналізатор це програма, яка обробляє вхідне речення згідно правил граматики і будує одну або декілька синтаксичних структур, які відповідають (узгоджуються) цій граматиці. Граматика, це декларативна специфікація певних закономірностей - це насправді стрічка (набір стрічок) а не програма. Аналізатор здійснює процедурну інтерпретації граматики – шукає серед лісу дерев, які може породжувати граматика одне потрібне дерево, яке відповідає вхідному реченню.

Практична частина

Варіант №2

2. В класі Tree реалізовано різноманітні корисні методи. Переглянути файл

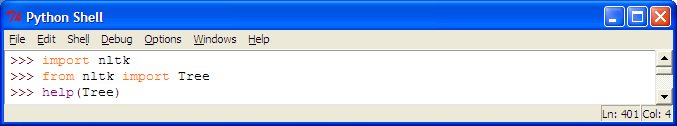
допомоги  Tree з документації та описати основні з цих методів (import

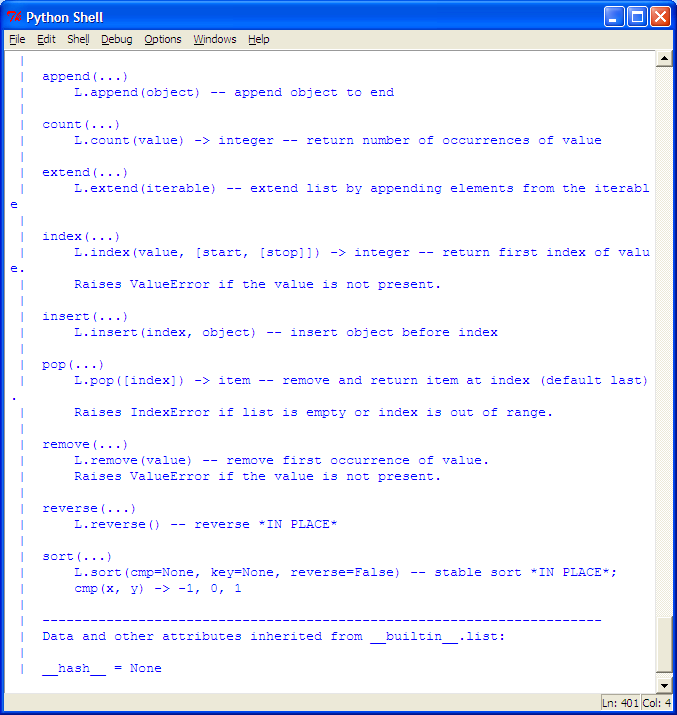
Tree,  help(Tree).

Викликавши файл допомоги Tree, видно, що основним методом є

chomsky\_normal\_form, оскільки його інтерпретації зустрічаються частіше, ніж

інших методів. Файл допомоги пояснює принцип поб





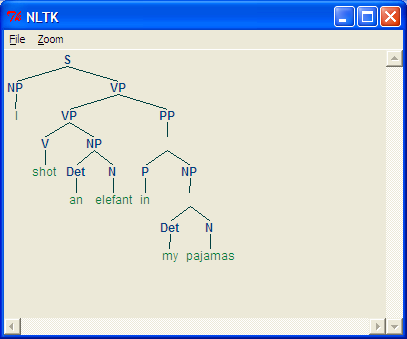
4. Перетворити всі дерева, які зустрічаються в методичних вказівкаx

зображені за допомогою дужок використовуючи  nltk.Tree() .

Використовувати draw() для побудови графічного зображення дерева.

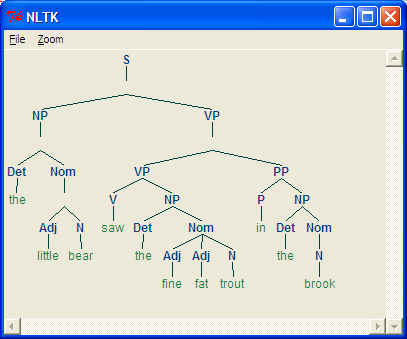
>>> t=nltk.Tree('(S(NP I)(VP(VP(V shot) (NP(Det an) (N elefant))) (PP((P in) (NP((Det my) (N pajamas)))))))')

>>> t.draw()



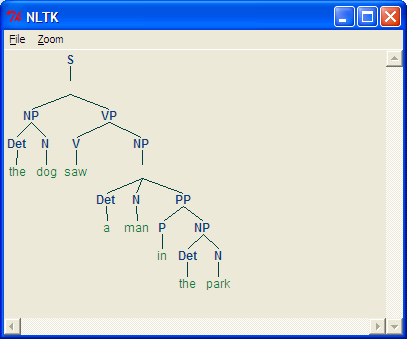
>>>t=nltk.Tree('(S((NP((Det the)(Nom((Adj little)(N bear)))))(VP((VP(V saw)(NP(Det the)(Nom(Adj fine)(Adj fat)(N trout))))(PP(P in)(NP(Det the)(Nom(N brook))))))))')

>>> t.draw()



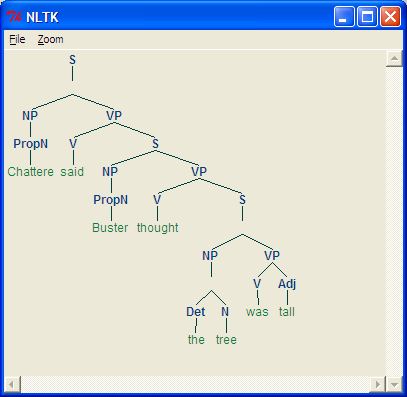
>>>t=nltk.Tree('(S((NP(Det the)(N dog))(VP(V saw)(NP((Det a)(N man)(PP(P in)(NP(Det the)(N park))))))))')

>>> t.draw()



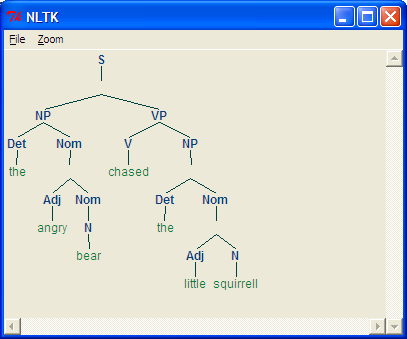
>>>t=nltk.Tree('(S((NP(PropN Chattere))(VP(V said)(S(NP(PropN Buster))(VP(V thought)(S((NP((Det the)(N tree)))(VP(V was)(Adj tall)))))))))')

>>> t.draw()



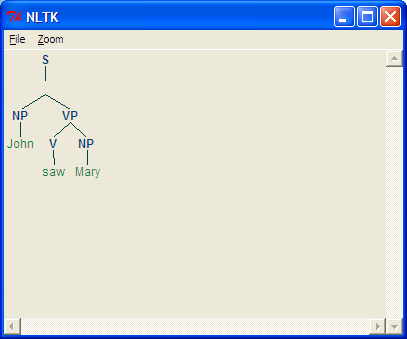
>>> t=nltk.Tree('(S((NP(Det the)(Nom((Adj angry)(Nom(N bear)))))(VP(V chased)(NP((Det the)(Nom((Adj little)(N squirrell))))))))')

>>> t.draw()

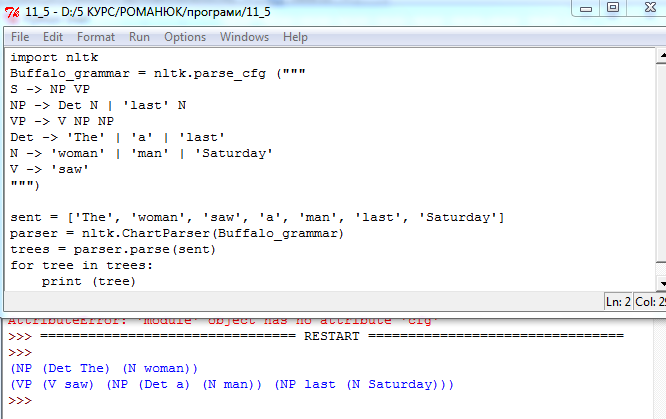


>>>t=nltk.Tree('(S((NP John)(VP(V saw)(NP Mary))))')

>>> t.draw()



5. Написати програму побудови дерев для речення  The woman saw a man last Thursday.



10. Здійснити аналіз послідовності слів: Buffalo buffalo Buffalo buffalo buffalo

buffalo Buffalo buffalo. Оскільки, згідно з

<http://en.wikipedia.org/wiki/Buffalo_buffalo_Buffalo_buffalo_buffalo_buffalo_Buff>

o\_buffalo це граматично правильне речення, напишіть контексно-вільну

граматику на основі дерева наведеного на цій сторінці з Інтернету. Здійсніть

нормалізацію слів (lowercase), для моделювання ситуації коли слухач сприймає

це речення на слух. Скільки дерев розбору може мати це дерево в такому

випадку?

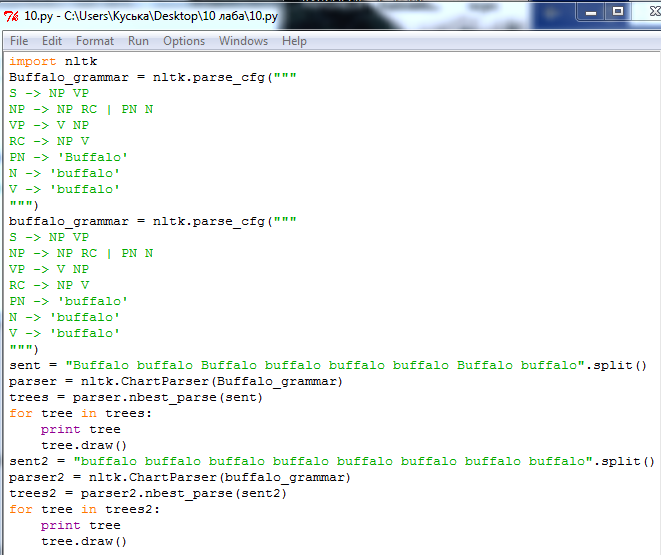


Рис.1 Виконання програми №10

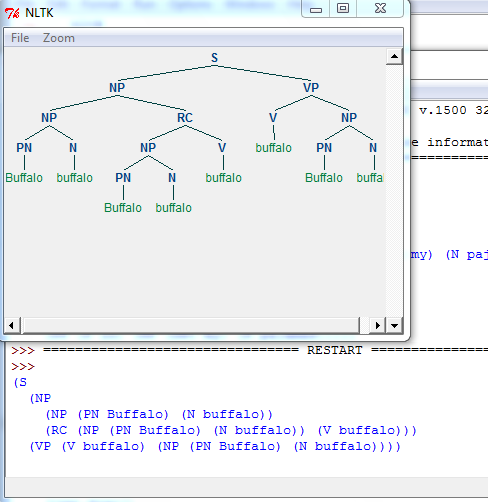


Рис.2 Результати виконання програми №10(а)

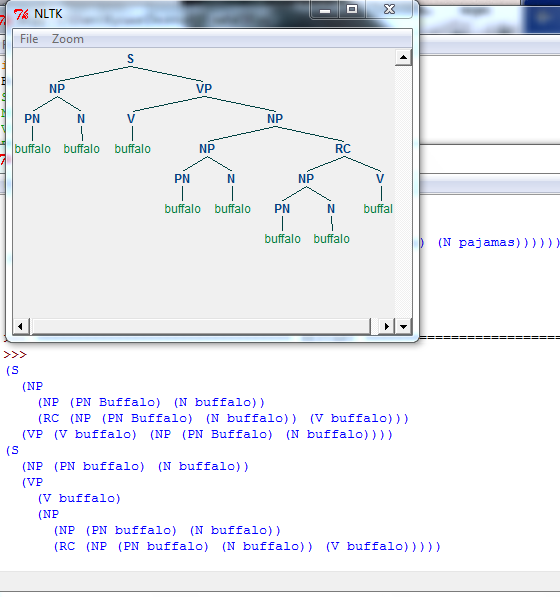


Рис.3 Результати виконання програми №10(б)

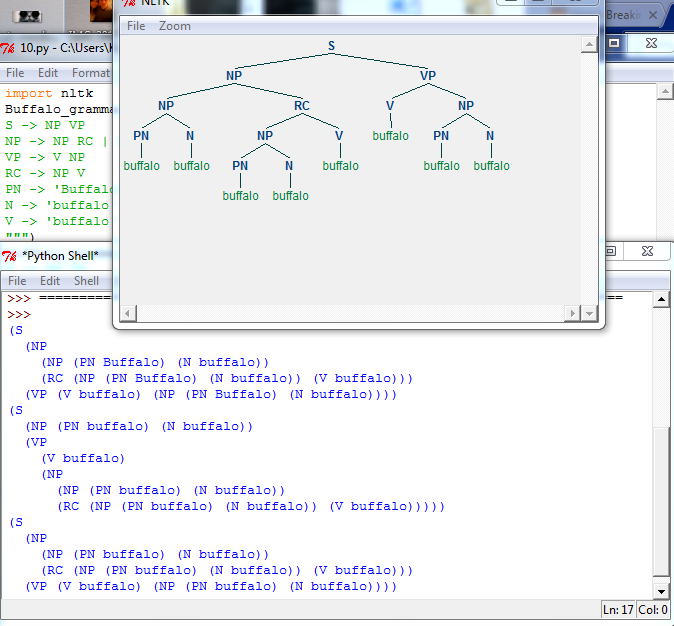


Рис.4 Результати виконання програми №10(в)

12. Написати програму порівняння швидкодії всіх аналізаторів, які згадувалися

в методичних. Використовувати timeit функцію для визначення часу

синтаксичного аналізу одного і того самого речення різними аналізаторами.

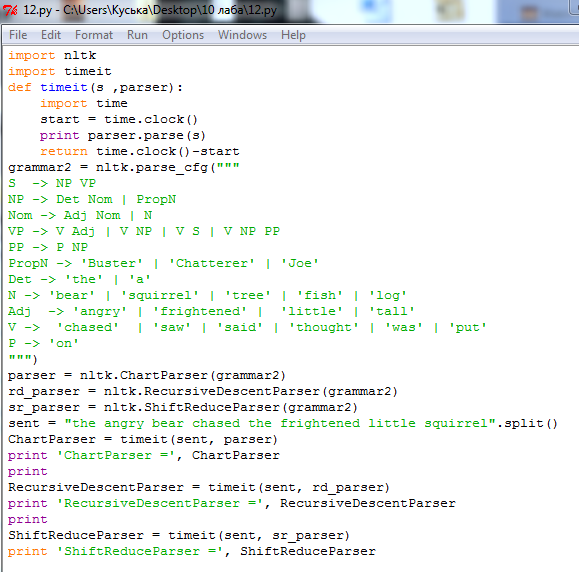


Рис.5 Текст програми №12

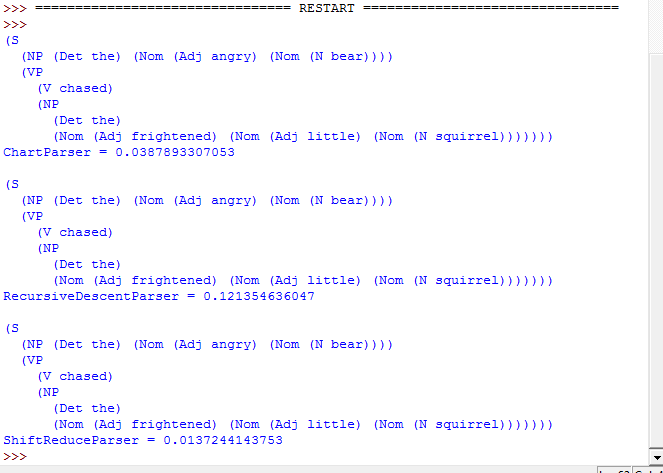


Рис.6 Результати виконання програми №12

13.Прочитати про "garden path" речення

http://en.wikipedia.org/wiki/Garden\_path\_sentence. Оцінити обчислювальну

складність аналізу таких речень в порівнянні з труднощами аналізу таких

речень людиною?

"Garden path sentences" - це граматично правильні речення, які починаються в такий спосіб, що в читачів тлумачення буде неправильним; вони здійснюють неправильний граматичний аналіз, який заводить їх в тупік. Ці речення використовуються в психолінгвістиці."Garden path" відноситься до висловлювання - "управляти садовою доріжкою" тобто "вводити в оману". Згідно з існуючою теорією в психолінгвістиці, коли читач читає  "garden path sentences" то він будує структуру значення одного слова в даний момент часу. В певний момент читачеві стає очевидно, що наступне слово чи фраза не може бути вставлено в структуру збудовану до цього часу: це не сумісно з доріжкою якою він управляє. Ці речення є поширені в аналітичній мові, де порядок в значній мірі залежить від встановлення граматичного відмінку в реченні.

Приклади "garden path sentence":

1) The author wrote the novel was likely to be a best-seller.

The author composed the novel...

The author wrote that the novel was likely to be a best-seller.

2) The man returned to his house was happy.

he man came back to his house...

Returned to his house, the man was happy.

3) The government plans to raise taxes were defeated.

The government is planning to raise taxes...

The government's plans to raise taxes were defeated.

Синтаксичний аналіз визначає як діляться фрази в "garden path sentences". Синтаксичний аналіз є загальним терміном, який використовується в психолінгвістиці описуючи розуміння мови. Людина скоріше(правильніше), ніж комп’ютер аналізує речення (визначає частини мови, синтаксичні зв’язки і т.д.)

Синтаксичний аналізатор, так як і людина визначає як поділити речення , але до одного і того ж речення можливо приписувати різні граматичні структури. Наприклад  “He ate the cookies on the couch,” це може означати, що він їсть печиво яке є на кушетці, або, що він сидить на кушетці в той час коли їсть печиво. Отже, як і людина так і машина може припуститися помилки, хоча людина правильніше аналізує речення.

**Висновок**: на цій лабораторній роботі я ознайомилася з автоматичним синтаксичним аналізом в NLTK, алгоритмами рекурсивного спуску(зверху-вниз) та переміщення-згортання (знизу-вверх), навчилися будувати дерева.