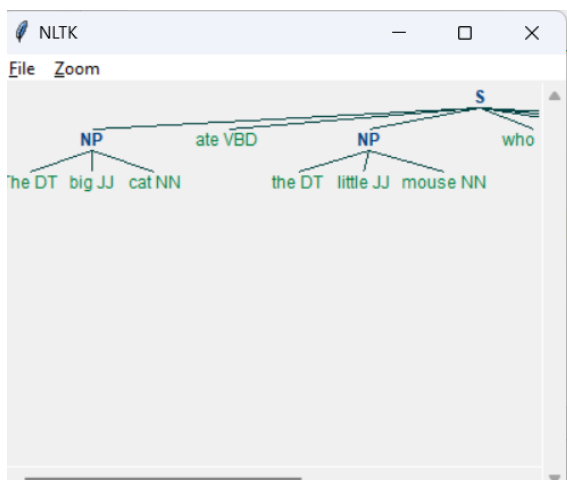


```
p12 x p111 x
C:\ALBINA\ml\venv\Scripts\python.exe C:/ALBINA/ml/p111.py
('welcome', 'to', 'amal')
('to', 'amal', 'jyothi')
('amal', 'jyothi', 'college')
('jyothi', 'college', 'of')
('college', 'of', 'engineering')

Process finished with exit code 0
```

```
p12 x
C:\ALBINA\ml\venv\Scripts\python.exe C:/ALBINA/ml/p12.py
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\hpl\AppData\Roaming\nltk_data...
[nltk_data] Unzipping tokenizers\punkt.zip.
['The', 'big', 'cat', 'ate', 'the', 'little', 'mouse', 'who', 'was', 'after', 'the', 'fresh', 'cheese']
[('The', 'DT'), ('big', 'JJ'), ('cat', 'NN'), ('ate', 'VBD'), ('the', 'DT'), ('little', 'JJ'), ('mouse', 'NN'), ('who', 'WP'), ('was', 'VBD'), ('after', 'IN'), ('the', 'DT'), ('fresh', 'JJ'), ('cheese', 'NN')]
(S
  (NP The/DT big/JJ cat/NN)
  ate/VBD
  (NP the/DT little/JJ mouse/NN)
  who/WP
  was/VBD
  after/IN
  (NP the/DT fresh/JJ cheese/NN))
```



Result

The program was executed and the result was successfully obtained. Thus CO5 was obtained.

Experiment No.: 12

Aim

Program to implement a simple web crawler

CO5

Implement programs for web data mining and natural language processing using NLTK

Procedure

import requests

from bs4 import BeautifulSoup

import csv

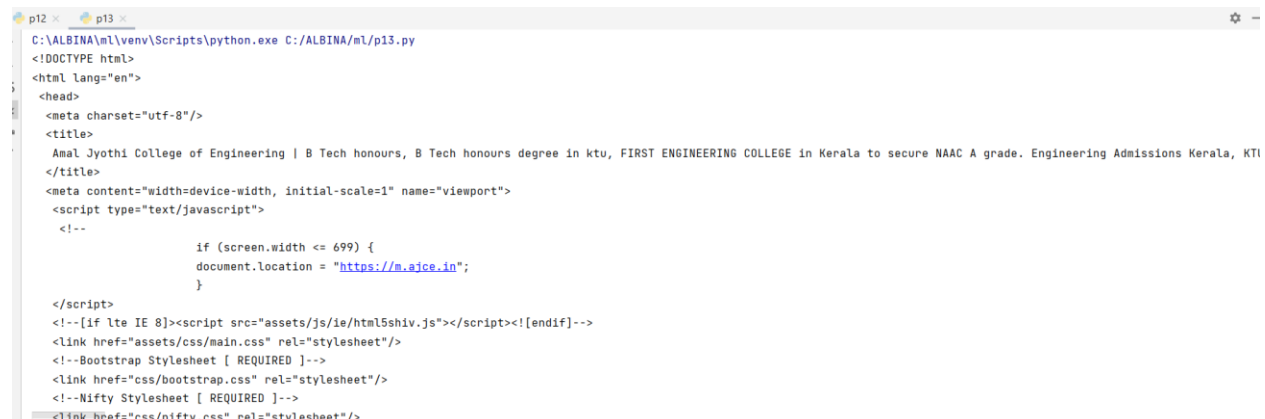
URL = "http://www.ajce.in"

r = requests.get(URL)

soup = BeautifulSoup(r.content, 'html.parser')

print(soup.prettify())

Output Screenshot



```
p12 x p13 x
C:\ALBINA\ml\venv\Scripts\python.exe C:/ALBINA/ml/p13.py
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="utf-8"/>
<title>
Amal Jyothi College of Engineering | B Tech honours, B Tech honours degree in ktu, FIRST ENGINEERING COLLEGE in Kerala to secure NAAC A grade. Engineering Admissions Kerala, KTI
</title>
<meta content="width=device-width, initial-scale=1" name="viewport">
<script type="text/javascript">
<!--
if (screen.width <= 699) {
document.location = "https://m.ajce.in";
}
</script>
<!--[if lte IE 8]><script src="assets/js/ie/html5shiv.js"></script><![endif]-->
<link href="assets/css/main.css" rel="stylesheet"/>
<!--Bootstrap Stylesheet [ REQUIRED ]-->
<link href="css/bootstrap.css" rel="stylesheet"/>
<!--Nifty Stylesheet [ REQUIRED ]-->
<link href="css/nifty.css" rel="stylesheet"/>
```

Result

The program was executed and the result was successfully obtained. Thus CO5 was obtained.