

Experiment No.: 12

Aim

Program to implement a simple web crawler program.

CO5

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

Procedure

import requests

from bs4 import BeautifulSoup

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

r = requests.get(URL)

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

print(soup)

Output Screenshot



```
<!DOCTYPE html>
<html itemscope="" itemtype="http://schema.org/WebPage" lang="en">
<head>
<meta content="Search the world's information, including webpages, images, videos and more. Google has many special features to help you find exactly what you"
<meta content="noodp" name="robots"/>
<meta content="text/html; charset=utf-8" http-equiv="Content-Type"/>
<meta content="/images/branding/googlelog/1x/googlelog_standard_color_128dp.png" itemprop="image"/>
<title>
Google
</title>
<script nonce="A3XVKUFHDD8TMTnbtCQ0tg">
(function(){window.google={kEI:'f3J4Y-KGLomYwbKpWbC_8As',kEXPI:'0,202563,1156846,6058,207,4804,2316,383,246,5,5367,1123753,1197787,380704,16114,28684,22431,
var f=this|self;var h,k=[];function l(a){for(var b;a&&(la.getAttribute(l(b=a.getAttribute("eid"))));a=a.parentNode;return b|h}function m(a){for(var b=null;a
function n(a,b,c,d,g){var e="";c||-1==b.search("&ei=")||e="&ei="+l(d),-1==b.search("&lei=")&&(d=m(d))&&(e+="&lei="+d));d="";!c&&f._csid&&-1===b.search("&cs
document.documentElement.addEventListener("submit",function(b){var a;if(a=b.target){var c=a.getAttribute("data-submitfalse");a="1"===c||"q"===c&&!a.elements.q.v
</script>
<style>
#gbg,#guser{font-size:13px;padding-top:1px !important;}#gbar{height:22px}#guser{padding-bottom:7px !important;text-align:right}.gbh,.gbd{border-top:1px sol
</style>
<style>
body,td,a,p,.h{font-family:arial,sans-serif}body{margin:0;overflow-y:scroll}#gog{padding:3px 8px 0}td{line-height:.8em}.gac_m td{line-height:17px}form{margin
</style>
<script nonce="A3XVKUFHDD8TMTnbtCQ0tg">
(function(){window.google.erd={jsr:1,bv:1691,de:true};
var h=this|self;var k,l=null!=(k=h.mei)?k:1,n,p=null!==(n=h.sdo)?n:!0,q=0,r,t=google.erd,v=t.jsr;google.ml=function(a,b,d,m,e){e=void 0===e?2:e;b&&(r=a&&a.messi
a.fileName;g&&(0<g.indexOf("-extension:/")&&(e=3),c+="%script="+b(g),f&&g===window.location.href&&(f=document.documentElement.outerHTML.split("\n")[f],c+="%cad
</script>
</head>
<body bgcolor="#fff">
```

Result

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