**Part 1:**

**I obtain my raw html by using my curl command and lynx command.**

```
atria:~/cs432/Asst3> python curlPrgrm.py
('Extracting source for ', 'http://www.pornalia.net')
('Extracting source for ', 'http://www.pornalia.net')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                 %   To    t a  l          % R  ec  ei  veDld oa% Xd fe Urdpl oa A
der  ag Te otSpal  e edSpe n t  T i  Lemeft      STipeedme
  T  im0 e      C0 ur  re 0n t
    0        0           0                0           0   - - Dlo:-ad-: - - 
-p-:loa-d-:  - -To ta--l:  -- :-Sp-e nt       0 Left  Speed
100    51  100    51    0     0    432      0 --:--:-- --:--:-- --:--:--   435
100    51  100    51    0     0    415      0 --:--:-- --:--:-- --:--:--   418
('Extracting source for ', 'http://www.patrickcines.com')
('Extracting source for ', 'http://www.sarawickham.com')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
('Extracting source for ', 'http://www.sarawickham.com')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 98809    0 98809    0     0  59987      0 --:--:--  0:00:01 --:--:-- 59993
100 98809    0 98809    0     0  23343      0 --:--:--  0:00:04 --:--:-- 23348
('Extracting source for ', 'http://www.lanzada.org')
('Extracting source for ', 'http://www.lanzada.org')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
                                 Dload  Upload   Total   Spent    Left  Speed
100 51554    0 51554    0     0  33408      0 --:--:--  0:00:01 --:--:-- 33411
('Extracting source for ', 'http://konogi.com/hen_1/')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100 51554    0 51554    0     0  31521      0 --:--:--  0:00:01 --:--:-- 31512
('Extracting source for ', 'http://konogi.com/hen_1/')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  5604  100  5604    0     0  10020      0 --:--:-- --:--:-- --:--:-- 10025
('Extracting source for ', 'http://Instagram.com/PINKLOVE0325')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  5604  100  5604    0     0  10303      0 --:--:-- --:--:-- --:--:-- 10320
('Extracting source for ', 'http://regio.newslocker.nl/gelderland/')
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
('Extracting source for ', 'http://www.geordy.nl')
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:--     0
                                 Dload  Upload   Total   Spent    Left  Speed
100   245  100   245    0     0   1249      0 --:--:-- --:--:-- --:--:--  1250
```

**Lynx:**

```
Extracting HTML data from: source705.txt
Extracting HTML data from: source606.txt
Extracting HTML data from: source967.txt
Extracting HTML data from: source59.txt
Extracting HTML data from: source1268.txt
Extracting HTML data from: source820.txt
Extracting HTML data from: source1458.txt
Extracting HTML data from: source806.txt
Extracting HTML data from: source1586.txt
Extracting HTML data from: source1433.txt
Extracting HTML data from: source192.txt
Extracting HTML data from: source1101.txt
Extracting HTML data from: source1039.txt
Extracting HTML data from: source1021.txt
Extracting HTML data from: source623.txt
Extracting HTML data from: source531.txt
Extracting HTML data from: source194.txt
Extracting HTML data from: source8.txt
Extracting HTML data from: source1576.txt
Extracting HTML data from: source1339.txt
Extracting HTML data from: source501.txt
Extracting HTML data from: source1608.txt
Extracting HTML data from: source68.txt
Extracting HTML data from: source366.txt
Extracting HTML data from: source1178.txt
Extracting HTML data from: source1070.txt
Extracting HTML data from: source1197.txt
Extracting HTML data from: source1385.txt
Extracting HTML data from: source1261.txt
Extracting HTML data from: source875.txt
Extracting HTML data from: source1419.txt
```

**Part 2:**

**For this section, I used the query term "politics" after using a grep command in order to find my links. For my IDF I multiplied 16 billion (amount of search results from Google) by 766,000,000 (amount of search results for the word politics). I found my TF by dividing the amount of times the word "politics" appeared on a website by the amount of words total on the link. Later I multiplied both the TF and IDF in order to obtain my TFIDF.**

| TFIDF | TF | IDF | URI |
|---|---|---|---|
| 0.00038 | 0.00008 | 4.48100 | http://blogos.com/ |
| 0.00403 | 0.00090 | 4.48100 | http://uzax.com/ |
| 0.00462 | 0.00103 | 4.48100 | http://www.express.co.uk |
| 0.01851 | 0.00413 | 4.48100 | http://artmatters.info |
| 0.00327 | 0.00073 | 4.48100 | http://www.goldmyne.tv |
| 0.00040 | 0.00009 | 4.48100 | http://defense-technologynews.blogspot.com/ |
| 0.00327 | 0.00073 | 4.48100 | http://www.mediagid.am |

| | | | |
|---|---|---|---|
| 0.00206 | 0.00046 | 4.48100 | http://mmb.moneycontrol.com |
| 0.02639 | 0.00589 | 4.48100 | http://www.cjscotland.co.uk |
| 0.00605 | 0.00135 | 4.48100 | http://www.abujafacts.ng |

**Part 3:**

**In order to find my page ranks I entered the website into** http://www.prchecker.info/check_page_rank.php.

| Page Rank | URI |
|---|---|
| 0.60000 | http://blogos.com/article/104352/ |
| 0.00000 | http://uzax.com/ |
| 0.70000 | http://www.express.co.uk |
| 0.40000 | http://artmatters.info |
| 0.10000 | http://www.goldmyne.tv |
| 0.30000 | http://defense-technologynews.blogspot.com/ |
| 0.50000 | http://www.mediagid.am |
| 0.50000 | http://mmb.moneycontrol.com |
| 0.40000 | http://www.cjscotland.co.uk |
| 0.00000 | http://www.abujafacts.ng |

**Part 4:**

**I was able to obtain my Kendall tau and "p" value by using my resources of this website http://www.wessa.net/rwasp_kendall.wasp#output , with this website I simply inputted my TFIDF as my X values and my page ranks as my Y values. After the website computed the values for me, which are highlighted below.**

| Kendall tau Rank Correlation | |
|---|---|
| Kendall tau | -0.209358960390091 |
| 2-sided p-value | 0.467181503772736 |
| Score | -9 |
| Var(Score) | 121.066665649414 |
| Denominator | 42.9883689880371 |