

## Part 1

Data for Part can be found on the other folders marked in the GitHub. However, they were done with a personal scripting python code that enabled both the “curl” command and “lynx –source” command. Can be seen below:

## Curl

```

C:\Program Files\Internet Explorer>python curlScriptKC.py
('Extracting source for ', 'https://github.com/bcoe/npm-tweets')
('Extracting source for ', 'http://tomaugspurger.github.io')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
  0     0    0     0    0     0      0     0  --:--:-- --:--:-- --:--:--  0
100 21416 100 21416  0     0 144k    0  --:--:-- --:--:-- --:--:-- 145k
100 59145  0 59145  0     0 378k    0  --:--:-- --:--:-- --:--:-- 379k
('Extracting source for ', 'http://www.CallBerb.com')
('Extracting source for ', 'http://www.stickam.jp/profile/ta30')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
  0     0    0     0    0     0      0     0  --:--:-- --:--:-- --:--:--  0
100   608 100   608  0     0 2515    0  --:--:-- --:--:-- --:--:-- 2522
('Extracting source for ', 'https://www.facebook.com/BebesticMultiMedia')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
100  268k  0  268k  0     0 156k    0  --:--:-- 0:00:01 --:--:-- 156k
100 43327  0 43327  0     0 21813   0  --:--:-- 0:00:01 --:--:-- 21805
('Extracting source for ', 'http://www.greatcareers.co/blog')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
  0     0    0     0    0     0      0     0  --:--:-- --:--:-- --:--:--  0
('Extracting source for ', 'https://www.sinsweety.co.uk')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
100 77244  0 77244  0     0 73234   0  --:--:-- 0:00:01 --:--:-- 73286
100 55042 100 55042  0     0 50898   0  --:--:-- 0:00:01 --:--:-- 50917
('Extracting source for ', 'http://germanlinux.blogspot.com')
('Extracting source for ', 'http://sylmercury.tumblr.com')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
  0     0    0     0    0     0      0     0  --:--:-- --:--:-- --:--:--  0
('Extracting source for ', 'http://www.mundo5ql.es')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
100 120k  0 120k  0     0 179k    0  --:--:-- --:--:-- --:--:-- 179k
('Extracting source for ', 'http://www.flexsource.ie/')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
100 48932 100 48932  0     0 40685   0  --:--:-- 0:00:01 --:--:-- 40708
100 88653  0 88653  0     0 34704   0  --:--:-- 0:00:02 --:--:-- 34697
('Extracting source for ', 'http://stores.ebay.co.uk/Ruby-redsky')
('Extracting source for ', 'http://katsu-tea.tumblr.com/')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
  0     0    0     0    0     0      0     0  --:--:-- --:~:~:~ --:~:~:~  0
100 50594  0 50594  0     0 56609   0  --:~:~:~ --:~:~:~ --:~:~:~ 56592
100 137k 100 137k  0     0 63256   0  --:~:~:~ 0:00:02 --:~:~:~ 63266
('Extracting source for ', 'http://unpocodetodo.info')
('Extracting source for ', 'http://www.collinaitaliana.com')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed
  0     0    0     0    0     0      0     0  --:~:~:~ --:~:~:~ --:~:~:~  0
100 14194 100 14194  0     0 36382   0  --:~:~:~ --:~:~:~ --:~:~:~ 36394
('Extracting source for ', 'http://ignitertechnical.com')
% Total    % Received % Xferd  Average Speed   Time    Time     Current
                                 Dload  Upload    Total   Spent    Left     Speed

```

## LYNX

```
atria.cs.odu.edu - PuTTY
atria:~/CS432/A3> python lynxScript.py
Extracting HTML data from: source182.txt
Extracting HTML data from: source3.txt

Can't Access `file:///localhost/home/mgomez/CS432/A3/source182.txt'
Alert!: Unable to access document.

lynx: Can't access startfile
Extracting HTML data from: source300.txt
Extracting HTML data from: source125.txt
```

## Part 2 – TFIDF Calculations

TFIDF	TF	IDF	URI
.26869	.00407	66.01813	<a href="http://www.thepakmedia.com/iphone-5se-and-iphone-7-concepts-video/2016/02/19">http://www.thepakmedia.com/iphone-5se-and-iphone-7-concepts-video/2016/02/19</a>
1.46560	.02220	66.01813	<a href="http://rollosphotos.com/shop/iphone+cases">http://rollosphotos.com/shop/iphone+cases</a>
1.23982	.01878	66.01813	<a href="http://www.gadgetoid.com/channels/phones/">http://www.gadgetoid.com/channels/phones/</a>
.06668	.00101	66.01813	<a href="http://www.ameanmagazine.com/search/label/phonegap">http://www.ameanmagazine.com/search/label/phonegap</a>
.09903	.00150	66.01813	<a href="file:///il-en/c/bags-and-purses/electronics-cases/phone-cases?ref=catnav-2938">file:///il-en/c/bags-and-purses/electronics-cases/phone-cases?ref=catnav-2938</a>
.09771	.00148	66.01813	<a href="http://robbgear.robreport.com/stuff/quad-pa-one-headphone-amplifier">http://robbgear.robreport.com/stuff/quad-pa-one-headphone-amplifier</a>
.05281	.0008	66.01813	<a href="http://filemakertoday.com/filemaker-13/4035-filemaker-13/caller-id-looks-up-filemaker-client-record-with-voip-telephone">http://filemakertoday.com/filemaker-13/4035-filemaker-13/caller-id-looks-up-filemaker-client-record-with-voip-telephone</a>
1.45636	.02206	66.01813	<a href="file:///walmartlabs-wmusiphone/ios-shared/pull/110">file:///walmartlabs-wmusiphone/ios-shared/pull/110</a>
.08978	.00136	66.01813	<a href="http://cssmatter.com/blog/responsive-website-on-iphone-unwanted-white-space-on-rotate-from-landscape-to-portrait/">http://cssmatter.com/blog/responsive-website-on-iphone-unwanted-white-space-on-rotate-from-landscape-to-portrait/</a>
.07196	.00109	66.01813	<a href="http://dinuguan.tumblr.com/post/139454880377/what-happens-when-you-boil-iphone">http://dinuguan.tumblr.com/post/139454880377/what-happens-when-you-boil-iphone</a>

### **Part 3 – Page Ranks**

Page Rank	URI
0	<a href="http://www.thepakmedia.com/iphone-5se-and-iphone-7-concepts-video/2016/02/19">http://www.thepakmedia.com/iphone-5se-and-iphone-7-concepts-video/2016/02/19</a>
.03	<a href="http://rollospotos.com/shop/iphone+cases">http://rollospotos.com/shop/iphone+cases</a>
.01	<a href="http://www.gadgetoid.com/channels/phones/">http://www.gadgetoid.com/channels/phones/</a>
0	<a href="http://www.ameanmagazine.com/search/label/phonegap">http://www.ameanmagazine.com/search/label/phonegap</a>
.03	<a href="https://www.twilio.com/blog/2016/02/how-to-verify-phone-numbers-in-python-with-the-twilio-lookup-api.html">https://www.twilio.com/blog/2016/02/how-to-verify-phone-numbers-in-python-with-the-twilio-lookup-api.html</a>
0	<a href="http://robbgear.robreport.com/stuff/quad-pa-one-headphone-amplifier">http://robbgear.robreport.com/stuff/quad-pa-one-headphone-amplifier</a>
.03	<a href="http://filemakertoday.com/filemaker-13/4035-filemaker-13/caller-id-looks-up-filemaker-client-record-with-voip-telephone">http://filemakertoday.com/filemaker-13/4035-filemaker-13/caller-id-looks-up-filemaker-client-record-with-voip-telephone</a>
0	<a href="http://www.webmastersid.com/2014/10/google-smartphone-mobile-friendly.html">http://www.webmastersid.com/2014/10/google-smartphone-mobile-friendly.html</a>
0	<a href="http://cssmatter.com/blog/responsive-website-on-iphone-unwanted-white-space-on-rotate-from-landscape-to-portrait/">http://cssmatter.com/blog/responsive-website-on-iphone-unwanted-white-space-on-rotate-from-landscape-to-portrait/</a>
.02	<a href="http://dinuguan.tumblr.com/post/139454880377/what-happens-when-you-boil-iphone">http://dinuguan.tumblr.com/post/139454880377/what-happens-when-you-boil-iphone</a>

### **Part 4 – Extra Credit – Kendall Calculations**

Kendall Tau Rank Correlation

TDIF multiplied with Page Rank (decimal)

Kendall Tau = 0

2-sided p-value = 1

Kendall tau	0
2-sided p-value	1