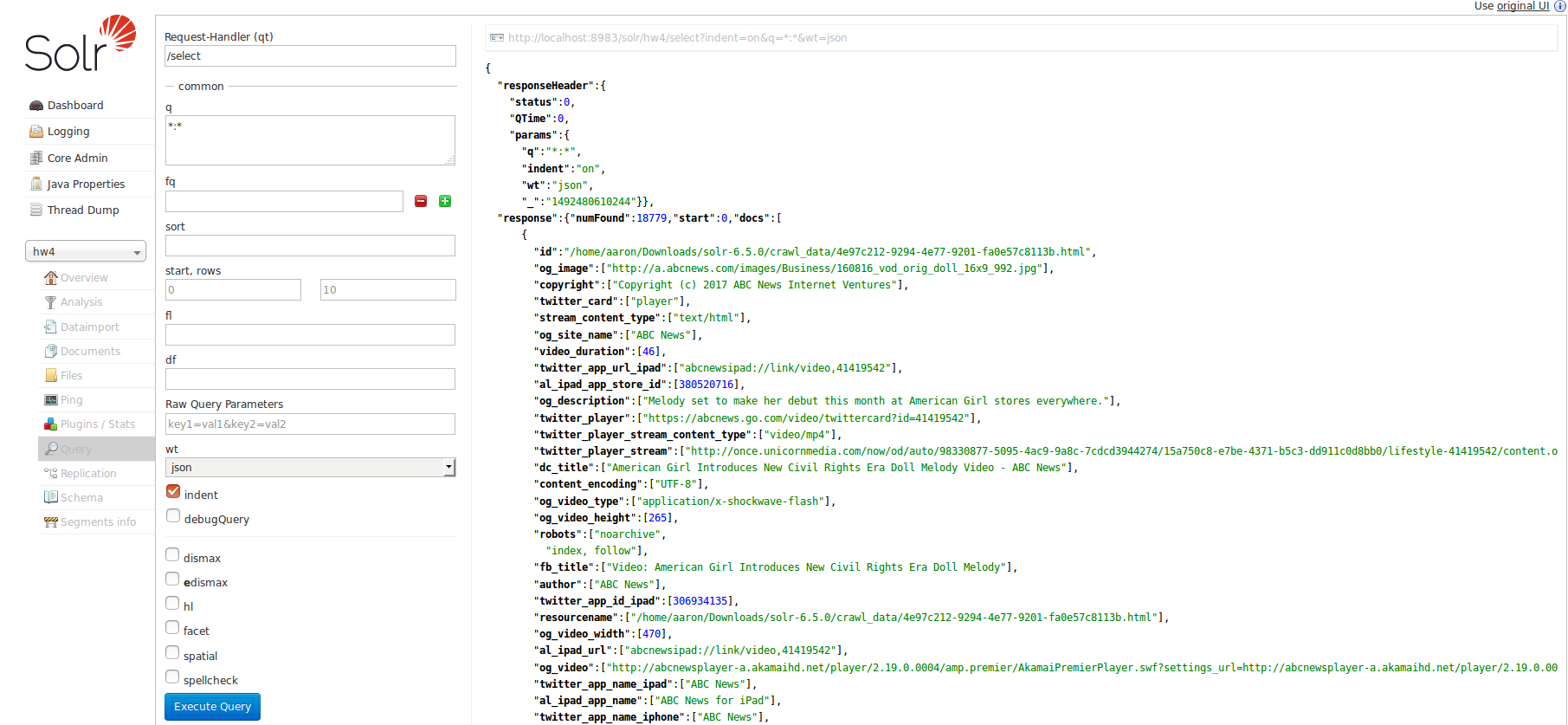
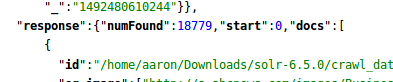
**CSCI572 Homework 4 Report**

**Name: Shurui Liu, 9999-9586-70**

**I resubmitted this report on April 18 after getting email permission from the professor.**

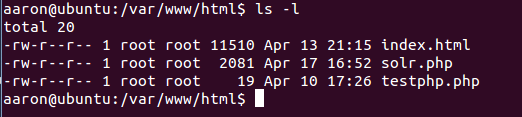
1. **Steps**
2. The first step is to use Solr to index all the HTML files that assigned to me. I am responsible for **ABCNews**. This screenshot shows you there are **18779** files are indexed by Solr.



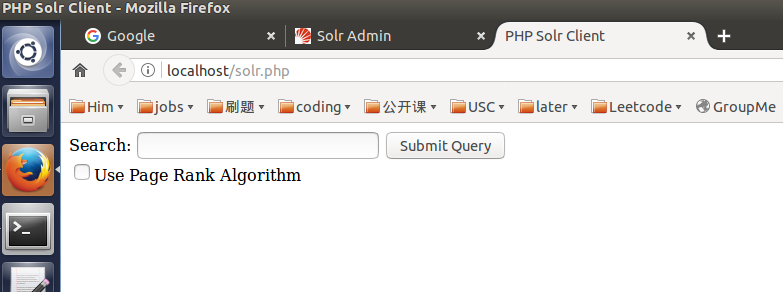


Before indexing these files, I made proper changes in **managed-schema** and **solrconfig.xml** based on the tutorial we got from 572 website.

1. The second step is setup the PHP page on my Apache server. I place my code, **solr.php**, at this location like the screenshot below. **/var/www/html**



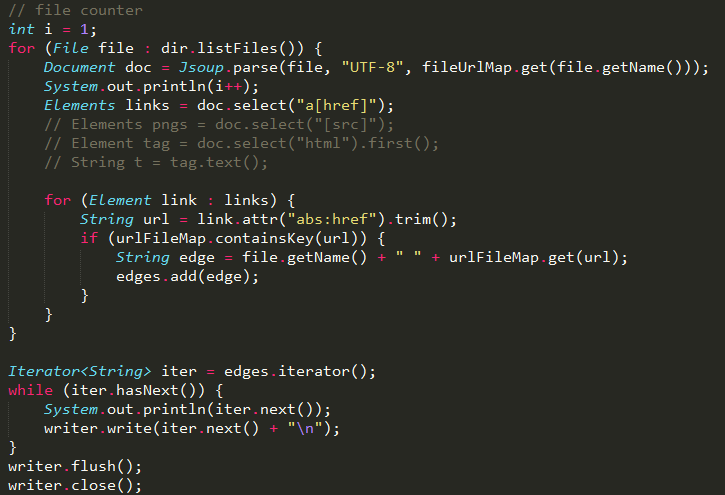
When both Apache and Solr are on, I have this page below.



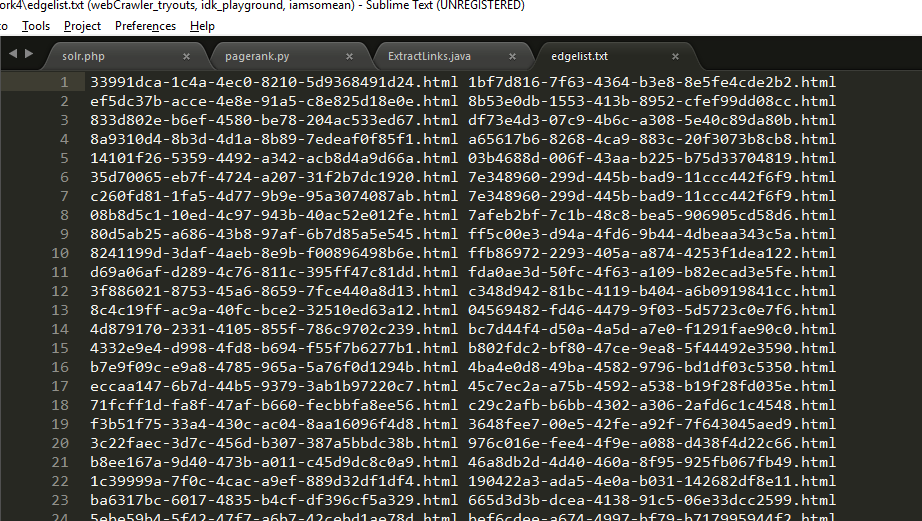
1. Then I write a Java code to calculate the edgelist and save it as a text file called edgelist.txt.

The Java code uses a library called Jsoup to calculate the edgelist. This screenshot below shows the critical part of the code.

For variable ***url***, we have to use **abs:href** for **link.attr**.

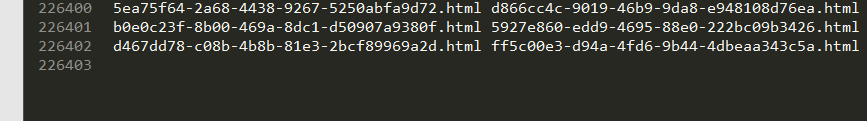


After running this code, I got an edgelist text file. Here is a screenshot of what it looks like.



The file means that there is outgoing links from the HTML code on the **left** side point to the HTML code on the **right** side.

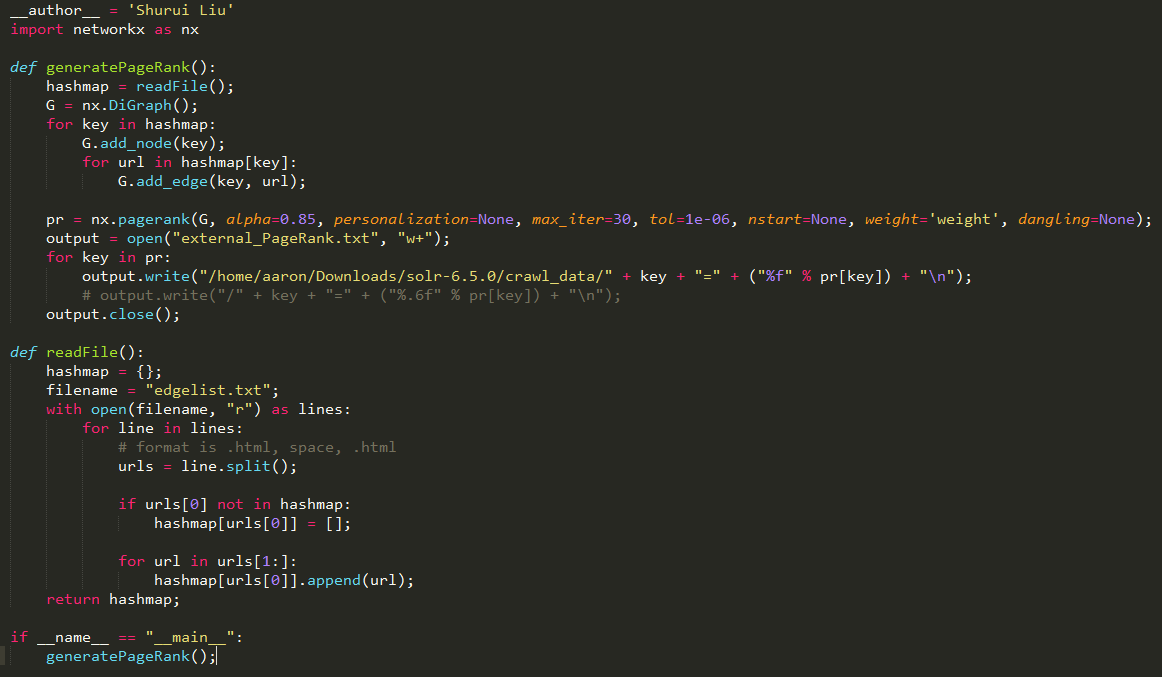
The number of HTML files is **18779**, and the number of edges in this edgelist.txt file is **226402**.

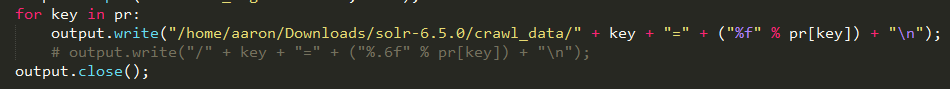


1. Based on the tutorial we got for the homework assignment, here is how I computed the pagerank.

***pr = nx.pagerank(G, alpha=0.85, personalization=None, max\_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None);***

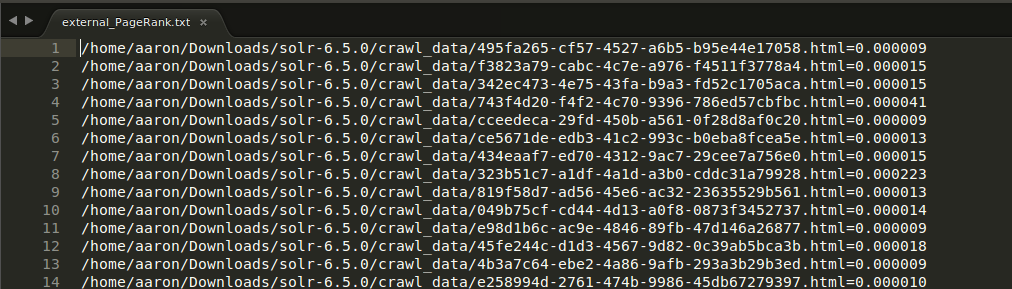
alpha is 0.85, personalization is None, max iteration time is 30, tolerance (decimal number) is 1e-06, nstart is None, weight is ‘weight’, and dangling is None. Code is below. Notice that I attached the crawl\_data folder path in the beginning of the file name. This is the way for Solr to be able to find those files in its folder on my computer.





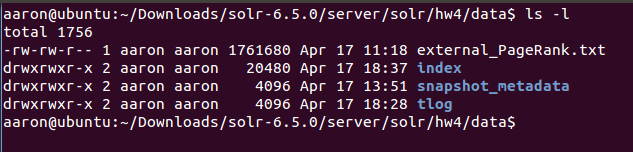
After running this code, I got a text file called **external\_pageRankFile.txt**, which is what we need to import into Solr for the PageRank algorithm. Here’s what it looks like below. The format is

**/filepath/filename.html=0.xxx**

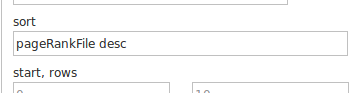


1. After having this **external\_pageRankFile.txt** file, I place it at here, **solr-6.5.0/server/solr/hw4/data**

Notice that **hw4** is my core name.



1. Then going back to the Solr UI, make a query with this change to apply the PageRank algorithm.



1. **Query results (normal Solr results first, then pageRank result)**

The way I check the overlap is to see whether there is any files have the **same size** from both two sets of results.

* **Brexit, 1 overlap**

Results 1 - 10 of 132:

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 46271

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 51693

Document Tony Blair's new mission is to change minds on Brexit

Author: ABC News | Size: 25167

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 43041

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 42704

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 42701

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 42761

Document Brexit European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 42761

Document AP Interview: Sweden PM: Brexit deal in 2 years 'very tough'

Author: ABC News | Size: 28917

**Document Tony Blair's new mission is to change minds on Brexit - ABC News**

**Author: ABC News | Size: 76964**

Results 1 - 10 of 132:

Document ABC News � Breaking News, Latest News, Headlines & Videos

Author: ABC News | Size: 187636

Document International News | Latest World News, Videos & Photos -ABC News - ABC News

Author: ABC News | Size: 183777

Document World News, Photos and Videos - ABC News

Author: None | Size: 141400

Document World News, Photos and Videos - ABC News

Author: None | Size: 141396

Document AP Interview: Sweden PM: Brexit deal in 2 years 'very tough' - ABC News

Author: ABC News | Size: 80049

Document Browse Topics that begin with B

Author: None | Size: 155155

**Document Tony Blair's new mission is to change minds on Brexit - ABC News**

**Author: ABC News | Size: 76964**

Document European Union Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 44859

Document London Stock Exchange Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48353

Document Unemployment Rate Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 45364

* **NASDAQ, 0 overlaps**

Results 1 - 10 of 67:

Document NASDAQ Composite Index Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 43866

Document NASDAQ 100 Index Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50615

Document NASDAQ Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 44503

Document Indexes News, Photos and Videos - ABC News

Author: None | Size: 29670

Document How major US stock market indexes fared on Friday

Author: ABC News | Size: 23618

Document Markets Right Now: Late push renews records for US indexes

Author: ABC News | Size: 23199

Document Stocks Party Like It�s 2000 During Record Highs - ABC News

Author: ABC News | Size: 76232

Document Russell 2000 Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 43605

Document Triple-Digit Decline for Dow Jones Industrials - ABC News

Author: ABC News | Size: 98595

Document Facebook Shares Back at the IPO Price - ABC News

Author: ABC News | Size: 77900

Results 1 - 10 of 67:

Document Stock.exchange News, Photos and Videos - ABC News

Author: None | Size: 31458

Document Facebook unveils job postings feature Video - ABC News

Author: ABC News | Size: 64222

Document London Stock Exchange Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48353

Document Bear Market Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50279

Document Flash Crash Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48756

Document Canadian Imperial Bank of Commerce Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 46569

Document Browse Topics that begin with N

Author: None | Size: 105817

Document Asian shares slip after Wall Street fall and Samsung arrest - ABC News

Author: ABC News | Size: 76523

Document Heartbleed Bug Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 47313

Document Xiaomi Inc. Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48052

* **NBA, 0 overlaps**

Results 1 - 10 of 526:

Document National Basketball Association NBA

Author: None | Size: 51512

Document NBA warns Texas over proposed 'bathroom bill'

Author: ABC News | Size: 20113

Document Stephen Curry, Warriors still top NBA merchandise sales - ABC News

Author: ABC News | Size: 69263

Document #NBArank Rising Stars: Your list of the NBA's top players under 25 - ABC News

Author: ABC News | Size: 68628

Document #NBArank Rising Stars: Your list of the NBA's top players under 25 - ABC News

Author: ABC News | Size: 68630

Document #NBArank Rising Stars: Your list of the NBA's top players under 25 - ABC News

Author: ABC News | Size: 68630

Document National Basketball Association NBA

Author: None | Size: 46904

Document National Basketball Association NBA

Author: None | Size: 46847

Document National Basketball Association NBA

Author: None | Size: 46104

Document National Basketball Association NBA

Author: None | Size: 46160

Results 1 - 10 of 526:

Document ABC News � Breaking News, Latest News, Headlines & Videos

Author: ABC News | Size: 187636

Document Sports News & Articles � Scores, Pictures, Videos - ABC News

Author: ABC News | Size: 72701

Document Sports News, Photos and Videos - ABC News

Author: None | Size: 28148

Document National Football League NFL

Author: None | Size: 45851

Document National Basketball Association NBA

Author: None | Size: 46332

Document The revival of Mahmoud Abdul-Rauf - ABC News

Author: ABC News | Size: 92097

Document Steve Kerr keeping 'Warriors + 1' All-Star choice close to vest - ABC News

Author: ABC News | Size: 71255

Document Trade talk: Best deadline deals for Cavs, Bulls, Bucks, Pacers, Pistons - ABC News

Author: ABC News | Size: 81408

Document Despite Pelicans' struggles, Anthony Davis 'here to stay' - ABC News

Author: ABC News | Size: 68054

Document NBA joins NFL in warning Texas over proposed 'bathroom bill' - ABC News

Author: ABC News | Size: 64450

* **Snapchat, 1 overlap**

Results 1 - 10 of 95:

Document Snapchat Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48462

Document Snapchat Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50407

**Document Snapchat Videos at ABC News Video Archive at abcnews.com**

**Author: None | Size: 50704**

Document Snapchat Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50406

Document Snapchat Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50409

Document Snapchat Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50405

Document Snapchat Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50849

Document Snapchat Filters Among Top Halloween Costume Trends of 2016 - ABC News

Author: ABC News | Size: 79753

Document Police: 200 Descend on Philadelphia Mall in 2nd Night of Violence; Fracas Organized on Snapchat - ABC News

Author: ABC News | Size: 78049

Document Snapchat Files for IPO Video - ABC News

Author: ABC News | Size: 65517

Results 1 - 10 of 95:

Document Technology News, Photos and Videos - ABC News

Author: None | Size: 37899

Document Technology News, Photos and Videos - ABC News

Author: None | Size: 37903

Document ARCHIVAL VIDEO: Meet Siri, Apple's Latest Invention Video - ABC News

Author: ABC News | Size: 65262

Document Cannes Film Festival

Author: None | Size: 52079

Document Two teenage girls killed while hiking in Indiana Video - ABC News

Author: ABC News | Size: 65894

Document Browse Topics that begin with S

Author: None | Size: 184449

Document Larry Birkhead on relationship with Anna Nicole Smith, raising their daughter Video - ABC News

Author: ABC News | Size: 73892

Document Food Computer: How Caleb Harper Inspires Us to Be Future Farmers Video - ABC News

Author: ABC News | Size: 66495

**Document Snapchat Videos at ABC News Video Archive at abcnews.com**

**Author: None | Size: 50704**

Document Finland Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 49431

* **Illegal Immigration, 0 overlaps**

Results 1 - 10 of 1694:

Document Illegal Immigration Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 45964

Document Illegal Aliens Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 46763

Document Arizona Immigration Law: Judge Susan Bolton Blocks Parts of SB 1070 - ABC News

Author: ABC News | Size: 77985

Document Nationality Law Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48561

Document Dying Dad Jesus Navarro Denied Kidney Transplant Over Immigration Status, but Supporters Try to Help - ABC News

Author: ABC News | Size: 79729

Document Dying Dad, Jesus Navarro, to Get Kidney Transplant Despite Undocumented Immigration Status - ABC News

Author: ABC News | Size: 76962

Document Immigration News, Photos and Videos - ABC News

Author: None | Size: 30757

Document Immigration News, Photos and Videos - ABC News

Author: None | Size: 30761

Document Judge Susan Bolton Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 46864

Document Businesses nationwide participate in Day Without Immigrants protest

Author: ABC News | Size: 27907

Results 1 - 10 of 1694:

Document News Topics | Hot Topics In The News - ABC News

Author: None | Size: 46197

Document This Week With George Stephanopoulos - ABC News

Author: ABC News | Size: 149556

Document Nightline: Late Evening News - ABC News

Author: ABC News | Size: 153176

Document U.S. News | Latest National News, Videos & Photos - ABC News - ABC News

Author: ABC News | Size: 210415

Document World News Tonight With David Muir - ABC News

Author: ABC News | Size: 149698

Document International News | Latest World News, Videos & Photos -ABC News - ABC News

Author: ABC News | Size: 183777

Document ABC News Videos & Live News Clips Online - ABC News

Author: ABC News | Size: 326272

Document Arizona to death-row inmates: Bring your own execution drugs - ABC News

Author: ABC News | Size: 79629

Document Russell Simmons teams up with rabbi, imam for anti-Trump rally dubbed 'I Am A Muslim, Too' - ABC News

Author: ABC News | Size: 77812

Document Latest Weather Today

Author: None | Size: 52664

* **Donald Trump, 0 overlaps**

Results 1 - 10 of 10141:

Document Donald Trump Jr. Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 48293

Document Donald Trump's Inauguration in Photos Photos - ABC News

Author: ABC News | Size: 122105

Document Tiffany Trump: News, Videos and Photos-ABC News

Author: None | Size: 51537

Document The Moment Donald Trump and His Family Knew He Won the Election - ABC News

Author: ABC News | Size: 78493

Document Trump's First 100 Days Photos - ABC News

Author: ABC News | Size: 125591

Document Donald Trump Latest News and Videos

Author: None | Size: 50757

Document ANALYSIS: Donald Trump Sees Calm After Storm in First DC Trip as President-Elect - ABC News

Author: ABC News | Size: 72805

Document GOP Strategist Likely Faces Hurdles in Defamation Suit Against Donald Trump - ABC News

Author: ABC News | Size: 74133

Document President Donald Trump Vows to Launch 'Major Investigation' Into Alleged Voter Fraud - ABC News

Author: ABC News | Size: 77476

Document OPINION: What Type of President Will Donald Trump Be - Lincoln or Buchanan? - ABC News

Author: ABC News | Size: 72281

Results 1 - 10 of 10141:

Document News, Photos and Videos - ABC News

Author: None | Size: 257947

Document This Week With George Stephanopoulos - ABC News

Author: ABC News | Size: 149556

Document 20/20 | Investigative Journalism & News Magazine - ABC News

Author: ABC News | Size: 132562

Document ABC News � Breaking News, Latest News, Headlines & Videos

Author: ABC News | Size: 187636

Document Entertainment News |Latest Celebrity News, Videos & Photos - ABC News - ABC News

Author: ABC News | Size: 223518

Document Nightline: Late Evening News - ABC News

Author: ABC News | Size: 153176

Document U.S. News | Latest National News, Videos & Photos - ABC News - ABC News

Author: ABC News | Size: 210415

Document World News Tonight With David Muir - ABC News

Author: ABC News | Size: 149698

Document Technology and Science News - ABC News

Author: ABC News | Size: 177600

Document How do I contact ABCNews.com? - ABC News

Author: ABC News | Size: 47112

* **Russia, 0 overlaps**

Results 1 - 10 of 1553:

Document Top Trump envoys signal no quick changes to US-Russia ties

Author: ABC News | Size: 28396

Document US secretary of state says Russia must honor Ukraine deal

Author: ABC News | Size: 25419

Document Russia Sanctions Explained: A Look at the Measures Donald Trump Could Roll Back - ABC News

Author: ABC News | Size: 81310

Document Russia Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 47345

Document US official says Russia deployed missile in treaty violation - ABC News

Author: ABC News | Size: 73777

Document What we know - and don't know - about the Trump team's contacts with Russia before the election - ABC News

Author: ABC News | Size: 78070

Document Nikki Haley Blasts Russia in First Remarks at UN Security Council - ABC News

Author: ABC News | Size: 78046

Document Trump: Reports that campaign had contact with Russia are 'fake news'

Author: ABC News | Size: 27173

Document EU Diplomats Back Continued Russia Sanctions - ABC News

Author: ABC News | Size: 75468

Document Danes See Russia as a 'Leader' With 'Advanced Capabilities' - ABC News

Author: ABC News | Size: 65616

Results 1 - 10 of 1553:

Document News, Photos and Videos - ABC News

Author: None | Size: 257947

Document This Week With George Stephanopoulos - ABC News

Author: ABC News | Size: 149556

Document Investigative Reporting | Blotter - ABC News

Author: ABC News | Size: 61099

Document ABC News � Breaking News, Latest News, Headlines & Videos

Author: ABC News | Size: 187636

Document Nightline: Late Evening News - ABC News

Author: ABC News | Size: 153176

Document World News Tonight With David Muir - ABC News

Author: ABC News | Size: 149698

Document Technology and Science News - ABC News

Author: ABC News | Size: 177600

Document International News | Latest World News, Videos & Photos -ABC News - ABC News

Author: ABC News | Size: 183777

Document ABC News Videos & Live News Clips Online - ABC News

Author: ABC News | Size: 326272

Document Trump hits the road, hoping rally speeches overcome troubles - ABC News

Author: ABC News | Size: 73501

* **NASA, 0 overlaps**

Results 1 - 10 of 4937:

Document NASA

Author: None | Size: 47764

Document Incredible Images Captured From Space Photos - ABC News

Author: ABC News | Size: 227987

Document Picture | Incredible Images Captured From Space - ABC News

Author: ABC News | Size: 228899

Document NASA Targets New Date for Mars InSight Drilling Mission - ABC News

Author: ABC News | Size: 65094

Document Two Earth-Size Planets Found by NASA Kepler Mission - ABC News

Author: ABC News | Size: 68683

Document Pluto's Moon Charon Had a 'Hulk' Period, New NASA Images Reveal - ABC News

Author: ABC News | Size: 66624

Document NASA Wants to Blast Your Art Work Into Space - ABC News

Author: ABC News | Size: 64284

Document Kepler Space Telescope Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 51246

Document Mars Exploration Videos at ABC News Video Archive at abcnews.com

Author: None | Size: 50943

Document Breathtaking Views of Earth From Above Photos - ABC News

Author: ABC News | Size: 100529

Results 1 - 10 of 4937:

Document News, Photos and Videos - ABC News

Author: None | Size: 257947

Document Live Streaming News Video - ABC News - ABC News

Author: ABC News | Size: 49610

Document Live Streaming News Video - ABC News - ABC News

Author: ABC News | Size: 49644

Document This Week With George Stephanopoulos - ABC News

Author: ABC News | Size: 149556

Document Investigative Reporting | Blotter - ABC News

Author: ABC News | Size: 61099

Document Live Streaming News Video - ABC News - ABC News

Author: ABC News | Size: 49589

Document 20/20 | Investigative Journalism & News Magazine - ABC News

Author: ABC News | Size: 132562

Document ABC News � Breaking News, Latest News, Headlines & Videos

Author: ABC News | Size: 187636

Document Entertainment News |Latest Celebrity News, Videos & Photos - ABC News - ABC News

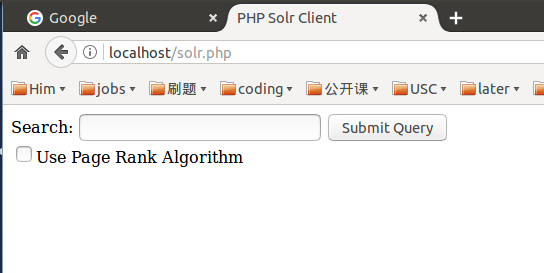
Author: ABC News | Size: 223518

Document Nightline: Late Evening News - ABC News

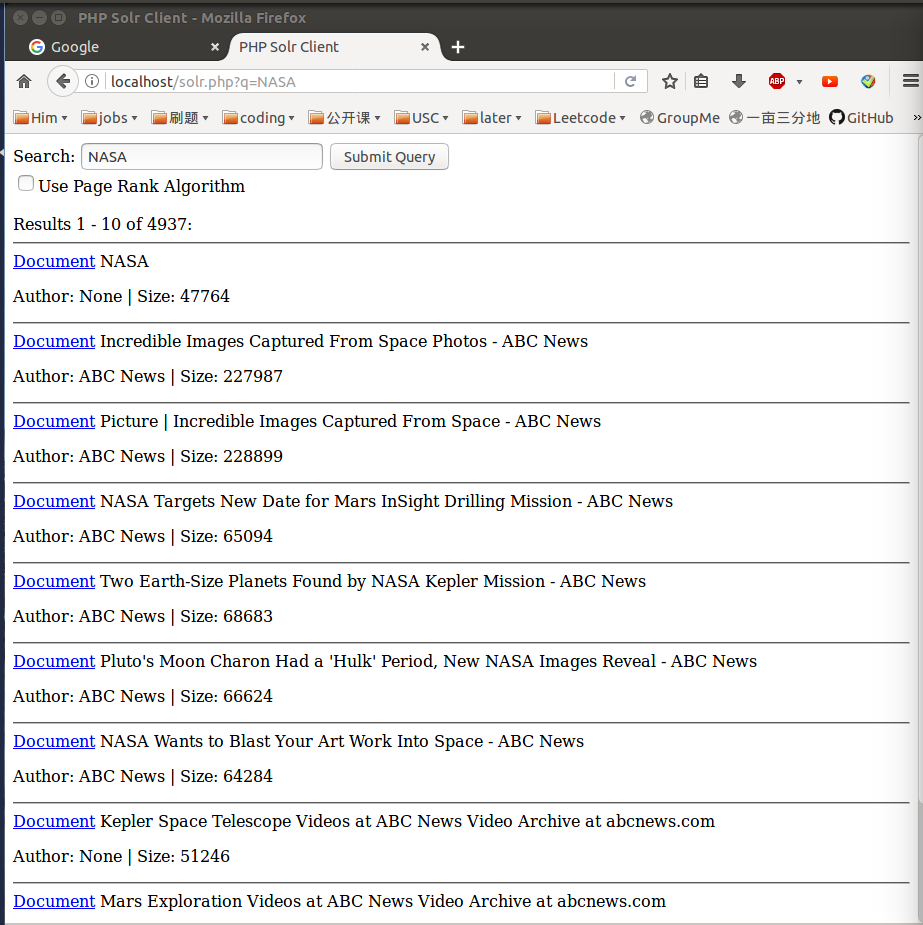
Author: ABC News | Size: 153176

1. **Screenshot**

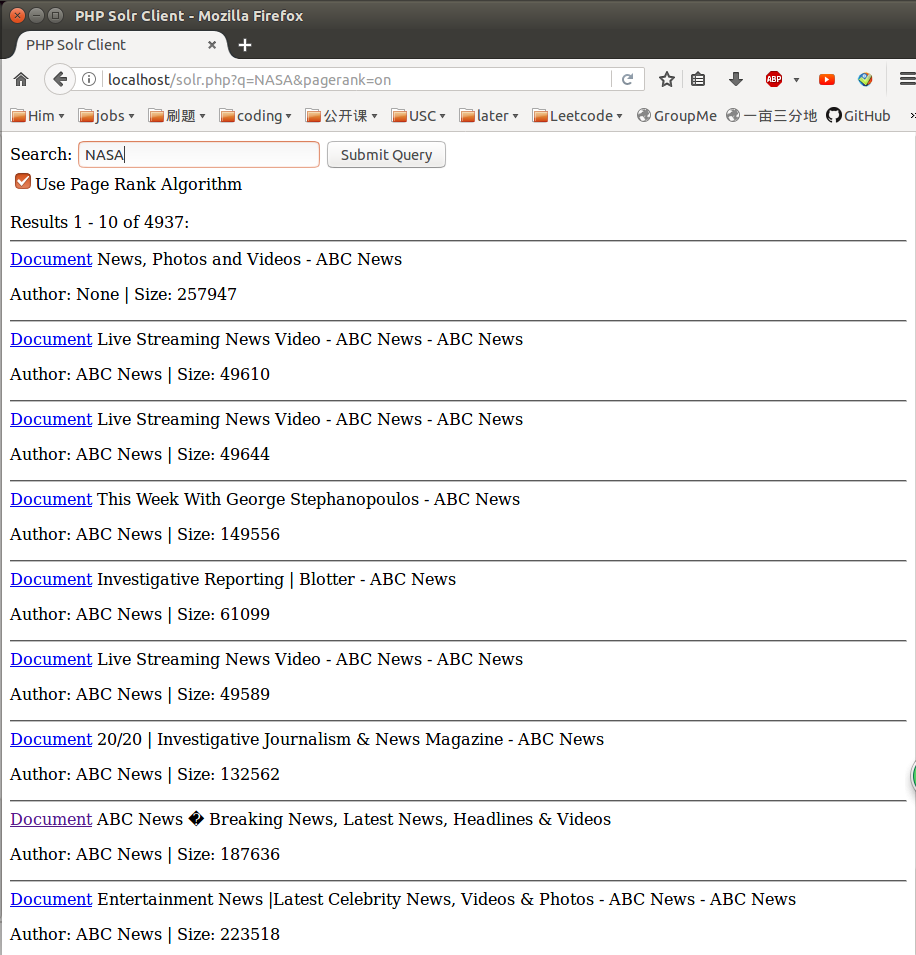
I am taking NASA as an example here. Default UI before hitting the submit button.



The default Solr searching algorithm result.

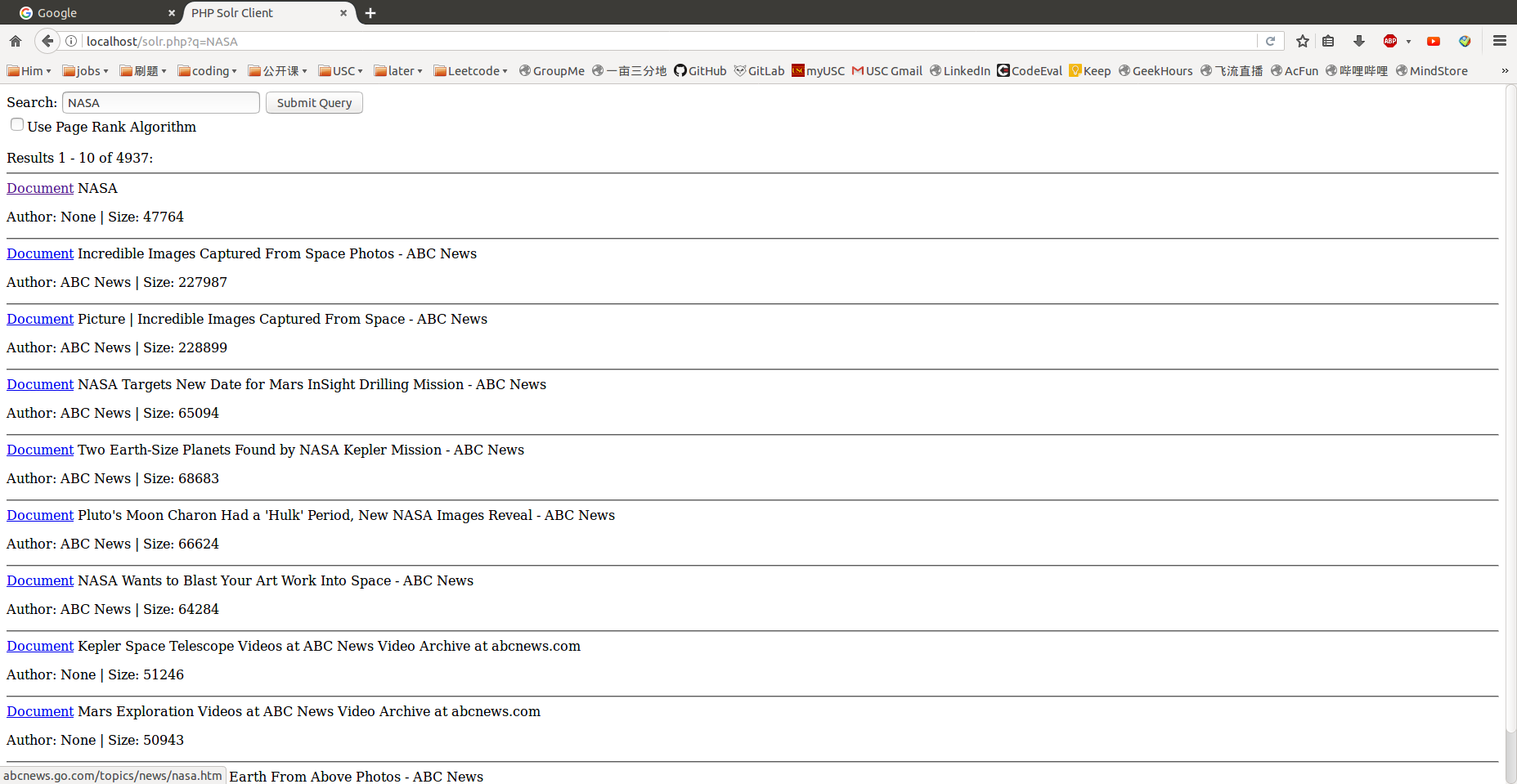


Then the PageRank searching algorithm, notice that the order changed.

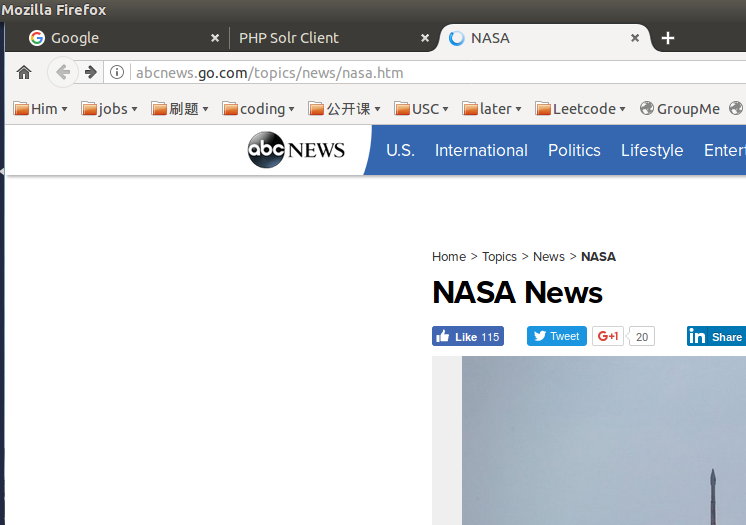


Then try to click on the first results from default searching algorithm. You can see the hyperlink shows on the left bottom corner of the screen.

<http://abcnews.go.com/topics/news/nasa.htm>



Click on this link, it takes me to this page below.



1. **Graph**

Here is a bar graph shows how many overlaps do we have between Lucene and PageRank algorithm. Only Brexit and Snapchat have overlaps between two ranking algorithms.

1. **Explanation about why some pages have higher page rank**

The logic of PageRank algorithm is ranking different webpages based on the number of outgoing links (outgoing edges) they have. For any pages except those isolated pages (with no outgoing links), if you have more numbers of outgoing links, you will have a relatively higher rank; but if you have more incoming links, it will lower your ranking. This is why some pages have higher rank and some pages have lower rank.