

# CSCI 572 - Information Retrieval and Web Search Engines

## Assignment 4 - REPORT

Vishnupriya Ravibalan – USC ID: 5597288287

### STEPS

1. Downloaded the HTML file folder (renamed crawl\_data) and the csv files from the google drive
2. Installed Solr and started solr on terminal
3. Created a core called myexample and using bin/post command I indexed the crawl\_data – 15727 indexed files for NBC News
4. Checked the indexed results on localhost:8983 by querying results.
5. I developed a simple UI using HTML and PHP to query the index. The frontend components included a search bar, submit button and two radio buttons labeled Lucene and PageRank respectively to give the user an option to pick the algorithm (Lucene or PageRank) to use while displaying the search results.
6. In order to implement PageRank, I installed JSoup in Eclipse (by adding external jar file) and ran Java code to extract all the outgoing links from every webpage and created an edge list.
7. Using the edgelist.txt previously created as an input, I wrote Python code in Eclipse using the networkx package to obtain the PageRank scores. The parameters I used were as specified in the instructions:-

*alpha=0.85, personalization=None, max\_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None*

8. I stored the output of the Python code in a file called external\_PageRankFile.txt and placed the file in the data folder of the core to make sure that this file is recognized by Solr. I also used the response ID from solr to make sure my pageRank results were recognized by the solr index and were of the format :  
*/Users/Vishnupriya/Sites/crawl\_data/76399c3ff6499aa2b60fc386f5162159.html=1.22732964297e-05*
9. I modified the managed-schema by adding an external field to bind the external\_PageRankFile to the Solr search function.
10. I modified the listener event in the solrconfig.xml file to “ExternalFieldFileReloader” to make sure Solr reads the PageRank file.
11. If the PageRank radio button was selected, I used \$additionalParameters to include an additional ‘sort’ parameter in Solr to sort the PageRank values in descending order.
12. While displaying results, each result consists of a clickable Title and URL, which redirect the user to the actual webpage which opens in a new tab. Each result also includes a description and document ID to mimic a google like user interface.
13. If the URL is missing, I traversed the given CSV file to find the URL and display the same according to FAQ 20. The Title and URL for this case will be a URL retrieved from the CSV.
14. If there is no title or description present in the file for a result, then “No Title Available” and “No Description available” are displayed.

### **Reason for higher PageRank scores of some pages:**

PageRank Algorithm is a popular search algorithm whose underlying principle is the importance of a page is determined by the number of incoming links to the webpage. The underlying assumption is that an important page will be pointed to by a good quantity of quality links, which are outgoing from pages that also have high page rank scores. Good pages lead to good pages.

The reason why some pages even though irrelevant have high Page Rank scores is because these pages have a lot of incoming links. Homepages and other pages in the upper hierarchy are common examples of pages with high PageRanks since they have a high number of incoming links.

### **Process Flow for a Query:**

Below are the screenshots showing the process flow for Query 1 – Bitcoin Blockchain

- Initial Page – Shows the search bar, two radio buttons and a submit button.



- Lucene results upon submitting the query.

PHP Solr Client for Lucene and x

localhost/?q=Bitcoin+Blockchain&Select=Lucene

Bitcoin Blockchain

Lucene Page Rank

Results 1 - 10 of 617:

- How Bitcoin and other cryptocurrencies could kill off cash**  
National governments are also looking into creating virtual currencies of their own.  
URL: <https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-nm857036> ID: /Users/Vishnupriya/Sites/crawl\_data/27640eb3f3bb03459f2330f02124860.html
- Bitcoin pays for entire wedding as its price soars east \$16 000**  
With bitcoin hitting a record high of over \$16,000, one wedding in Indiana just got bigger.  
URL: <https://www.nbcnews.com/tech/innovation/bitcoin-cant-entire-wedding-its-price-soars-east-16-000-n826776> ID: /Users/Vishnupriya/Sites/crawl\_data/c346a4b2d9558ea93d980c10cc15b5db.html
- What is Bitcoin? And should you invest in it?**  
Only 24 percent of Americans are familiar with Bitcoin. Here's what you need to know about this "cryptocurrency".  
URL: <https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-nrcal44108> ID: /Users/Vishnupriya/Sites/crawl\_data/9ba27c9494397c038fe869cb1e7ba27.html
- Middle America is Crazy in Love With Bitcoin**  
In May, there were more searches for Bitcoin than Beyonce. Cryptocoins are going mainstream as the middleclass wants in on the "Wild West" action.  
URL: <https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011> ID: /Users/Vishnupriya/Sites/crawl\_data/b4f26b0cd98731708d5e1972cd268f7.html
- How Bitcoin and other cryptocurrencies could kill off cash**  
National governments are also looking into creating virtual currencies of their own.  
URL: <https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-nm857036> ID: /Users/Vishnupriya/Sites/crawl\_data/bd4b4542a95df351b03e78a7296f9dc3.html
- The hidden dangers of buying virtual currency go beyond a simple hack**  
Cyberthieves can break into crypto exchanges, drain crypto wallets, and infect computers with malware.  
URL: <https://www.nbcnews.com/business/business-news/hidden-dangers-buying-virtual-currency-go-beyond-simple-hack-n852708> ID: /Users/Vishnupriya/Sites/crawl\_data/58f9c26a85da5d7984e10878193b681.html
- Bitcoin pays for entire wedding as its price soars east \$16 000**  
With bitcoin hitting a record high of over \$16,000, one wedding in Indiana just got bigger.  
URL: <https://www.nbcnews.com/tech/innovation/bitcoin-cant-entire-wedding-its-price-soars-east-16-000-n826776> ID: /Users/Vishnupriya/Sites/crawl\_data/80da735982094cc96cb46de189b7fb3.html
- Middle America is Crazy in Love With Bitcoin**  
In May, there were more searches for Bitcoin than Beyonce. Cryptocoins are going mainstream as the middleclass wants in on the "Wild West" action.  
URL: <https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011> ID: /Users/Vishnupriya/Sites/crawl\_data/6593288f09983334451215de9c27e60.html
- What is Bitcoin? And should you invest in it?**  
Only 24 percent of Americans are familiar with Bitcoin. Here's what you need to know about this "cryptocurrency".  
URL: <https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-nrcal44108> ID: /Users/Vishnupriya/Sites/crawl\_data/ea75aac08f02b9c0df0f1721c149.h9.html
- The hidden dangers of buying virtual currency go beyond a simple hack**  
Cyberthieves can break into crypto exchanges, drain crypto wallets, and infect computers with malware.  
URL: <https://www.nbcnews.com/business/business-news/hidden-dangers-buying-virtual-currency-go-beyond-simple-hack-n852708> ID: /Users/Vishnupriya/Sites/crawl\_data/a614919dae023dbcf160a7180a301a.html

- PageRank results upon submitting the query.

PHP Solr Client for Lucene and x

localhost/?q=Bitcoin+Blockchain&Select=PageRank

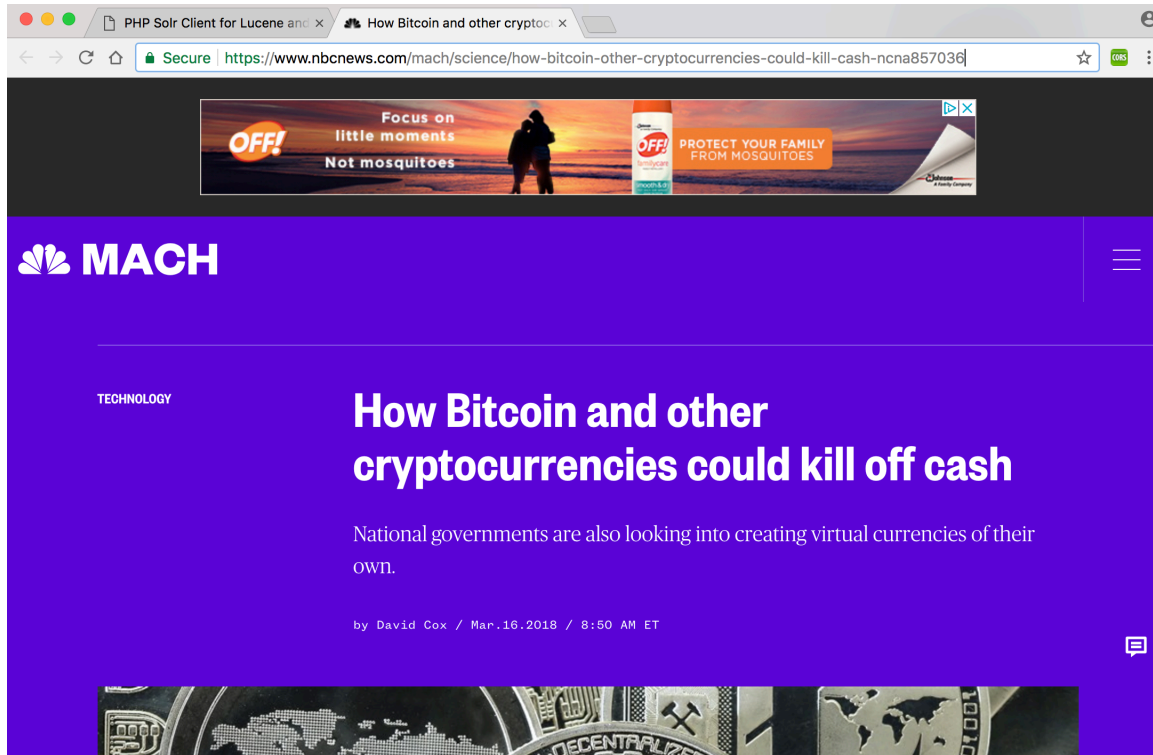
Bitcoin Blockchain

Lucene Page Rank

Results 1 - 10 of 617:

- NBC News - Breaking News & Top Stories - Latest World, US & Local News**  
Go to NBCNews.com for breaking news, videos, and the latest top stories in world news, business, politics, health and pop culture.  
URL: <https://www.nbcnews.com/> ID: /Users/Vishnupriya/Sites/crawl\_data/77b86639d1384acd8141803ecd5d27cd0.html
- MACH: The best and latest in Tech & Innovation | NBC News**  
MACH asks big questions and celebrates big ideas. It champions science & innovation, challenges technology, and explores our universe.  
URL: <https://www.nbcnews.com/mach> ID: /Users/Vishnupriya/Sites/crawl\_data/582a22b26788d93835aeb8770aa4a973.html
- Business and Finance News, Reports & Videos - NBC News**  
Find the latest news, videos, and photos on finance, industry trends, money, and more on NBCNews.com. Read business reports and watch industry-specific videos online.  
URL: <https://www.nbcnews.com/business> ID: /Users/Vishnupriya/Sites/crawl\_data/c74d9a1aa0c112a26381d5a6986a355.html
- Tech: Latest US & Global News, Innovations, Gadgets & More - NBC News**  
Find the latest tech news stories, photos, and videos on NBCNews.com. Read headlines covering cyber attacks, top new tech products, tech tips & more.  
URL: <https://www.nbcnews.com/tech> ID: /Users/Vishnupriya/Sites/crawl\_data/76d165f6b6366d3f0f8ce66437ba.html
- Science, Space, Environment & Trends - NBC News**  
Find the latest science news articles, photos and videos covering space, the environment, human development and more on NBCNews.com.  
URL: <https://www.nbcnews.com/science> ID: /Users/Vishnupriya/Sites/crawl\_data/1713bf741b955c2752ab0fae655a1a7.html
- This futuristic jet pack that lets users fly up to 10,000 feet in the sky**  
Zapata is the company best known for its hydro-flyer. It's now designed the Flyboard Air which is a jet-powered personal aerial vehicle that can go up to 10,000 feet in the sky.  
URL: <https://www.nbcnews.com/mach/video/this-futuristic-jet-pack-lets-users-fly-up-to-10-000-feet-in-the-sky-1189436995946> ID: /Users/Vishnupriya/Sites/crawl\_data/36a22c0f582c86dc79405426bda0f6.html
- Why YouTube's algorithm boosted conspiracy theory videos in the wake of the Parkland shooting**  
YouTube's recommendation algorithm has come under scrutiny before, but the recent school shooting in Parkland, Florida, launched a fresh spate of criticism. Following the shooting, the No. 1 trending video on the platform was a clip claiming one of the survivors was an actor. YouTube eventually removed the video, saying it had been misclassified by the company's systems, and that YouTube would make efforts to fix the larger issues. But the controversy highlights the larger issue of information and truth, and not only a polarizing political climate but also the era of the social media echo-chamber.  
URL: <https://www.nbcnews.com/mach/video/why-youtube-s-algorithm-boosted-conspiracy-theory-videos-in-the-wake-of-the-parkland-shooting-1187660356671> ID: /Users/Vishnupriya/Sites/crawl\_data/88992172578eb6d14b9a50eb345091.html
- From space dream to space junk: The Vanguard 1 satellite completes 60 years in space today**  
Just months after the Sputnik launch, the U.S government began testing the Vanguard project. Exactly 60 years ago, on March 17th 1958, Vanguard 1 was sent into space from Cape Canaveral, Florida, becoming the second satellite launched by the U.S and the first ever to use solar cell power. After it lost communication in 1964, it is now the oldest piece of man-made space debris.  
URL: <https://www.nbcnews.com/mach/video/from-space-dream-to-space-junk-the-vanguard-1-satellite-completes-60-years-in-space-today-1188094531644> ID: /Users/Vishnupriya/Sites/crawl\_data/b0e8908022cda26f7a5df8c281e8e49e.html
- Pi in the Sky: How a NASA scientist uses Pi to look for life beyond Earth**  
Pi is a mathematical constant that is embedded in most calculations. Dr. Luisa Rebul, who is a research scientist at Caltech, is studying Kepler data to learn more about young stars and if there are planets around them. She explains how Pi is crucial to her calculations.  
URL: <https://www.nbcnews.com/mach/video/pi-in-the-sky-how-a-nasa-scientist-uses-pi-to-look-for-life-beyond-earth-1185847363999> ID: /Users/Vishnupriya/Sites/crawl\_data/7ea0f0f6260f8a910e455d90b3c0f.html
- Astronaut Scott Kelly's genes show long-term changes after a year in space**  
NASA recently released preliminary findings from its Twin Study involving astronauts Scott Kelly and Mark Kelly. Researchers now know that 93% of Scott's genes returned to normal after spending almost a year in space. However, the remaining 7% point to possible longer term changes in his genes.  
URL: <https://www.nbcnews.com/mach/video/astonaut-scott-kelly-s-genes-show-long-term-changes-after-a-year-in-space-1188626819784> ID: /Users/Vishnupriya/Sites/crawl\_data/b01bed47761728964946a72b5287d85.html

- Webpage for Lucene query number 3 (Link opened in new tab).



## URLs of top ten results to compare Lucene and PageRank:

### Query 1 – Bitcoin Blockchain

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-ncna857036">https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-ncna857036</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/tech/innovation/bitcoin-pays-entire-wedding-its-price-soars-past-16-000-n826776">https://www.nbcnews.com/tech/innovation/bitcoin-pays-entire-wedding-its-price-soars-past-16-000-n826776</a>	<a href="https://www.nbcnews.com/mach">https://www.nbcnews.com/mach</a>
3.	<a href="https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106">https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106</a>	<a href="https://www.nbcnews.com/business">https://www.nbcnews.com/business</a>
4.	<a href="https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011">https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011</a>	<a href="https://www.nbcnews.com/tech">https://www.nbcnews.com/tech</a>
5.	<a href="https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-ncna857036">https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-ncna857036</a>	<a href="https://www.nbcnews.com/science">https://www.nbcnews.com/science</a>
6.	<a href="https://www.nbcnews.com/business/business-">https://www.nbcnews.com/business/business-</a>	<a href="https://www.nbcnews.com/mach/video/this-">https://www.nbcnews.com/mach/video/this-</a>

	news/hidden-dangers-buying-virtual-currency-go-beyond-simple-hack-n852706	futuristic-jet-pack-that-lets-users-fly-up-to-10-000-feet-in-the-sky-1189436995946
7.	<a href="https://www.nbcnews.com/tech/innovation/bitcoin-pays-entire-wedding-its-price-soars-past-16-000-n826776">https://www.nbcnews.com/tech/innovation/bitcoin-pays-entire-wedding-its-price-soars-past-16-000-n826776</a>	<a href="https://www.nbcnews.com/mach/video/why-youtube-s-algorithm-boosted-conspiracy-theory-videos-in-the-wake-of-the-parkland-shooting-1187660355671">https://www.nbcnews.com/mach/video/why-youtube-s-algorithm-boosted-conspiracy-theory-videos-in-the-wake-of-the-parkland-shooting-1187660355671</a>
8.	<a href="https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011">https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011</a>	<a href="https://www.nbcnews.com/mach/video/from-space-dream-to-space-junk-the-vanguard-1-satellite-completes-60-years-in-space-today-1188094531644">https://www.nbcnews.com/mach/video/from-space-dream-to-space-junk-the-vanguard-1-satellite-completes-60-years-in-space-today-1188094531644</a>
9.	<a href="https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106">https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106</a>	<a href="https://www.nbcnews.com/mach/video/pi-in-the-sky-how-a-nasa-scientist-uses-pi-to-look-for-life-beyond-earth-1185847363999">https://www.nbcnews.com/mach/video/pi-in-the-sky-how-a-nasa-scientist-uses-pi-to-look-for-life-beyond-earth-1185847363999</a>
10.	<a href="https://www.nbcnews.com/business/business-news/hidden-dangers-buying-virtual-currency-go-beyond-simple-hack-n852706">https://www.nbcnews.com/business/business-news/hidden-dangers-buying-virtual-currency-go-beyond-simple-hack-n852706</a>	<a href="https://www.nbcnews.com/mach/video/astronaut-scott-kelly-s-genes-show-long-term-changes-after-a-year-in-space-1186826819794">https://www.nbcnews.com/mach/video/astronaut-scott-kelly-s-genes-show-long-term-changes-after-a-year-in-space-1186826819794</a>

## Query 2 – Stephen Hawking

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/mach/science/brief-history-stephen-hawking-cambridge-cultural-icon-ncsl856456">https://www.nbcnews.com/mach/science/brief-history-stephen-hawking-cambridge-cultural-icon-ncsl856456</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/news/world/british-physicist-stephen-hawking-among-world-s-greatest-minds-science-n702301">https://www.nbcnews.com/news/world/british-physicist-stephen-hawking-among-world-s-greatest-minds-science-n702301</a>	<a href="https://www.nbcnews.com/meet-the-press">https://www.nbcnews.com/meet-the-press</a>
3.	<a href="https://www.nbcnews.com/think/opinion/stephen-hawking-s-life-was-made-possible-access-great-health-ncna856971">https://www.nbcnews.com/think/opinion/stephen-hawking-s-life-was-made-possible-access-great-health-ncna856971</a>	<a href="https://www.nbcnews.com/mach">https://www.nbcnews.com/mach</a>
4.	<a href="https://www.nbcnews.com/science/space/i-m-atheist-stephen-hawking-god-space-travel-n210076">https://www.nbcnews.com/science/space/i-m-atheist-stephen-hawking-god-space-travel-n210076</a>	<a href="https://www.nbcnews.com/think">https://www.nbcnews.com/think</a>
5.	<a href="https://www.nbcnews.com/science/space/i-m-atheist-stephen-hawking-god-space-travel-n210076">https://www.nbcnews.com/science/space/i-m-atheist-stephen-hawking-god-space-travel-n210076</a>	<a href="https://www.nbcnews.com/health">https://www.nbcnews.com/health</a>
6.	<a href="https://www.nbcnews.com/news/world/british-physicist-stephen-hawking-among-world-s-greatest-minds-science-n702301">https://www.nbcnews.com/news/world/british-physicist-stephen-hawking-among-world-s-greatest-minds-science-n702301</a>	<a href="https://www.nbcnews.com/news/us-news">https://www.nbcnews.com/news/us-news</a>
7.	<a href="https://www.nbcnews.com/mach/science/stephen-hawking-says-humans-must-leave-earth-within-600-years-ncna818926">https://www.nbcnews.com/mach/science/stephen-hawking-says-humans-must-leave-earth-within-600-years-ncna818926</a>	<a href="https://www.nbcnews.com/science">https://www.nbcnews.com/science</a>
8.	<a href="https://www.nbcnews.com/science/science-news/six-years-after-zero-g-flight-stephen-hawking-still-space-f6C9703084">https://www.nbcnews.com/science/science-news/six-years-after-zero-g-flight-stephen-hawking-still-space-f6C9703084</a>	<a href="https://www.nbcnews.com/news/photo">https://www.nbcnews.com/news/photo</a>
9.	<a href="https://www.nbcnews.com/mach/science/what-stephen-hawking-taught-us-about-">https://www.nbcnews.com/mach/science/what-stephen-hawking-taught-us-about-</a>	<a href="https://www.nbcnews.com/video">https://www.nbcnews.com/video</a>

	black-holes-ncna857471	
10.	<a href="https://www.nbcnews.com/mach/science/stephen-hawking-says-humans-must-leave-earth-within-600-years-ncna818926">https://www.nbcnews.com/mach/science/stephen-hawking-says-humans-must-leave-earth-within-600-years-ncna818926</a>	<a href="https://www.nbcnews.com/news">https://www.nbcnews.com/news</a>

### **Query 3 – Falcon Heavy**

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/mach/science/spacex-s-falcon-heavy-launch-gives-big-boost-elon-musk-ncna845256">https://www.nbcnews.com/mach/science/spacex-s-falcon-heavy-launch-gives-big-boost-elon-musk-ncna845256</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/mach/science/watch-spacex-s-falcon-heavy-rocket-make-its-maiden-flight-ncna845096">https://www.nbcnews.com/mach/science/watch-spacex-s-falcon-heavy-rocket-make-its-maiden-flight-ncna845096</a>	<a href="https://www.nbcnews.com/meet-the-press">https://www.nbcnews.com/meet-the-press</a>
3.	<a href="https://www.nbcnews.com/mach/science/watch-spacex-s-falcon-heavy-rocket-make-its-maiden-flight-ncna845096">https://www.nbcnews.com/mach/science/watch-spacex-s-falcon-heavy-rocket-make-its-maiden-flight-ncna845096</a>	<a href="https://www.nbcnews.com/nightly-news">https://www.nbcnews.com/nightly-news</a>
4.	<a href="https://www.nbcnews.com/mach/science/spacex-s-falcon-heavy-launch-gives-big-boost-elon-musk-ncna845256">https://www.nbcnews.com/mach/science/spacex-s-falcon-heavy-launch-gives-big-boost-elon-musk-ncna845256</a>	<a href="https://www.nbcnews.com/news/world">https://www.nbcnews.com/news/world</a>
5.	<a href="https://www.nbcnews.com/mach/science/spacex-s-falcon-feat-may-be-just-prelude-main-event-ncna846446">https://www.nbcnews.com/mach/science/spacex-s-falcon-feat-may-be-just-prelude-main-event-ncna846446</a>	<a href="https://www.nbcnews.com/news/latino">https://www.nbcnews.com/news/latino</a>
6.	<a href="https://www.nbcnews.com/mach/science/heres-what-will-happen-tesla-spacex-shot-space-ncna845566">https://www.nbcnews.com/mach/science/heres-what-will-happen-tesla-spacex-shot-space-ncna845566</a>	<a href="https://www.nbcnews.com/science">https://www.nbcnews.com/science</a>
7.	<a href="https://www.nbcnews.com/mach/science/month-space-pictures-cosmic-penguin-playing-mars-ncsl852081">https://www.nbcnews.com/mach/science/month-space-pictures-cosmic-penguin-playing-mars-ncsl852081</a>	<a href="https://www.nbcnews.com/news/weather">https://www.nbcnews.com/news/weather</a>
8.	<a href="https://www.nbcnews.com/mach/science/spacex-s-falcon-feat-may-be-just-prelude-main-event-ncna846446">https://www.nbcnews.com/mach/science/spacex-s-falcon-feat-may-be-just-prelude-main-event-ncna846446</a>	<a href="https://www.nbcnews.com/news/photo">https://www.nbcnews.com/news/photo</a>
9.	<a href="https://www.nbcnews.com/science/space">https://www.nbcnews.com/science/space</a>	<a href="https://www.nbcnews.com/video">https://www.nbcnews.com/video</a>
10.	<a href="https://www.nbcnews.com/mach/science/how-spacex-aims-bring-high-speed-internet-whole-world-ncna853766">https://www.nbcnews.com/mach/science/how-spacex-aims-bring-high-speed-internet-whole-world-ncna853766</a>	<a href="https://www.nbcnews.com/news">https://www.nbcnews.com/news</a>

### **Query 4 – Syrian War**

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/storyline/syrias-children">https://www.nbcnews.com/storyline/syrias-children</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/storyline/syrias-children/syrian-civil-war-photo-photographer-s-rescue-goes-viral-n748146">https://www.nbcnews.com/storyline/syrias-children/syrian-civil-war-photo-photographer-s-rescue-goes-viral-n748146</a>	<a href="https://www.nbcnews.com/meet-the-press">https://www.nbcnews.com/meet-the-press</a>
3.	<a href="https://www.nbcnews.com/storyline/aleppos">https://www.nbcnews.com/storyline/aleppos</a>	<a href="https://www.nbcnews.com/nightly-news">https://www.nbcnews.com/nightly-news</a>

	children	
4.	<a href="https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851">https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851</a>	<a href="https://www.nbcnews.com/mach">https://www.nbcnews.com/mach</a>
5.	<a href="https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851">https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851</a>	<a href="https://www.nbcnews.com/politics">https://www.nbcnews.com/politics</a>
6.	<a href="https://www.nbcnews.com/news/other/barrel-bombs-just-another-symbol-brutality-syrian-civil-war-f2D11765640">https://www.nbcnews.com/news/other/barrel-bombs-just-another-symbol-brutality-syrian-civil-war-f2D11765640</a>	<a href="https://www.nbcnews.com/think">https://www.nbcnews.com/think</a>
7.	<a href="https://www.nbcnews.com/slideshow/children-among-injured-besieged-damascus-suburb-n834526">https://www.nbcnews.com/slideshow/children-among-injured-besieged-damascus-suburb-n834526</a>	<a href="https://www.nbcnews.com/feature/nbc-out">https://www.nbcnews.com/feature/nbc-out</a>
8.	<a href="https://www.nbcnews.com/slideshow/children-among-injured-besieged-damascus-suburb-n834526">https://www.nbcnews.com/slideshow/children-among-injured-besieged-damascus-suburb-n834526</a>	<a href="https://www.nbcnews.com/business">https://www.nbcnews.com/business</a>
9.	<a href="https://www.nbcnews.com/storyline/syrias-suffering-families">https://www.nbcnews.com/storyline/syrias-suffering-families</a>	<a href="https://www.nbcnews.com/news/world">https://www.nbcnews.com/news/world</a>
10.	<a href="https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851">https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851</a>	<a href="https://www.nbcnews.com/news/latino">https://www.nbcnews.com/news/latino</a>

#### **Query 5 – LA Rams**

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/news/sports/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366">https://www.nbcnews.com/news/sports/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/news/sports/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366">https://www.nbcnews.com/news/sports/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366</a>	<a href="https://www.nbcnews.com/meet-the-press">https://www.nbcnews.com/meet-the-press</a>
3.	<a href="https://www.nbcnews.com/news/sports/sports-broadcasting-legend-dick-enberg-dies-82-n832076">https://www.nbcnews.com/news/sports/sports-broadcasting-legend-dick-enberg-dies-82-n832076</a>	<a href="https://www.nbcnews.com/news/nbcblk">https://www.nbcnews.com/news/nbcblk</a>
4.	<a href="https://www.nbcnews.com/slideshow/black-quarterbacks-who-ve-played-super-bowl-n513226">https://www.nbcnews.com/slideshow/black-quarterbacks-who-ve-played-super-bowl-n513226</a>	<a href="https://www.nbcnews.com/news/latino">https://www.nbcnews.com/news/latino</a>
5.	<a href="https://www.nbcnews.com/news/sports/sports-broadcasting-legend-dick-enberg-dies-82-n832076">https://www.nbcnews.com/news/sports/sports-broadcasting-legend-dick-enberg-dies-82-n832076</a>	<a href="https://www.nbcnews.com/tech">https://www.nbcnews.com/tech</a>
6.	<a href="https://www.nbcnews.com/pop-culture/music/bernie-casey-pro-football-player-turned-fan-favorite-actor-dies-n803291">https://www.nbcnews.com/pop-culture/music/bernie-casey-pro-football-player-turned-fan-favorite-actor-dies-n803291</a>	<a href="https://www.nbcnews.com/news/us-news/nbc-affiliates-n19981">https://www.nbcnews.com/news/us-news/nbc-affiliates-n19981</a>
7.	<a href="https://www.nbcnews.com/pop-culture/music/bernie-casey-pro-football-player-turned-fan-favorite-actor-dies-n803291">https://www.nbcnews.com/pop-culture/music/bernie-casey-pro-football-player-turned-fan-favorite-actor-dies-n803291</a>	<a href="https://www.nbcnews.com/news">https://www.nbcnews.com/news</a>

8.	<a href="https://www.nbcnews.com/meet-the-press/meet-press-transcript-october-11-2015-n442476">https://www.nbcnews.com/meet-the-press/meet-press-transcript-october-11-2015-n442476</a>	<a href="https://www.nbcnews.com/better/video/how-mika-brzezinski-found-her-value-and-got-paid-what-she-s-worth-1189457987619">https://www.nbcnews.com/better/video/how-mika-brzezinski-found-her-value-and-got-paid-what-she-s-worth-1189457987619</a>
9.	<a href="https://www.nbcnews.com/meet-the-press/meet-press-transcript-october-11-2015-n442476">https://www.nbcnews.com/meet-the-press/meet-press-transcript-october-11-2015-n442476</a>	<a href="https://www.nbcnews.com/better/video/how-not-to-choke-in-sports-and-other-pressure-charged-situations-1186761795616">https://www.nbcnews.com/better/video/how-not-to-choke-in-sports-and-other-pressure-charged-situations-1186761795616</a>
10.	<a href="https://www.nbcnews.com/politics/2016-election/la">https://www.nbcnews.com/politics/2016-election/la</a>	<a href="https://www.nbcnews.com/better/video/how-to-lose-weight-and-why-it-s-so-hard-1187543619775">https://www.nbcnews.com/better/video/how-to-lose-weight-and-why-it-s-so-hard-1187543619775</a>

### **Query 6 – Black Panther**

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/news/nbcblk/black-panther-s-wakanda-sheds-light-black-excellence-n844801">https://www.nbcnews.com/news/nbcblk/black-panther-s-wakanda-sheds-light-black-excellence-n844801</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/news/nbcblk/black-panther-sets-fandango-record-top-pre-selling-marvel-movie-n836631">https://www.nbcnews.com/news/nbcblk/black-panther-sets-fandango-record-top-pre-selling-marvel-movie-n836631</a>	<a href="https://www.nbcnews.com/nightly-news">https://www.nbcnews.com/nightly-news</a>
3.	<a href="https://www.nbcnews.com/news/nbcblk/black-panther-novelist-journalism-sci-fi-bring-hidden-light-n848011">https://www.nbcnews.com/news/nbcblk/black-panther-novelist-journalism-sci-fi-bring-hidden-light-n848011</a>	<a href="https://www.nbcnews.com/mach">https://www.nbcnews.com/mach</a>
4.	<a href="https://www.nbcnews.com/tech/internet/plot-destroy-black-panther-ratings-thwarted-facebook-n844226">https://www.nbcnews.com/tech/internet/plot-destroy-black-panther-ratings-thwarted-facebook-n844226</a>	<a href="https://www.nbcnews.com/think">https://www.nbcnews.com/think</a>
5.	<a href="https://www.nbcnews.com/pop-culture/movies/black-panther-blows-away-box-office-192-million-weekend-n849246">https://www.nbcnews.com/pop-culture/movies/black-panther-blows-away-box-office-192-million-weekend-n849246</a>	<a href="https://www.nbcnews.com/health">https://www.nbcnews.com/health</a>
6.	<a href="https://www.nbcnews.com/news/nbcblk/black-panther-black-comic-creators-spent-years-breaking-down-barriers-n848461">https://www.nbcnews.com/news/nbcblk/black-panther-black-comic-creators-spent-years-breaking-down-barriers-n848461</a>	<a href="https://www.nbcnews.com/feature/nbc-out">https://www.nbcnews.com/feature/nbc-out</a>
7.	<a href="https://www.nbcnews.com/news/nbcblk/black-panther-actress-janeshia-adams-ginyard-talks-living-stuntwoman-life-n847646">https://www.nbcnews.com/news/nbcblk/black-panther-actress-janeshia-adams-ginyard-talks-living-stuntwoman-life-n847646</a>	<a href="https://www.nbcnews.com/pop-culture">https://www.nbcnews.com/pop-culture</a>
8.	<a href="https://www.nbcnews.com/think/opinion/defeat-killmonger-black-panther-holds-lessons-israel-ncna850831">https://www.nbcnews.com/think/opinion/defeat-killmonger-black-panther-holds-lessons-israel-ncna850831</a>	<a href="https://www.nbcnews.com/news/nbcblk">https://www.nbcnews.com/news/nbcblk</a>
9.	<a href="https://www.nbcnews.com/news/nbcblk/black-panther-novelist-journalism-sci-fi-bring-hidden-light-n848011">https://www.nbcnews.com/news/nbcblk/black-panther-novelist-journalism-sci-fi-bring-hidden-light-n848011</a>	<a href="https://www.nbcnews.com/news/asian-america">https://www.nbcnews.com/news/asian-america</a>
10.	<a href="https://www.nbcnews.com/news/nbcblk/stars-black-panther-discuss-hope-marvel-film-impact-n847706">https://www.nbcnews.com/news/nbcblk/stars-black-panther-discuss-hope-marvel-film-impact-n847706</a>	<a href="https://www.nbcnews.com/tech">https://www.nbcnews.com/tech</a>



### **Query 7 – Austin Bombings**

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/news/us-news/police-identify-teen-killed-austin-package-bombing-investigation-grows-n856371">https://www.nbcnews.com/news/us-news/police-identify-teen-killed-austin-package-bombing-investigation-grows-n856371</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/news/us-news/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786">https://www.nbcnews.com/news/us-news/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786</a>	<a href="https://www.nbcnews.com/meet-the-press">https://www.nbcnews.com/meet-the-press</a>
3.	<a href="https://www.nbcnews.com/news/us-news/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786">https://www.nbcnews.com/news/us-news/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786</a>	<a href="https://www.nbcnews.com/nightly-news">https://www.nbcnews.com/nightly-news</a>
4.	<a href="https://www.nbcnews.com/widget/video-embed/1184733251954">https://www.nbcnews.com/widget/video-embed/1184733251954</a>	<a href="https://www.nbcnews.com/dateline">https://www.nbcnews.com/dateline</a>
5.	<a href="https://www.nbcnews.com/news/us-news/police-identify-teen-killed-austin-package-bombing-investigation-grows-n856371">https://www.nbcnews.com/news/us-news/police-identify-teen-killed-austin-package-bombing-investigation-grows-n856371</a>	<a href="https://www.nbcnews.com/better">https://www.nbcnews.com/better</a>
6.	<a href="https://www.nbcnews.com/news/us-news/austin-texas-edge-after-spate-package-bombs-kill-two-n855956">https://www.nbcnews.com/news/us-news/austin-texas-edge-after-spate-package-bombs-kill-two-n855956</a>	<a href="https://www.nbcnews.com/mach">https://www.nbcnews.com/mach</a>
7.	<a href="https://www.nbcnews.com/news/us-news/austin-texas-edge-after-spate-package-bombs-kill-two-n855956">https://www.nbcnews.com/news/us-news/austin-texas-edge-after-spate-package-bombs-kill-two-n855956</a>	<a href="https://www.nbcnews.com/leftfield">https://www.nbcnews.com/leftfield</a>
8.	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>	<a href="https://www.nbcnews.com/think">https://www.nbcnews.com/think</a>
9.	<a href="https://www.nbcnews.com/news/latino/california-affirmative-action-bill-sparks-fierce-debate-n87426">https://www.nbcnews.com/news/latino/california-affirmative-action-bill-sparks-fierce-debate-n87426</a>	<a href="https://www.nbcnews.com/news/us-news">https://www.nbcnews.com/news/us-news</a>
10.	<a href="https://www.nbcnews.com/news/latino/celia-cruz-be-remembered-grammy-lifetime-achievement-award-n516956">https://www.nbcnews.com/news/latino/celia-cruz-be-remembered-grammy-lifetime-achievement-award-n516956</a>	<a href="https://www.nbcnews.com/news/nbcblk">https://www.nbcnews.com/news/nbcblk</a>

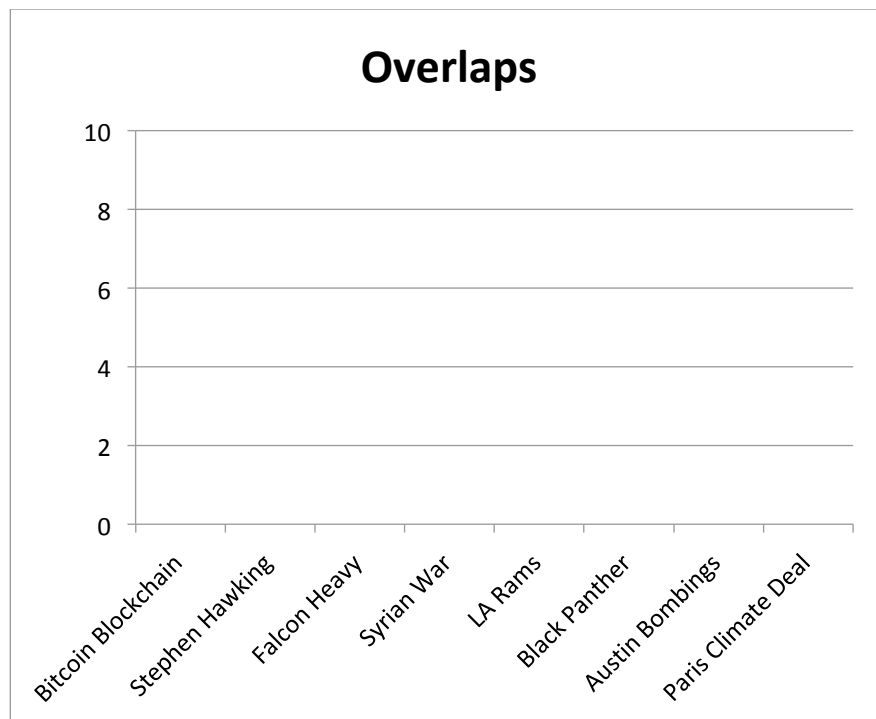
### **Query 8 – Paris Climate Deal**

	Lucene	PageRank
1.	<a href="https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-declines-endorsing-paris-climate-change-deal-g7-summit-will-n765406">https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-declines-endorsing-paris-climate-change-deal-g7-summit-will-n765406</a>	<a href="https://www.nbcnews.com/">https://www.nbcnews.com/</a>
2.	<a href="https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-declines-endorsing-paris-climate-change-deal-g7-summit-will-n765406">https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-declines-endorsing-paris-climate-change-deal-g7-summit-will-n765406</a>	<a href="https://www.nbcnews.com/meet-the-press">https://www.nbcnews.com/meet-the-press</a>
3.	<a href="https://www.nbcnews.com/politics/white-house/trump-pulls-u-s-out-paris-climate-">https://www.nbcnews.com/politics/white-house/trump-pulls-u-s-out-paris-climate-</a>	<a href="https://www.nbcnews.com/dateline">https://www.nbcnews.com/dateline</a>

	agreement-n767066	
4.	<a href="https://www.nbcnews.com/politics/white-house/here-s-what-happens-if-trump-leaves-paris-climate-agreement-n766761">https://www.nbcnews.com/politics/white-house/here-s-what-happens-if-trump-leaves-paris-climate-agreement-n766761</a>	<a href="https://www.nbcnews.com/mach">https://www.nbcnews.com/mach</a>
5.	<a href="https://www.nbcnews.com/politics/politics-news/forget-trump-u-s-storms-ahead-climate-change-never-n801826">https://www.nbcnews.com/politics/politics-news/forget-trump-u-s-storms-ahead-climate-change-never-n801826</a>	<a href="https://www.nbcnews.com/health">https://www.nbcnews.com/health</a>
6.	<a href="https://www.nbcnews.com/meet-the-press/meet-press-june-4-2017-n768046">https://www.nbcnews.com/meet-the-press/meet-press-june-4-2017-n768046</a>	<a href="https://www.nbcnews.com/investigations">https://www.nbcnews.com/investigations</a>
7.	<a href="https://www.nbcnews.com/news/world/analysis-trump-s-america-first-foreign-policy-isolates-u-s-n833046">https://www.nbcnews.com/news/world/analysis-trump-s-america-first-foreign-policy-isolates-u-s-n833046</a>	<a href="https://www.nbcnews.com/science">https://www.nbcnews.com/science</a>
8.	<a href="https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-end-first-foreign-trip-pledges-unwavering-support-u-s-n765446">https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-end-first-foreign-trip-pledges-unwavering-support-u-s-n765446</a>	<a href="https://www.nbcnews.com/news/weather">https://www.nbcnews.com/news/weather</a>
9.	<a href="https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/g-7-summit-leaders-brace-clash-trump-trade-climate-n764971">https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/g-7-summit-leaders-brace-clash-trump-trade-climate-n764971</a>	<a href="https://www.nbcnews.com/video">https://www.nbcnews.com/video</a>
10.	<a href="https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-end-first-foreign-trip-pledges-unwavering-support-u-s-n765446">https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-end-first-foreign-trip-pledges-unwavering-support-u-s-n765446</a>	<a href="https://www.nbcnews.com/storyline/bill-cosby-scandal">https://www.nbcnews.com/storyline/bill-cosby-scandal</a>

### Overlaps:

The graph below depicts the overlaps between the Lucene and PageRank results for the different queries. There were no overlaps between the two ranking results.



# CSCI 572 - Information Retrieval and Web Search Engines

## Assignment 5 - REPORT

Vishnupriya Ravibalan – USC ID: 5597288287

### STEPS

- Used my homework 4 files as a base for this assignment.
- Used the tutorial on adding autocomplete to Solr and made the required changes to the solrconfig.xml file, thus enabling solr's autocomplete feature.
- Implemented the front end using JQuery and JQuery UI plugins in the front end to implement the autocomplete.
- Every time a user types in the search box a request is sent to Solr Admin (Query used contains most importantly “/suggest?”) and the suggestions sent back from the server are displayed.
- Downloaded Peter Norvig's SpellChecker.php and used it to implement spelling correction.
- The SpellChecker.php requires a big.txt file, which is generated using a Java Program, which uses Apache Tika. I displayed the corrected query as “Did you mean:” and the correct query results.
- I used the Meta data from the html files to display the title, snippet, URL and ID for each result (top results) on the results page. I also highlighted the occurrence of every query term in the snippet.
- Whenever there is no snippet available, I used the description from the HTML metadata and whenever there is no URL available I retrieved it from the CSV file using the document ID.

### Results:-

Spell Correct:

	Input	Corrected Output
1.	juge	judge
2.	orge	urge
3.	amerca	america
4.	losang	losing
5.	worledl	world

Autocomplete:

	Query	Suggestions
1.	ju	Just,jul15,justice,jul
2.	Black pa	black parsed, black page, black pages, black path
3.	la	Label,last,latino,large
4.	ad	ad,advertise,adchoices,advertisement
5.	fo	for,format,footer,form