#### **CSCI 572 Assignment 4 Report**

#### 1. Steps followed in the assignment:

- Installed Oracle VirtualBox and Ubuntu 16.04 LTS
- Downloaded and unzipped Solr 7.3.3 in Ubuntu
- ➤ Downloaded the google drive folder with HTML files for fox news website along with the mapping csv file.
- > Started Solr server, created a new core 'myexample'. Modified config files, 'managed-scheme.xml'.
- > Tried indexing based on just 5 html files as a test.
- ➤ Indexed all the HTML files in the 'foxnews' folder.
- > Tried some basic queries on this index.
- Installed Apache and PHP on ubuntu using instructions given in FAQ no.48.
- ➤ Cloned the git repository (https://github.com/PTCInc/solr-php-client.git) for making HTTPS requests to Solr into a folder.
- ➤ Modified the php file given in the writeup to only display items that are required (Title, URL, ID, Desc)
- ➤ Placed my front-end.php file into the above git folder and started php server from this location.
- ➤ Searched queries using the PHP user interface, made sure all fields are displayed as required in the assignment write up.
- ➤ Modified the ExtractLinks.java file given in the write-up.
- Created 2 hashmaps, one from URL to csv file and vice versa.
- ➤ Extracted all links from html files from folder and matched them with csv file name equivalents.

- Created and outputted edgelist as edges.txt
- ➤ Wrote python file 'MakePageRanks.py' using networkx library to convert edgelist into dictionary containing PageRank. PageRank values were computed using the PageRank function of networkx library using the following parameters:

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

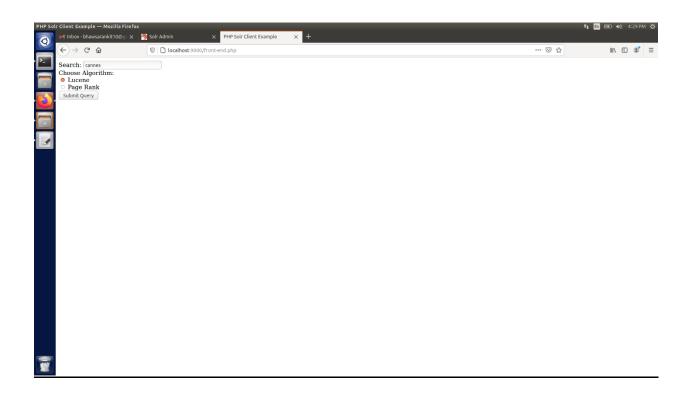
- Stored page rank value and file names in external\_pageRankFile.txt
- ➤ Place external page rank file in the solr core folder inside the data folder.
- Modified config files, 'managed-schema.xml' and 'solr-config.xml' and verified working of PageRank algorithm in Solr.
- ➤ Added radio buttons functionality in php front end.

#### 2. Why some pages have higher PageRank values than others:

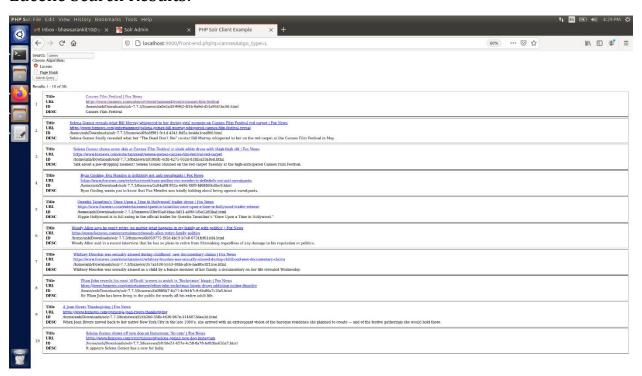
PageRank scores are a way to rank webpages. The higher the PageRank of a page, the more authoritative it is. Some pages have higher page rank than others because more webpages link to it. For example, a page having more in-links will be seen as reputable (as a lot a page's reference it), and the reputable page should have a high page rank. On the other hand, if a webpage has few inward facing links, it means its content is not verified because not a lot of people cite it. Such a page should have a low page rank score.

#### 3. Flow of single query:

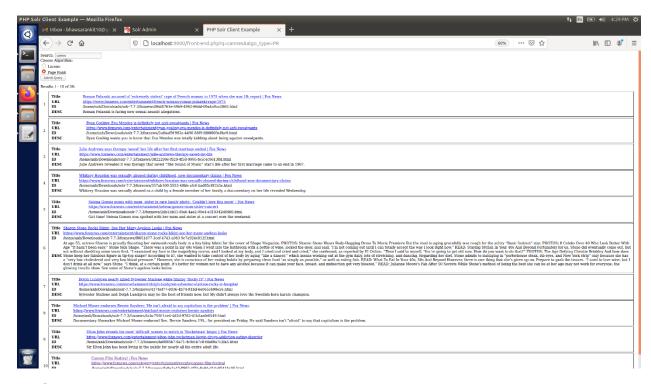
**Home Page:** 



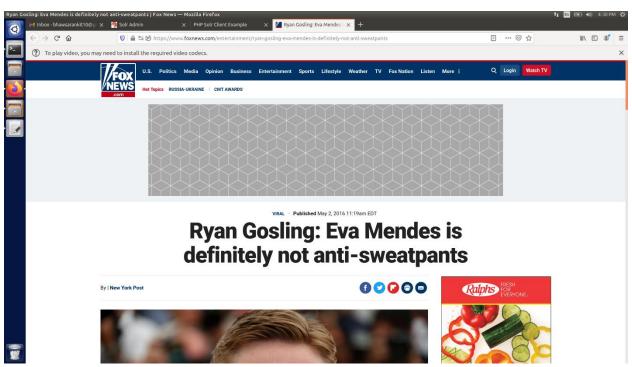
### Lucene Search Results:



#### PageRank Results:



#### Webpage:



#### 4. Table for comparing algorithms:

Excel Sheet for this table has also been provided for easy-to-read access.

Query	S r. N o.	Lucene Search	PageRank Search
Canne	1	https://www.foxnews.com/categor y/entertainment/events/cannes- film-festival	https://www.foxnews.com/enterta inment/french-woman-roman- polanski-rape-1975
	2	https://www.foxnews.com/enterta inment/selena-gomez-bill-murray- whispered-cannes-film-festival- reveal	https://www.foxnews.com/enterta inment/ryan-gosling-eva-mendes- is-definitely-not-anti-sweatpants
	3	https://www.foxnews.com/enterta inment/selena-gomez-cannes-film- festival-red-carpet	https://www.foxnews.com/enterta inment/julie-andrews-therapy- saved-my-life
	4	https://www.foxnews.com/enterta inment/ryan-gosling-eva-mendes- is-definitely-not-anti-sweatpants	https://www.foxnews.com/enterta inment/whitney-houston-was- sexually-abused-during-childhood- new-documentary-claims
	5	https://www.foxnews.com/enterta inment/quentin-tarantino-once- upon-a-time-in-hollywood-trailer- release	https://www.foxnews.com/enterta inment/selena-gomez-mom-sister- concert
	6	https://www.foxnews.com/enterta inment/woody-allen-retire-family- politics	https://www.foxnews.com/enterta inment/sharon-stone-rocks-bikini- see-her-many-ageless-looks
	7	https://www.foxnews.com/enterta inment/whitney-houston-was- sexually-abused-during-childhood- new-documentary-claims	https://www.foxnews.com/enterta inment/dolph-lundgren-sylvester- stallone-rocky-iv-hospital

	8	https://www.foxnews.com/enterta inment/elton-john-rocketman- biopic-drugs-addiction-eating- disorder	https://www.foxnews.com/enterta inment/michael-moore-endorses- bernie-sanders
	9	https://www.foxnews.com/opinion/a-joan-rivers-thanksgiving	https://www.foxnews.com/enterta inment/elton-john-rocketman- biopic-drugs-addiction-eating- disorder
		https://www.foxnews.com/enterta inment/selena-gomez-new-dog- instagram	https://www.foxnews.com/categor y/entertainment/events/cannes- film-festival
	1	https://www.foxnews.com/us/nav y-vet-donates-gold-star-families- stories-to-library-of-congress	https://www.foxnews.com/opinion
	2	https://www.foxnews.com/opinion /judge-andrew-napolitano- congress-constitutional-power- declare-war	https://www.foxnews.com/tech
Congr	3	https://www.foxnews.com/politics/jon-stewart-blasts-congress-over-9-11-fund-hearing-drove-me-nuts-there-were-empty-seats	https://www.foxnews.com/categor y/world/conflicts
	4	https://www.foxnews.com/opinion/rep-andy-biggs-congress-debt-ceiling-budget-deal	https://www.foxnews.com/categor y/politics/foreign-policy
	5	https://www.foxnews.com/opinion/gutfeld-congress-attacking-zuckerberg	https://www.foxnews.com/official- polls

	6	https://www.foxnews.com/opinion/andy-biggs-congress-judges-immigration-laws	https://www.foxnews.com/categor y/politics/executive
	7	https://www.foxnews.com/faith- values/bible-school-drew-brees- day-congress	https://www.foxnews.com/categor y/tech/topics/security
	8	https://www.foxnews.com/categor y/us/congress	https://www.foxnews.com/opinion /mike-huckabee-lets-fight-for- fabricated-structural-steel- manufacturing-jobs
	9	https://www.foxnews.com/opinion /bradley-blakeman-trade-pact- with-canada-and-mexico-offers- incredible-benefits-congress- should-approve-it	http://www.foxnews.com/trending
	1 0	https://www.foxnews.com/politics/trump-tells-congress-to-immediately-approve-mexico-canada-trade-deal-says-us-moving-past-stupid-years-of-trade-policy	https://www.foxnews.com/media/ elijah-cummings-seat-kimberly- klacik-baltimore-election- republican
Demo	1	https://www.foxnews.com/opinion/rnc-chair-mcdaniel-democrats-impeachment-vote-is-baseless-political-vendetta-trump-did-nothing-wrong	https://www.foxnews.com/opinion
crats	2	https://www.foxnews.com/politics /democrats-arent-actually- impeaching-the-president-yet-but- they-are-digging-deep	https://www.foxnews.com/us
	3	https://www.foxnews.com/opinion/david-bossie-baseless-trump-	https://www.foxnews.com

	impeachment-inquiry-by- desperate-democrats-makes- dramatic-flip-flop	
4	https://www.foxnews.com/politics/federal-judge-house-democrats-mazars-subpoena-trump-financial-record	https://www.foxnews.com/politics
5	https://www.foxnews.com/opinion/rep-andy-biggs-lipstick-impeachment	https://www.foxnews.com/categor y/politics/elections
6	https://www.foxnews.com/media/democrats-disdain-working-class-trump-jr-fox-friends	https://www.foxnews.com/categor y/world/scandals
7	https://www.foxnews.com/opinion/liz-harrington-democrats-holding-soviet-style-show-trial-to-overthrow-trump-not-an-impeachment-inquiry	https://www.foxnews.com/categor y/world/conflicts
8	https://www.foxnews.com/opinion/donald-trump-jr-democrats-immigration-radical-ideology	https://www.foxnews.com/categor y/politics/foreign-policy
9	https://www.foxnews.com/opinion/hans-von-spakovsky-in-trump-impeachment-probe-democrats-refuse-to-follow-nixon-and-clinton-precedents	https://www.foxnews.com/categor y/politics/senate
	https://www.foxnews.com/opinion/david-bossie-trump-impeachment-vote-is-democratic-declaration-of-war-republicans-must-declare-war-on-dems	https://www.foxnews.com/categor y/politics/house-of-representatives

	1 2	https://www.foxnews.com/politics/trump-antifa-terrorist-portland-protests  https://www.foxnews.com/world/iranian-cruise-missiles-drones-saudi-oil  https://www.foxnews.com/media/	https://www.foxnews.com/science https://www.foxnews.com
	3	fox-nation-patriot-awards- american-heroes-st-petersburg	https://www.foxnews.com/tech
	4	https://www.foxnews.com/media/ fox-nation-patriot-awards	https://www.foxnews.com/categor y/entertainment/movies
Patrio t	5	https://www.foxnews.com/media/ marine-vet-most-valuable-patriot- awards-congress	https://www.foxnews.com/categor y/science/planet-earth
Move ment	6	https://www.foxnews.com/faith- values/veterans-serve-disaster- relief	https://www.foxnews.com/enterta inment/gwyneth-paltrow- relationship-harvey-weinstein
	7	https://www.foxnews.com/media/ tomi-lahren-patriot-awards-fox- nation	https://www.foxnews.com/politics/aoc-bloomberg-purchase-election
	8	https://www.foxnews.com/enterta inment/netflix-pulls-patriot-act- with-hasan-minhaj-episode-after- getting-valid-legal-complaint-from- saudi-arabia	https://www.foxnews.com/world/ ukraine-russia-separatists-pullback
	9	https://www.foxnews.com/media/ vice-president-mike-pence-fox- nation-patriot-awards	https://www.foxnews.com/science/earths-mantle-and-crust-are-in-a-fiery-battle-to-the-death-of-supercontinents

		https://www.foxnews.com/tech/ne tflix-defends-pulling-patriot-act- episode-in-saudi-arabia	https://www.foxnews.com/auto/2 020-ram-1500-built-to-serve- pickup-salutes-the-u-s-military
	1	https://www.foxnews.com/politics/house-republicans-storm-impeachment-inquiry-deposition-what-went-down	https://www.foxnews.com/opinion
	2	https://www.foxnews.com/opinion/tom-del-beccaro-republicans-must-win-the-impeachment-trial-and-they-can-by-following-these-five-steps	https://www.foxnews.com
Repu	3	https://www.foxnews.com/politics/republicans-hunter-biden-ukraine-whisteleblower-impeachment-witness	https://www.foxnews.com/politics
blican s	4	https://www.foxnews.com/politics/house-republicans-plan-to-call-adam-schiff-to-testify-in-impeachment-inquiry-call-him-a-fact-witness	https://www.foxnews.com/categor y/politics/elections
	5	https://www.foxnews.com/politics/trump-settles-score-with-republicans-who-didnt-embrace-him-in-midterms	https://www.foxnews.com/categor y/world/scandals
	6	https://www.foxnews.com/politics/republicans-impeachment-hearings-lineup	https://www.foxnews.com/categor y/world/conflicts
	7	https://www.foxnews.com/politics/bill-weld-thrilled-other-	https://www.foxnews.com/categor y/politics/senate

		republicans-are-challeneging- trump-in-primary	
	8	https://www.foxnews.com/enterta inment/chelsea-handler- republicans-colorado-school- shooting	https://www.foxnews.com/categor y/politics/house-of-representatives
	9	https://www.foxnews.com/opinion/matt-gorman-kavanaugh-republicans	https://www.foxnews.com/official- polls
		https://www.foxnews.com/opinion/newt-gingrich-house-republicans-2020-retirements	https://www.foxnews.com/politics
	1	https://www.foxnews.com/politics/trump-senate-trial-impeachment-inquiry-reporters-notebook	https://www.foxnews.com/opinion
	2	https://www.foxnews.com/categor y/politics/2020-senate-races	https://www.foxnews.com/world
Senat e	3	https://www.foxnews.com/media/ jeff-sessions-former-embattled- trump-attorney-general- announces-bid-for-old-us-senate- seat	https://www.foxnews.com/us
	4	https://www.foxnews.com/opinion/newt-gingrich-left-viciously-attacks-mcconnell-because-hessuccessful-senate-leader-and-trump-ally	https://www.foxnews.com/science
	5	https://www.foxnews.com/faith- values/pastor-andrew-brunson- senate-trump-prayer	https://www.foxnews.com/shows

	6	https://www.foxnews.com/politics /jeff-sessions-alabama-senate- campaign-expected	https://www.foxnews.com/health
	7	https://www.foxnews.com/politics/trump-sessions-against-senate-run-alabama	https://www.foxnews.com/travel
	8	https://www.foxnews.com/categor y/politics/elections/senate	https://www.foxnews.com/enterta inment
	9	https://www.foxnews.com/politics /virginia-democrat-accused-of-sex- with-teen-secretary-wins-state- senate-seat	https://www.foxnews.com
		https://www.foxnews.com/media/ trump-senate-mike-pence-joe- scarborough	https://www.foxnews.com/tech
	1	https://www.foxnews.com/sports/alex-morgan-2020-olympics-child-birth	https://www.foxnews.com/opinion
Olym	2	https://www.foxnews.com/sports/naomi-osaka-us-citizenship-japan-2020-olympics	https://www.foxnews.com/enterta inment
pics 2020	3	https://www.foxnews.com/categor y/sports/olympics	https://www.foxnews.com
	4	https://www.foxnews.com/tech/russian-hackers-attack-possible-olympic-ban	https://www.foxnews.com/tech
	5	https://www.foxnews.com/categor y/sports/soccer	https://www.foxnews.com/politics

	6	https://www.foxnews.com/person/c/barnini-chakraborty	https://www.foxnews.com/categor y/health/heart-health
	7	https://www.foxnews.com/categor y/person/alex-morgan	https://www.foxnews.com/lifestyle
	8	https://www.foxnews.com/categor y/sports/tennis	https://www.foxnews.com/categor y/entertainment/tv
	9	https://www.foxnews.com/sports/kikkan-randall-new-york-city-marathon-skiing-cancer	https://www.foxnews.com/categor y/tech/topics/video-games
	1 0	https://www.foxnews.com/sports	https://www.foxnews.com/categor y/politics/elections
	1	https://www.foxnews.com/tech/jef f-bezos-to-be-no-2-richest-behind- bill-gates-after-amazon-stock-drop	https://www.foxnews.com
	2	https://www.foxnews.com/transcript/cavuto-looking-back-at-the-stock-market-crash-30-years-ago	https://www.foxnews.com/categor y/us/economy
Stock	3	https://www.foxnews.com/transcript/cavuto-the-markets-are-easier-to-take-stock-of-than-life	https://www.foxnews.com/science/revolutionary-war-rifle-stolen-50-years-ago-on-display
	4	https://www.foxnews.com/categor y/shows/your- world/transcript/common-sense	https://www.foxnews.com/enterta inment/demi-lovato-first- interview-year
	5	https://www.foxnews.com/categor y/tech/companies/amazon	https://www.foxnews.com/media/ chris-wallace-elizabeth-warren- medicare-for-all-plan
	6	https://www.foxnews.com/person/c/gordon-chang	https://www.foxnews.com/tech/hp-recalls-more-laptop-lithium-ion-

			<u>batteries-after-reports-of-</u>
			overheating-melting-and-charring
			https://www.foxnews.com/tech/he
	7	https://www.foxnews.com/categor y/tech/companies/microsoft	<u>res-how-the-trump-tariffs-hike-</u> <u>may-affect-prices-for-pc-parts</u>
		y/tecn/companies/inicrosorc	inay-anect-prices-ior-pc-parts
	Ω	https://www.foxnews.com/categor y/person/sheryl-sandberg	https://www.foxnews.com/categor y/tech/companies/amazon
	0		y/tecn/companies/amazon
		https://www.foxnews.com/auto/fi	
		<u>at-chrysler-sales-boss-reid-bigland-</u> <u>sues-company-for-violating-</u>	https://www.foxnews.com/opinion
	9	whistleblower-protections	/max-lucado-fearless-parenting
		https://www.fovnovyc.com/cnorts/	
	1	https://www.foxnews.com/sports/ royals-hire-ex-cardinals-skipper-	https://www.foxnews.com/auto/el
		mike-matheny-as-manager	ectric-ford-mustang-lithium
		https://www.foxnews.com/tech/co	
	_	<u>mputer-virus-cripples-tribune-</u>	
	1	<u>publishing-los-angeles-times</u>	https://www.foxnews.com/science
		https://www.foxnews.com/science	
	ว	/deadly-seal-virus-spreading-	https://www.foxnews.com/categor
	2	melting-arctic-ice	<u>y/science/wild-nature</u>
Virus		https://www.foxnews.com/health/	
	3	<u>connecticut-third-mosquito-eee-</u> <u>death</u>	https://www.foxnews.com/categor y/tech/topics/computers
	<i>J</i>		
	1	https://www.foxnews.com/categor	https://www.foxnews.com/categor
	4	<u>y/tech/topics/virus</u>	<u>y/science/planet-earth</u>
		https://www.foxnews.com/health/	https://www.foxnews.com/science
	5	<u>eee-eastern-equine-encephalitis-</u> <u>massachusetts-man-killed</u>	<u>/deadly-seal-virus-spreading-melting-arctic-ice</u>
	ر	massachusetts-man-kmeu	mercing arene-ice

6	https://www.foxnews.com/health/what-is-norovirus-3-things-to-know-about-the-vomit-inducing-bug	https://www.foxnews.com/science/sandra-orangutan-legal-personhood-new-home-florida
7	https://www.foxnews.com/health/ pregnant-women-who-get-flu-shot- are-less-likely-to-be-hospitalized- from-virus-cdc-says	https://www.foxnews.com/science/ancient-shipwrecks-stone-pyramid-anchors-discovered-greek-island
8	https://www.foxnews.com/categor y/health/infectious-disease/zika	https://www.foxnews.com/science/earths-mantle-and-crust-are-in-a-fiery-battle-to-the-death-of-supercontinents
9	https://www.foxnews.com/health/flu-season-2019-vaccine	https://www.foxnews.com/science /mammoth-traps-prehistoric- mexico
	https://www.foxnews.com/health/ british-mom-son-flu	https://www.foxnews.com/science /mercury-transit-2019-what-to- know

## 5. Overlap Graph:

After running a python script on the above table, I found the following overlap statistics:

{'Cannes': 4,

'Congress': 0,

'Democrats': 0,

'Patriot Movement': 0,

'Republicans': 0,

'Senate': 0,

'Olympics 2020': 0,

'Stock': 1,

'Virus': 1}

# Overlap Graph:

