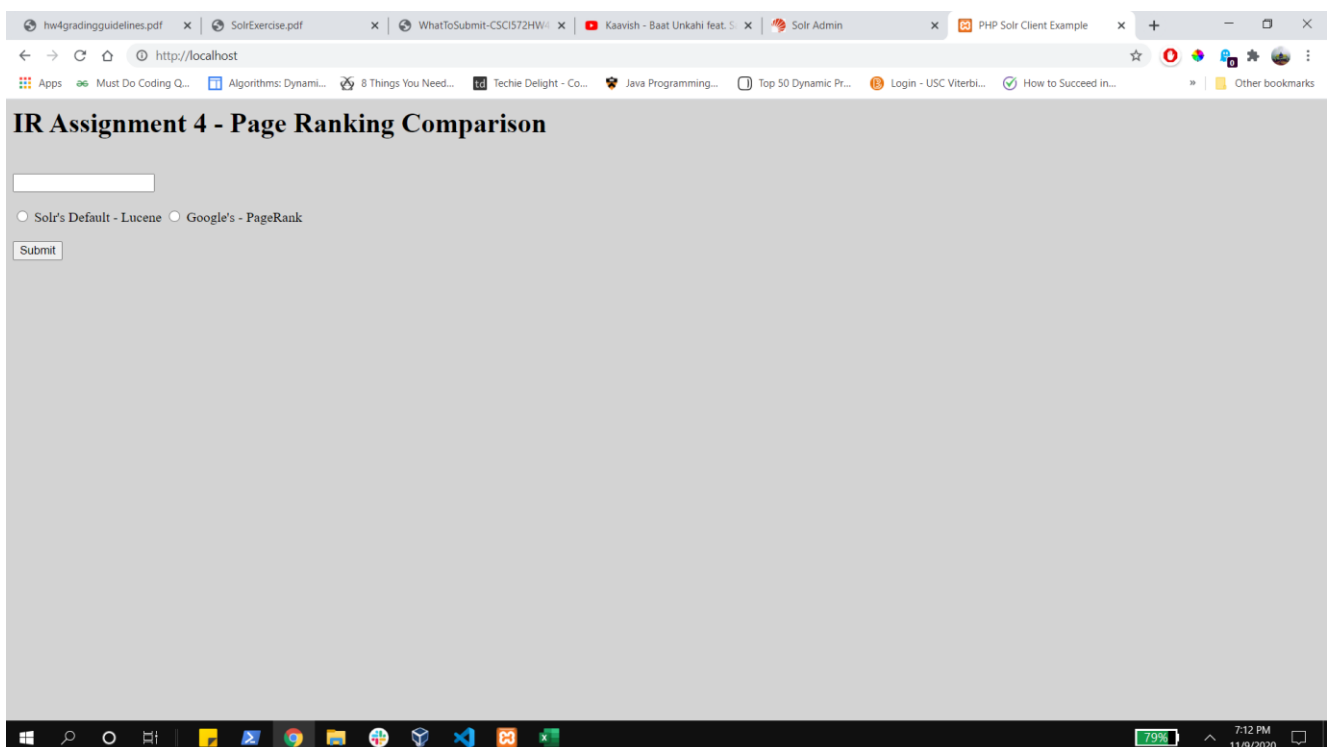


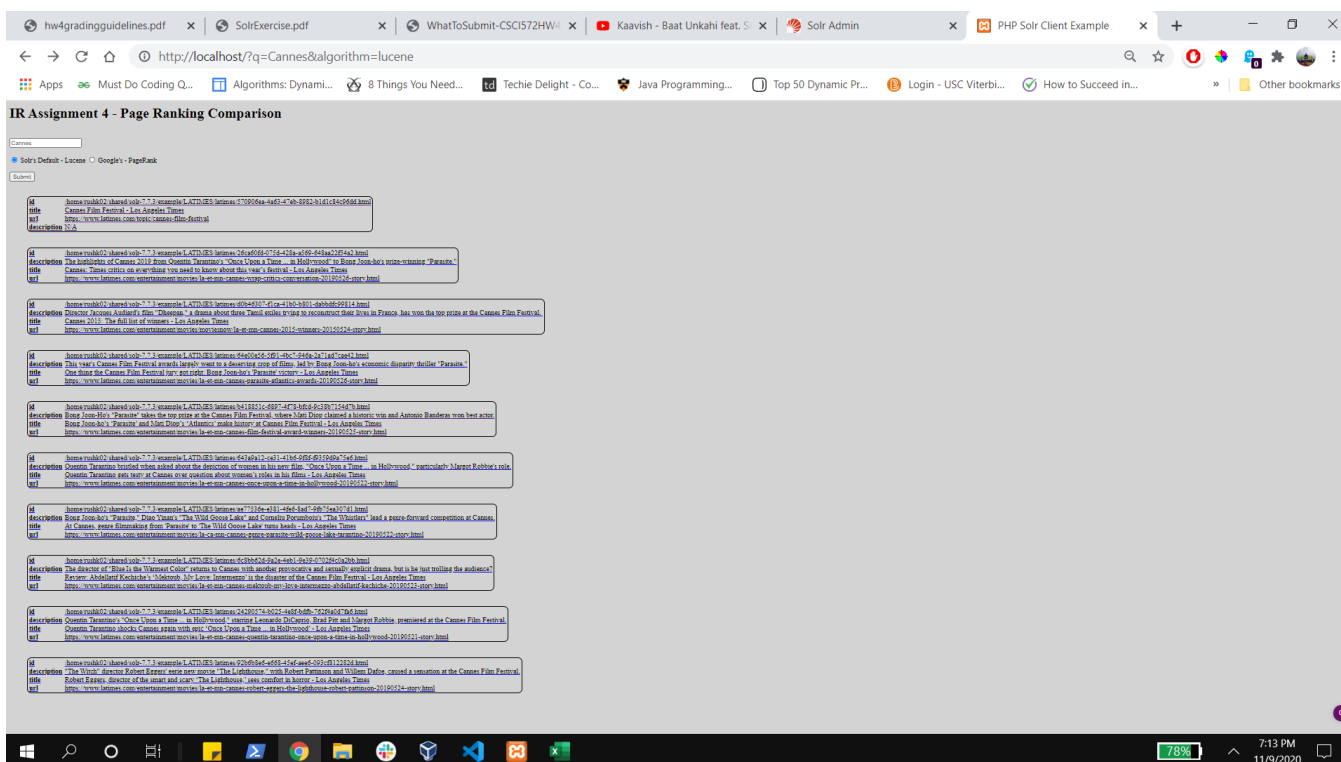
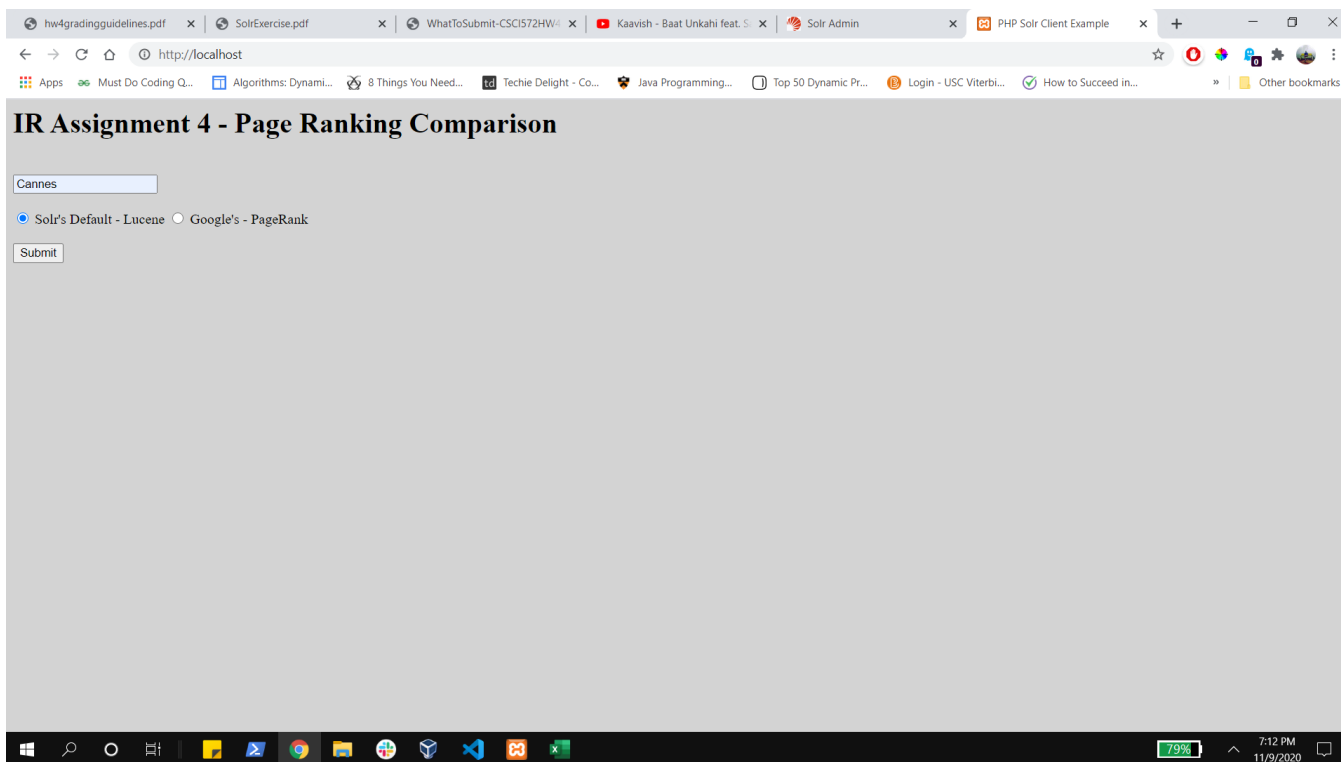
Homework 4 Report

- Steps Followed:

Initially I installed VirtualBox, ubuntu. Then I installed JDK 8 on ubuntu followed by Solr v7.7.3 for the assignment. To make the assignment simpler, I enabled bidirectional data transfer and set a shared folder where I stored all the HTML files and the required assignment files. I just ran the Solr server for indexing the HTML files in ubuntu and performed all the EdgeList and PageRank file calculations on my windows laptop. First, I created a java file for generating the EdgeList by using the CSV and the HTML files. I used the mappings to find the html files and then parsed the file using Jsoup to find all the outgoing links to other files, writing each link into a new EdgeList. After getting the EdgeList, I ran the python script to use networkx and run the pagerank algorithm to generate all the edges. I used the parameters $\alpha=0.85$, $\text{personalization}=\text{None}$, $\text{max_iter}=30$, $\text{tol}=1e-06$, $\text{nstart}=\text{None}$, $\text{weight}=\text{'weight'}$, $\text{dangling}=\text{None}$, mentioned in the homework description document. After generating the pageRank file, I followed the instructions mentioned to setup Solr with pageRank values and used the template PHP code and the internet to create the PHP website for the assignment.

- Some of the pages have higher pageRank due to the number of quality web pages hyperlinking to it, the web page content and keyword tags the page has. There are additional factors as well but the primary reason for some pages to have higher pageRank is having higher number of inlinks to the page. As we saw from the assignment, the sitemap and the general category webpages had a lot of inlinks and so, had high pageRank.
- Images documenting the flow –







1. Query: Cannes

LUCENE	PAGERANK
https://www.latimes.com/topic/cannes-film-festival	https://www.latimes.com/sitemap
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-wrap-critics-conversation-20190526-story.html	https://www.latimes.com/espanol/entretenimiento/articulo/2019-10-04/por-fin-llega-joker-aqui-un-resumen-de-todo-el-drama-que-ha-provocado
https://www.latimes.com/entertainment/movies/moviesnow/la-et-mn-cannes-2015-winners-20150524-story.html	https://www.latimes.com/entertainment-arts/movies/story/2019-11-08/documentaries-apollo-11-the-cave-sea-of-shadows-honeyland
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-parasite-atlantics-awards-20190526-story.html	https://www.latimes.com/lifestyle/gallery/chloe-sevignys-five-favorite-frocks
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-film-festival-award-winners-20190525-story.html	https://www.latimes.com/topic/film-festivals
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-once-upon-a-time-in-hollywood-20190522-story.html	https://www.latimes.com/topic/cannes-film-festival
https://www.latimes.com/entertainment/movies/la-ca-mn-cannes-genre-parasite-wild-goose-lake-tarantino-20190522-story.html	https://www.latimes.com/people/genaro-molina
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-mektoub-my-love-intermezzo-abdellatif-kechiche-20190523-story.html	https://www.latimes.com/entertainment-arts/music/story/2019-11-06/whitney-houston-lesbian-relationship-robyn-crawford
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-quentin-tarantino-once-upon-a-time-in-hollywood-20190521-story.html	https://www.latimes.com/entertainment-arts/movies/story/2019-10-31/apple-disney-hbomax-streaming-movies
https://www.latimes.com/entertainment/movies/la-et-mn-cannes-robert-eggers-the-lighthouse-robert-pattinson-20190524-story.html	https://www.latimes.com/opinion/story/2019-09-17/patt-morrison-dana-thomas-fast-fashion-environment

2. Query: Congress

LUCENE	PAGERANK
https://www.latimes.com/politics/story/2019-10-28/jerry-brown-to-testify-to-congress-rebutting-trumps-criticism-of-california	https://www.latimes.com/opinion
https://www.latimes.com/opinion/enterthefray/la-ol-new-york-rhode-island-congress-marijuana-20190116-story.html	https://www.latimes.com/espanol/politica
https://www.latimes.com/politics/la-na-pol-congress-sexual-harassment-20181212-story.html	https://www.latimes.com/politics/story/2019-11-08/michael-bloomberg-files-papers-for-2020-democratic-presidential-primary

https://www.latimes.com/politics/la-pol-ca-richest-california-lawmakers-20180305-story.html	https://www.latimes.com/opinion/story/2019-11-09/facebook-twitter-political-ads-lies
https://www.latimes.com/politics/story/2019-09-09/congress-gun-control-government-shutdown	https://www.latimes.com/politics/story/2019-11-08/elizabeth-warren-used-bare-knuckle-tactics-to-take-down-an-obama-nominee
https://www.latimes.com/opinion/story/2019-10-16/democratic-debate-trump-congress-executive-action	https://www.latimes.com/opinion/story/2019-11-09/devin-nunes-impeachment-witness-list
https://www.latimes.com/politics/story/2019-10-14/trump-ukraine-aid-congress-impeachment	https://www.latimes.com/environment/story/2019-11-08/eastern-sierra-towns-recreation-tourism-forest-service
https://www.latimes.com/politics/la-na-pol-congress-bioweapons-detection-system-20190414-story.html	https://www.latimes.com/politics/story/2019-11-08/steve-bannon-roger-stone-trump-campaign-wikileaks
https://www.latimes.com/politics/la-na-pol-congress-mueller-pressure-impeachment-20190529-story.html	https://www.latimes.com/opinion/story/2019-11-08/impeachment-trump-republicans-toomey
https://www.latimes.com/politics/la-pol-ca-richest-in-congress-darrell-issa-story.html	https://www.latimes.com/politics/story/2019-11-08/alexander-vindman-fiona-hill-impeachment-transcripts

3. Query: Democrats

LUCENE	PAGERANK
https://www.latimes.com/politics/story/2019-11-07/house-democrats-subpoena-mick-mulvaney-in-impeachment-probe	https://www.latimes.com/california
https://www.latimes.com/opinion/story/2019-11-05/impeachment-donald-trump-democrats-political-cost-elections	https://www.latimes.com/opinion
https://www.latimes.com/politics/story/2019-11-03/iowa-democrats-candidate-beat-trump	https://www.latimes.com/politics
https://www.latimes.com/politics/story/2019-10-17/democrats-quick-impeachment-timing-complicated	https://www.latimes.com/politics/story/2019-11-08/michael-bloomberg-files-papers-for-2020-democratic-presidential-primary
https://www.latimes.com/politics/story/2019-10-08/impeachment-trump-democrats-whistleblower-identity-testimony	https://www.latimes.com/politics/story/2019-11-08/democratic-primary-debate-moves-from-ucla-to-loyola-marymount-university
https://www.latimes.com/politics/story/2019-09-24/democrats-and-pelosi-appear-close-to-tipping-point-on-impeachment	https://www.latimes.com/politics/story/2019-11-08/elizabeth-warren-used-bare-knuckle-tactics-to-take-down-an-obama-nominee
https://www.latimes.com/politics/story/2019-09-29/trump-allies-and-democrats-reveal-the-deep-divisions-over-impeachment-inquiry	https://www.latimes.com/opinion/story/2019-11-09/elizabeth-warren-trump

https://www.latimes.com/politics/story/2019-08-01/assignment-post-debate-health-care-analysis-new-story	https://www.latimes.com/opinion/story/2019-11-09/devin-nunes-impeachment-witness-list
https://www.latimes.com/nation/ct-democrats-economic-plan-20170824-story.html	https://www.latimes.com/politics/story/2019-11-08/steve-bannon-roger-stone-trump-campaign-wikileaks
https://www.latimes.com/california/story/2019-10-21/california-independent-voters-can-cast-ballots-for-democrats-not-trump-march-2020-primary	https://www.latimes.com/opinion/story/2019-11-08/impeachment-trump-republicans-toomey

4. Query: Patriot Movement

LUCENE	PAGERANK
https://www.latimes.com/archives/la-xpm-2012-oct-23-la-me-russell-means-20121023-story.html	https://www.latimes.com/environment
https://www.latimes.com/staff/megan-garvey	https://www.latimes.com/science
https://www.latimes.com/archives/la-xpm-2001-jun-10-mn-8792-story.html	https://www.latimes.com/california/story/2019-11-09/prop-187-anniversary-commemoration
https://www.latimes.com/entertainment-arts/books/story/2019-10-25/shadowlands-anthony-mccann-oregon-standoff	https://www.latimes.com/entertainment-arts/story/2019-11-09/classical-music-things-to-do-in-la-this-week-nov-10-17-magic-flute-la-opera
https://www.latimes.com/sports/highschool/story/2019-11-06/girls-tennis-southern-section-playoff-results-and-updated-pairings	https://www.latimes.com/entertainment-arts/story/2019-11-09/theater-things-to-do-in-la-this-week-nov-10-17-dr-seuss-how-the-grinch-stole-christmas-old-globe-key-largo-andy-garcia-geffen-playhouse
https://www.latimes.com/politics/story/2019-09-20/bernie-sanders-muslim-voters-2020	https://www.latimes.com/food/story/2019-11-07/kung-pao-chicken-history-recipe-gong-bao
https://www.latimes.com/sports/nba/la-sp-nba-best-game-ever-20181222-story.html	https://www.latimes.com/obituaries/story/2019-11-04/james-stern-obit-black-activist-who-led-neo-nazi-group-dies-amid-bid-to-destroy-it
https://www.latimes.com/sports/highschool/story/2019-11-06/girls-tennis-southern-section-playoff-results-and-updated-pairings	https://www.latimes.com/topic/column-one
https://www.latimes.com/entertainment/movies/la-et-mn-july-4-box-office-history-20180629-story.html	https://www.latimes.com/topic/mexico-america
https://www.latimes.com/entertainment/movies/la-et-mn-july-4-box-office-history-20180629-story.html	https://www.latimes.com/topic/museums

5. Query: Republicans

LUCENE	PAGERANK
https://www.latimes.com/opinion/story/2019-10-23/house-republicans-storm-hearing-impeachment	https://www.latimes.com/
https://www.latimes.com/politics/story/2019-10-30/democrats-hoped-theyd-win-over-republicans-on-impeachment-but-its-not-looking-that-way-so-far	https://www.latimes.com/opinion
https://www.latimes.com/opinion/story/2019-11-08/impeachment-trump-republicans-toomey	https://www.latimes.com/opinion/la-letter-to-the-editor-htmlstory.html
https://www.latimes.com/politics/story/2019-10-04/vulnerable-senate-republicans-impeachment	https://www.latimes.com/opinion/story/2019-11-09/facebook-twitter-political-ads-lies
https://www.latimes.com/california/story/2019-10-16/california-republicans-democrats-wildfires-homelessness	https://www.latimes.com/opinion/story/2019-11-08/berlin-wall-30th-anniversary-cold-war-donald-trump
https://www.latimes.com/politics/story/2019-10-23/impeachment-deposition-room-stormed-by-republicans	https://www.latimes.com/politics/story/2019-11-08/elizabeth-warren-used-bare-knuckle-tactics-to-take-down-an-obama-nominee
https://www.latimes.com/opinion/story/2019-09-17/california-republicans-trump-fundraising-president	https://www.latimes.com/opinion/story/2019-11-09/elizabeth-warren-trump
https://www.latimes.com/politics/story/2019-10-28/katie-hill-resignation-trump-obstacle-for-republicans	https://www.latimes.com/opinion/story/2019-11-09/devin-nunes-impeachment-witness-list
https://www.latimes.com/opinion/story/2019-11-06/trump-republicans-democrats-predictions-polling-2020	https://www.latimes.com/opinion/story/2019-11-09/sat-uc-college-admissions
https://www.latimes.com/politics/story/2019-11-06/republicans-election-trump-candidate-loses-kentucky	https://www.latimes.com/opinion/story/2019-11-08/impeachment-trump-republicans-toomey

6. Query: Senate

LUCENE	PAGERANK
https://www.latimes.com/politics/story/2019-11-07/gop-is-already-thinking-about-how-to-turn-a-senate-impeachment-trial-to-trumps-advantage	https://www.latimes.com/

https://www.latimes.com/politics/story/2019-10-30/california-donors-spend-millions-on-2020-senate-races-across-the-country	https://www.latimes.com/opinion
https://www.latimes.com/politics/story/2019-10-04/vulnerable-senate-republicans-impeachment	https://www.latimes.com/california/story/2019-11-09/prop-187-anniversary-commemoration
https://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html	https://www.latimes.com/opinion/la-letter-to-the-editor-htmlstory.html
https://www.latimes.com/nation/la-pol-scotus-confirmation-votes-over-the-years-20181005-htmlstory.html	https://www.latimes.com/opinion/story/2019-11-09/facebook-twitter-political-ads-lies
https://www.latimes.com/politics/story/2019-08-14/2020-senate-control-presidential-race	https://www.latimes.com/opinion/story/2019-11-08/berlin-wall-30th-anniversary-cold-war-donald-trump
https://www.latimes.com/politics/story/2019-08-14/john-hickenlooper-quits-presidential-race-for-senate-run	https://www.latimes.com/politics/story/2019-11-08/elizabeth-warren-used-bare-knuckle-tactics-to-take-down-an-obama-nominee
https://www.latimes.com/politics/story/2019-10-30/california-donors-spend-millions-on-2020-senate-races-across-the-country	https://www.latimes.com/opinion/story/2019-11-09/elizabeth-warren-trump
https://www.latimes.com/politics/la-na-pol-william-barr-senate-confirm-attorney-general-20190214-story.html	https://www.latimes.com/opinion/story/2019-11-09/devin-nunes-impeachment-witness-list
https://www.latimes.com/opinion/story/2019-10-15/wmcaleen-homeland-security-trump-nielsen-senate-acting-secretary	https://www.latimes.com/opinion/story/2019-11-09/sat-uc-college-admissions

7. Query: Olympics 2020

LUCENE	PAGERANK
https://www.latimes.com/sports/olympics/story/2019-08-15/2020-tokyo-olympics-searching-for-answers-amid-heat-wave	https://www.latimes.com/
https://www.latimes.com/sports/olympics	https://www.latimes.com/sitemap
https://www.latimes.com/people/david-wharton	https://www.latimes.com/sports
https://www.latimes.com/people/david-wharton	https://www.latimes.com/opinion

https://www.latimes.com/sports/olympics/story/2019-10-16/2020-tokyo-olympics-marathon-sapporo-avoid-heat	https://www.latimes.com/business
https://www.latimes.com/sports/olympics/story/2019-07-31/deadly-heat-wave-2020-summer-olympics-tokyo	https://www.latimes.com/entertainment-arts
https://www.latimes.com/sports/story/2019-11-05/ioc-anti-doping-2020-tokyo-olympics	https://www.latimes.com/world-nation
https://www.latimes.com/sports/olympics/story/2019-10-10/naomi-osaka-chooses-japan-over-u-s-2020-tokyo-olympics	https://www.latimes.com/environment
https://www.latimes.com/sports/story/2019-07-23/tokyos-rough-road-to-2020-summer-olympics	https://www.latimes.com/politics
https://www.latimes.com/people/david-wharton	https://www.latimes.com/newsroom-directory

8. Query: Stock

LUCENE	PAGERANK
https://www.latimes.com/business/story/2019-11-06/uber-lock-up-period-ends-with-falling-stock-protests	https://www.latimes.com/
https://www.latimes.com/business/la-fi-lyft-stock-20190401-story.html	https://www.latimes.com/business
https://www.latimes.com/business/la-fi-hy-tesla-stock-20160518-snap-story.html	https://www.latimes.com/business/real-estate
https://www.latimes.com/business/story/2019-11-06/uber-lock-up-period-ends-with-falling-stock-protests https://www.latimes.com/business/story/2019-11-09/stock-managers-who-played-defense-in-2019-are-left-scrambling-to-make-up-ground	https://www.latimes.com/business/real-estate/story/2019-11-09/hulu-looking-for-alaska-landry-bender https://www.latimes.com/business/story/2019-11-08/trump-china-tariffs
https://www.latimes.com/business/la-fi-uber-ipo-stock-trading-price-20190510-story.html	https://www.latimes.com/business/real-estate/story/2019-11-09/hot-property-very-brady-sale-malibu
https://www.latimes.com/business/story/2019-09-09/at-t-stock-surges-elliott-reveals-3-2-billion-stake	https://www.latimes.com/politics/story/2019-11-08/elizabeth-warren-used-bare-knuckle-tactics-to-take-down-an-obama-nominee

https://www.latimes.com/business/la-fi-stock-market-status-update-20180209-story.html	https://www.latimes.com/business/real-estate/story/2019-11-09/hot-property-newsletter-rifleman-leads-the-posse-on-a-little-mosey-down-memory-lane
https://www.latimes.com/business/la-fi-stock-market-correction-20180212-story.html	https://www.latimes.com/business/story/2019-11-09/stock-managers-who-played-defense-in-2019-are-left-scrambling-to-make-up-ground
https://www.latimes.com/business/la-fi-stock-market-cpi-roundup-20180214-story.html	https://www.latimes.com/business/story/2019-11-08/gap-ceo-was-done-in-by-fashion-missteps-and-fading-brands

9. Query: Virus

LUCENE	PAGERANK
https://www.latimes.com/espanol/internacional/articulo/2019-11-09/monos-se-volvieron-inmunes-tras-inocularles-virus-del-ebola-con-mutacion	https://www.latimes.com/terms-of-service
https://www.latimes.com/science/sciencenow/la-sci-science-of-zika-five-ways-20160223.htmlstory.html	https://www.latimes.com/espanol/
https://www.latimes.com/espanol/internacional/articulo/2019-11-09/monos-se-volvieron-inmunes-tras-inocularles-virus-del-ebola-con-mutacion	https://www.latimes.com/science
https://www.latimes.com/science/story/2019-10-21/virus-afm-illness-paralyzing-kids	https://www.latimes.com/espanol/internacional
https://www.latimes.com/science/story/2019-10-21/virus-afm-illness-paralyzing-kids	https://www.latimes.com/espanol/deportes/articulo/2019-11-09/como-ser-campeon-de-league-of-legends
https://www.latimes.com/socal/daily-pilot/news/tn-dpt-me-west-nile-20170707-story.html	https://www.latimes.com/business/story/2019-10-13/body-parts-harvesting-hinders-coroner-autopsies
https://www.latimes.com/science/sciencenow/la-sci-sn-acute-flaccid-myelitis-polio-20181017-story.html	https://www.latimes.com/espanol/mexico/articulo/2019-11-09/crean-en-mexico-plataforma-para-medir-impacto-ambiental-de-basura-electronica
https://www.latimes.com/local/lanow/la-me-In-virulent-newcastle-disease-outbreak-in-southern-california-20190607-story.html	https://www.latimes.com/espanol/internacional/articulo/2019-11-09/el-nuevo-mitsubishi-mirage-pudiera-arribar-a-mediados-del-proximo-ano

https://www.latimes.com/world-nation/story/2019-09-04/anti-vaxxers-helping-polio-comeback-pakistan	https://www.latimes.com/espanol/internacional/articulo/2019-11-09/monos-se-volvieron-inmunes-tras-inocularles-virus-del-ebola-con-mutacion
https://www.latimes.com/local/obituaries/la-me-deborah-asnis-20150921-story.html	https://www.latimes.com/espanol/internacional/articulo/2019-11-09/una-tortuga-tecnologica-para-detectar-microplasticos-en-los-oceanos

- Overlap graph

Query	Overlap Number
Cannes	1
Congress	0
Democrats	0
Patriot Movement	0
Republicans	1
Senate	0
Olympics 2020	0
Stock	1
Virus	1

