Steps

First I install Solr 7.7.3 as instructed and started it on my existing Linux Mint VM. The newest Solr version does not work so I have to use the old version according to Piazza. Then I created new core called myexample. I downloaded and indexed the files with TIKA following the instructions of indexing files. After that I used python BeautifulSoup to look for links for each file and create the edgelist file in a similar fashion of the Jsoup instructions. I used the given csv file to create index dictionaries and then parse the links for each html file, and create an edge for each link using the dictionaries. I used python instead of Java because of convenience. Then I computed Pagerank by NetworkX using Python according to the instructions with the following parameters  
  
alpha=0.85, personalization=None, max\_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None  
  
I put the Pagerank result file to the data folder of the core and modified the configurationsin the managed-schema and the solrconfig.xml files as given, then I used the given PHP code modified it for this assignment. I also tested using the “fl”:”\*, field(pagrRankFile)” to verify that the pagerank search is working as expected. I modified so that it shows the id, title, url and description fields in the search result. I added a checkbox for pagerank and also the additional parameters of “pageRankFile desc” to sort by pagerank in descending order. Therefore if the checkbox is uncheck the default Lucene ranking will be used. I put the PHP file in var/www/html folder and started the Apache server. Finally I ran the 8 queries using both versions and compare the results of both, and plotted the overlap graph according to the results. The Pagerank search usually returns irrelevant results because those pages have high Pagerank but not that relevant to the search term.

Explanation

Some pages have higher Pagerank values than others because those pages are referenced more often by the rest of the pages. For example, the sitemap and the Home page of LA times have the highest pagerank because almost all pages have a link to those pages. As a result, since PageRank is a vote of how important the page is according to all other pages, the Homepage and sitemap pages will have very high Pagerank scores as expected.

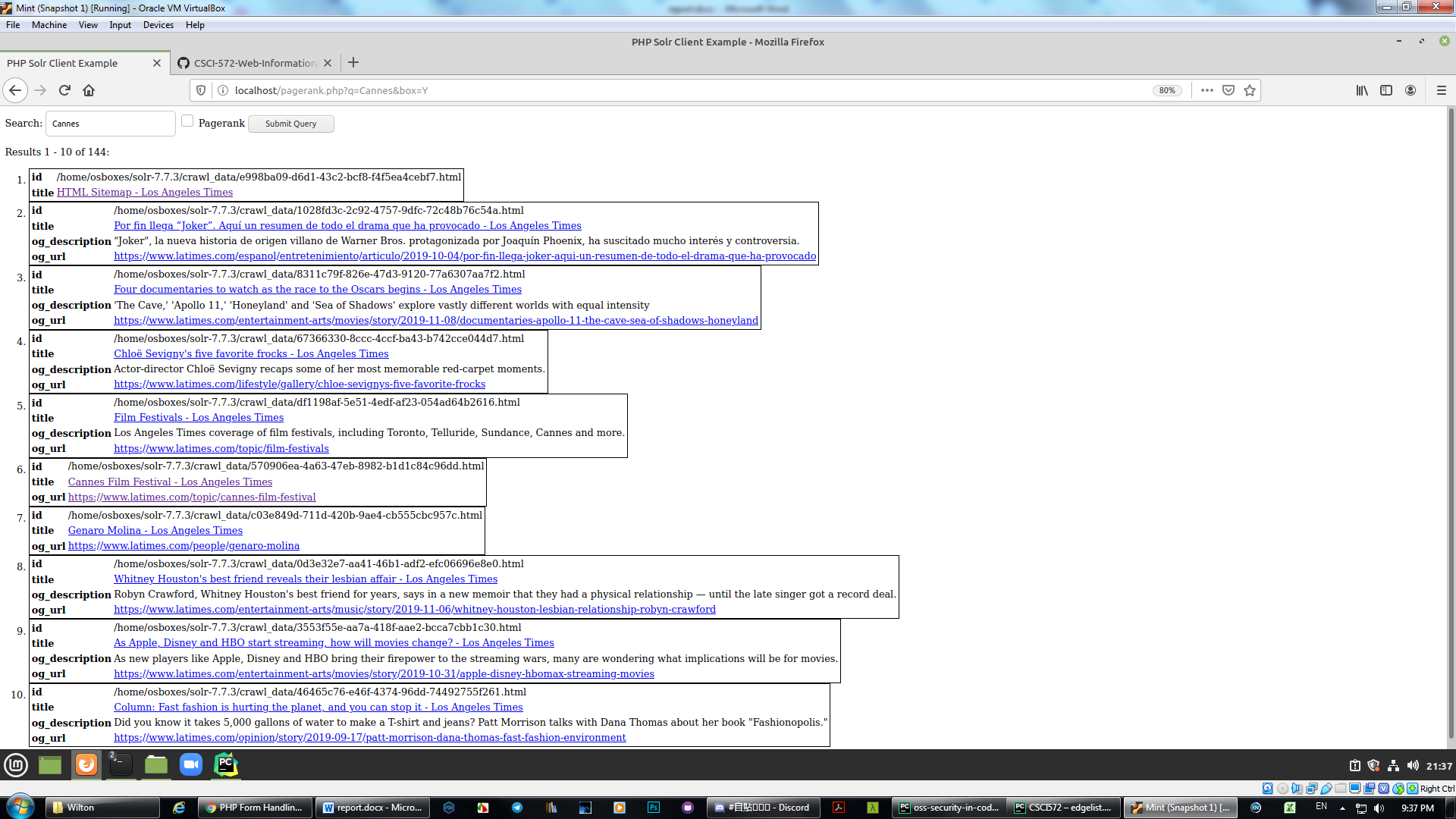
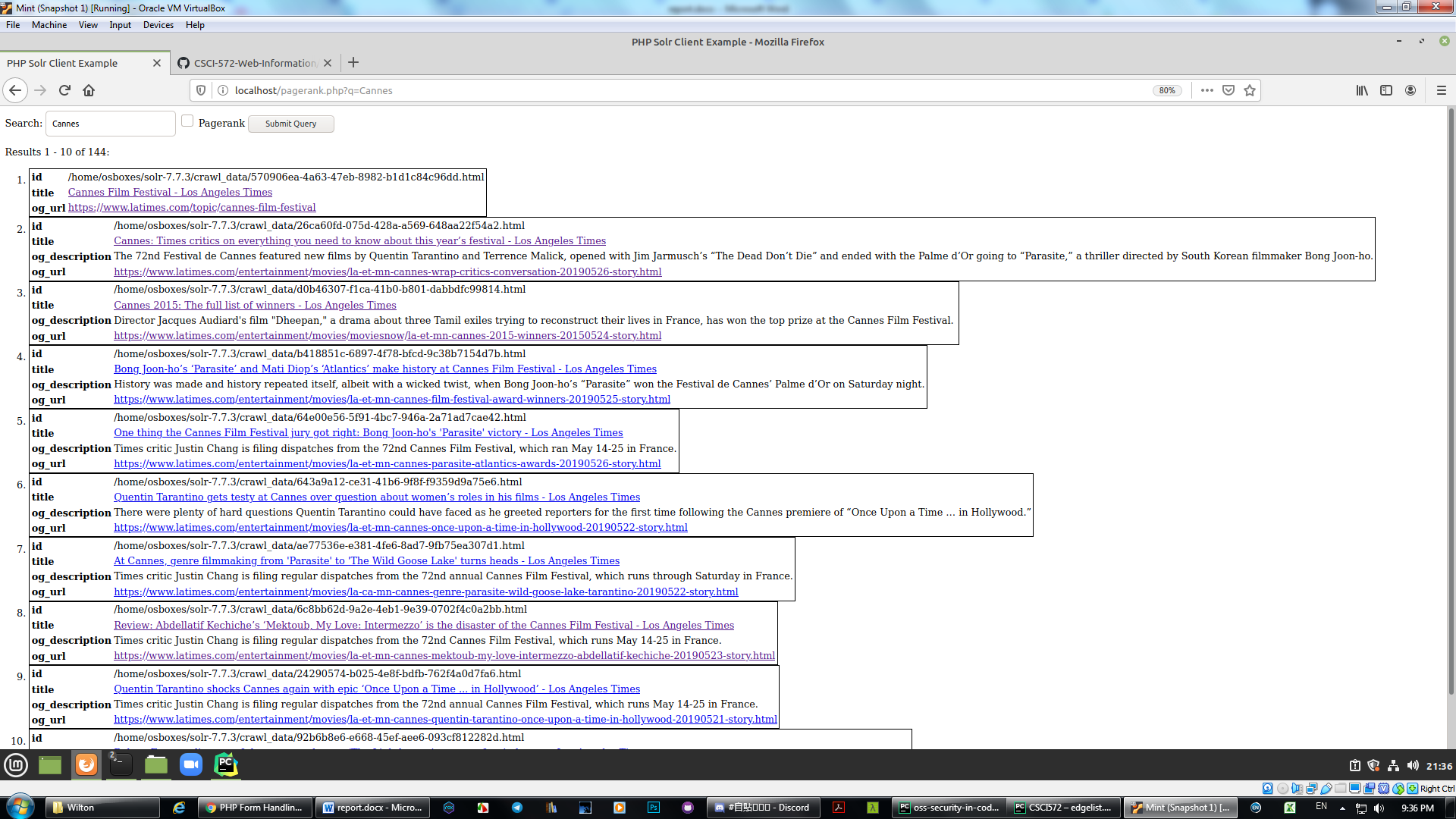
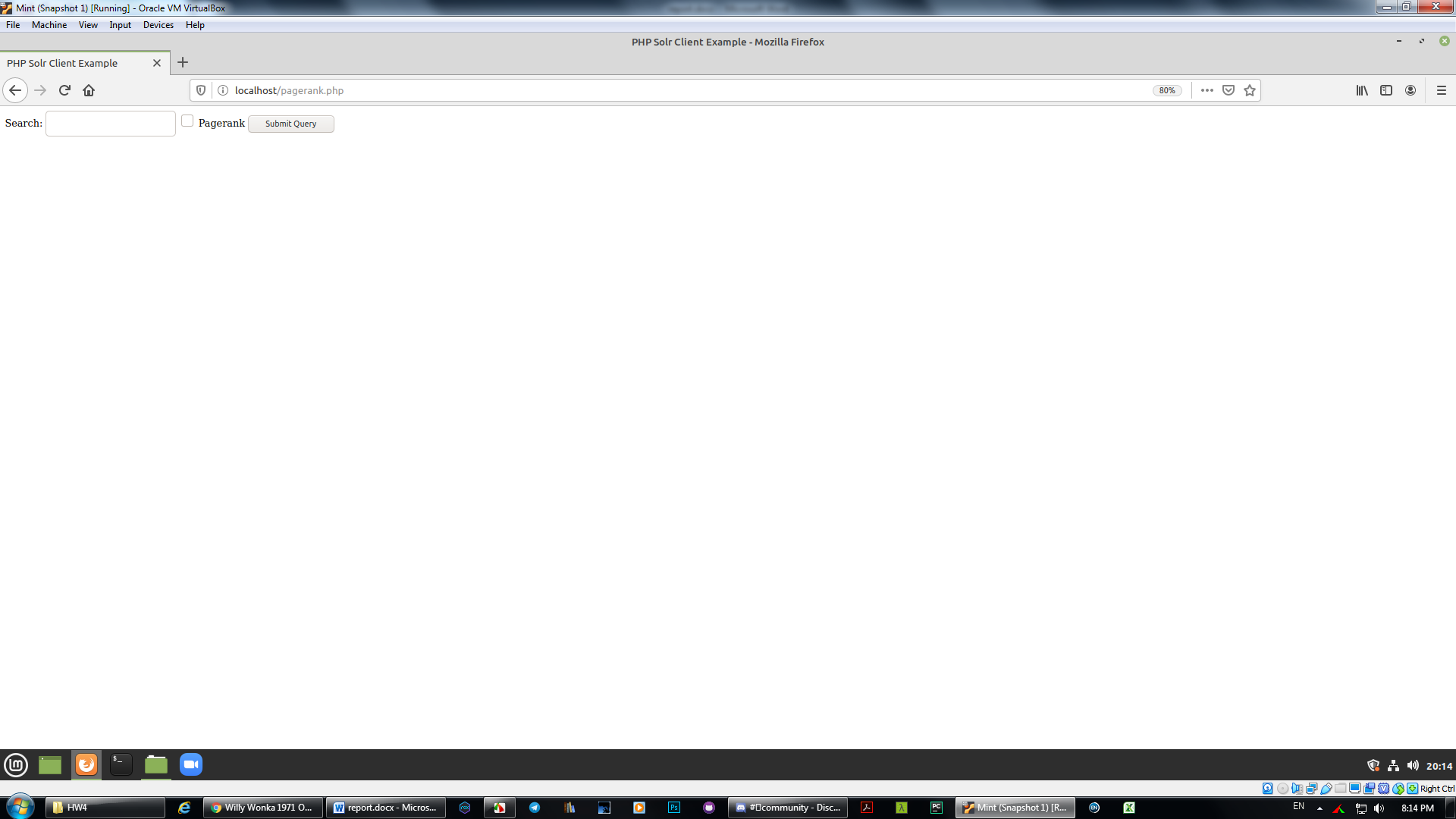
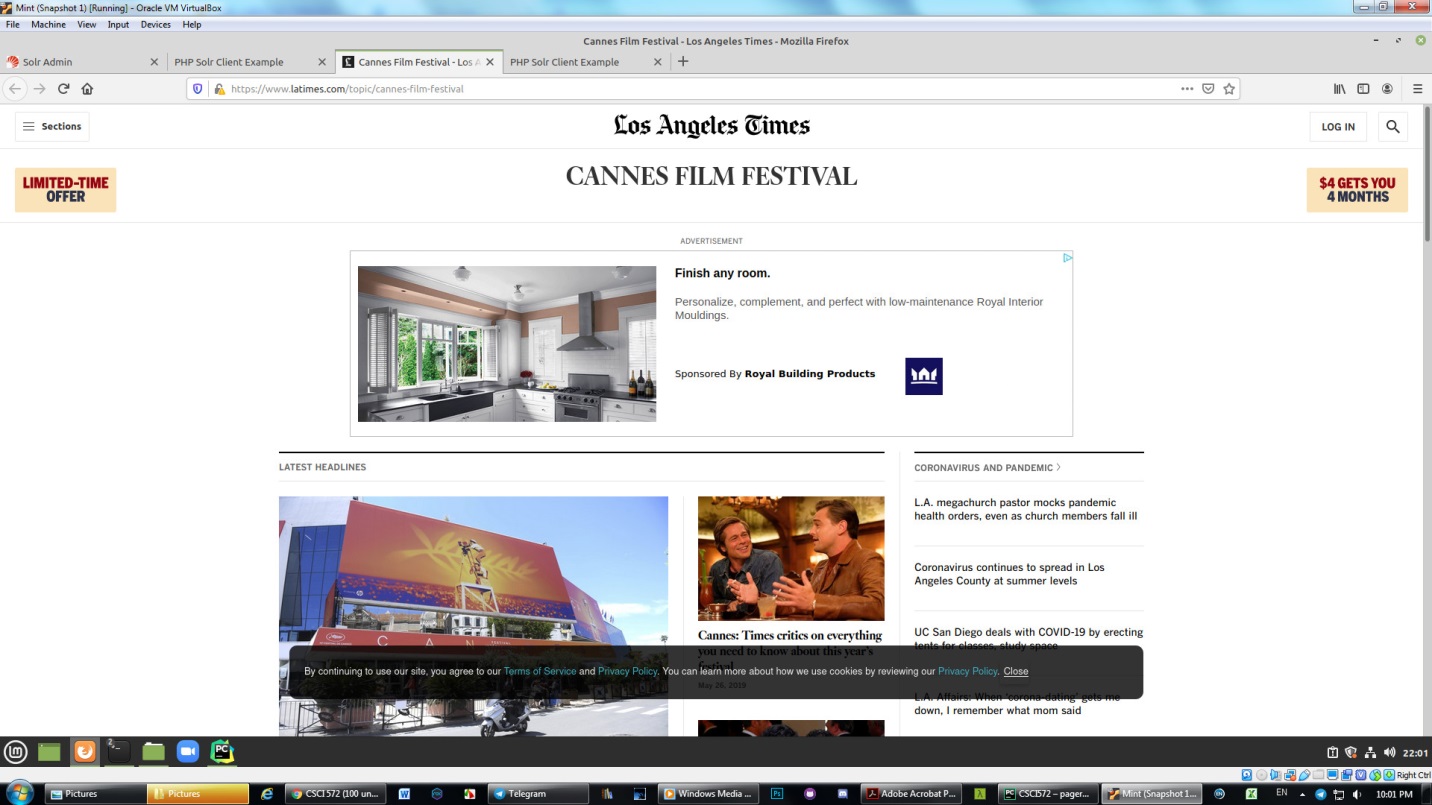
Screenshots  

Table of URLs

|  |  |
| --- | --- |
| Lucene | Pagerank |
| Cannes |  |
| 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-film-festival-award-winners-20190525-story.html | https://www.latimes.com/lifestyle/gallery/chloe-sevignys-five-favorite-frocks |
| https://www.latimes.com/entertainment/movies/la-et-mn-cannes-parasite-atlantics-awards-20190526-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 |
| Congress |  |
| 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/politics/story/2019-11-08/steve-bannon-roger-stone-trump-campaign-wikileaks |
| https://www.latimes.com/politics/la-na-pol-congress-bioweapons-detection-system-20190414-story.html | 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-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 |
| Democrats |  |
| 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 |
| Patriot Movement |  |
| 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-seusss-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-americas |
| https://www.latimes.com/entertainment/movies/la-et-mn-july-4-box-office-history-20180629-story.html | https://www.latimes.com/topic/museums |
| Republicans |  |
| 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/sat-uc-college-admissions |
| https://www.latimes.com/opinion/story/2019-11-06/trump-republicans-democrats-predictions-polling-2020 | https://www.latimes.com/opinion/story/2019-11-09/devin-nunes-impeachment-witness-list |
| 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 |
| Senate |  |
| 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/sat-uc-college-admissions |
| 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/devin-nunes-impeachment-witness-list |
| Olympics 2020 |  |
| 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/entertainment-arts |
| https://www.latimes.com/sports/olympics/story/2019-07-31/deadly-heat-wave-2020-summer-olympics-tokyo | https://www.latimes.com/business |
| 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/newsroom-directory |
| https://www.latimes.com/people/david-wharton | https://www.latimes.com/politics |
| Stock |  |
| 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/2019-11-08/trump-china-tariffs |
| 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/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/story/2019-11-08/gap-ceo-was-done-in-by-fashion-missteps-and-fading-brands |
| https://www.latimes.com/business/la-fi-stock-market-correction-20180212-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-cpi-roundup-20180214-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 |
| Virus |  |
| 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-ln-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