Assignment 4
Indexing the web using Solr
Sajal Kumar Goel
USC ID:8451918719

I) Installation:

1) Virtual Box and Ubuntu Installation
  a. Installed virtual box and Ubuntu 16.04
  b. Downloaded files from the google drive link:
  https://drive.google.com/drive/folders/14qtJywCRk00wEbmt-Pv3efyPIBqwcoZr
  c. Installed solr-8.7.0
    i. Used to index the data
    ii. Downloaded the file and unzipped it using unzip -q solr-8.7.0
    iii. Created a core by going into the /solr-8.7.0 directory and then running
       bin/solr -c create myexample
    iv. Made the foll changes in the managed-schema file in conf directory:
        1. Uncommented <copyfield source = "*" dest = "_text _" />
    v. Posted the files on solr server using the bin/post -c myexample -filetypes
       html /path/to/crawl/folder
    vi. Uncommented the line <str name ="df">_text_</str>
    vii. Reloaded the core to verify changes made were correct

2) Computing the Edge List:
  a. Wrote a program in Java to extract the link structure from downloaded web
     pages to compute EdgeList file, using JSoup library in Java.
  b. Web page files are the vertices and edge between two indicates a link between
     them.

3) Calculating Page Rank
  a. Used NetworkX library in python to compute the pagerank of webpages, from
     Edgelist(obtained from Java program above)
  b. Loaded web graph into NetwokX using read_edgelist() function with edgelist
     and digraph as parameters.
  c. Used pagerank() function to calculate Page Rank and output the result to
     external_pageRankFile.txt
  d. Following parameters were used in pagerank(): alpha=0.85,
     personalization=None, max_iter=30, tol=1e-06, nstart=None,
     weight='weight',dangling=None
  e. The external_pageRankFile.txt has format
     <document_id>=<page_rank_score>

4) Changes to managed-schema and solrconfig.xml for external page rank file
  a. Made the changes in managed-schema for using the external PageRank file for
     ranking
     <fieldType name="external" keyField="id" defVal="0" class="solr.ExternalFileField"/>

<field name="pageRankFile" type="external" stored="false" indexed="false"/>
   b. Added these lines in solrconfig.xml to reload the external file when a file is either reloaded or a new file is loaded
   <listener event = "newSearcher" class = "org.apache.solr.schema.ExternalFileFieldReloader"/>
   <listener event = "firstSearcher" class = "org.apache.solr.schema.ExternalFileFieldReloader"/>

5) Apache Server, Php Installation
   a. Installed the apache server and php on Ubuntu.
   b. Cloned the git repository in appropriate folder
   c. Started with the sample code then added radio buttons, aligned the elements and added php script to check and handle if url, id, title and description were null. The core name and path carefully written
   d. Added additional parameters 'sort' => 'pageRankFile desc' for sorting the webpage results based on the external_ pageRankFile.txt file

II) Why PageRank of some pages is more than Page Rank of others?

PageRank is an algorithm used by search engines to rank websites in their search engine results and is a way of measuring the importance of website pages. The page algorithm assigns a score to the web pages based on the number of in-links and out-links (Though recently Google uses many other factors to assign a score). If a web page is pointed by other web pages with a higher page rank score, its page rank score tends to increase. The EdgeList was used to construct a graph using the Networkx library in Python. We computed the page rank based on this graph using the PageRank function. Thus, some pages may have a higher page rank if they are pointed by other pages with high scores and relatively fewer out-links.

III) Screenshots depicting flow of process:
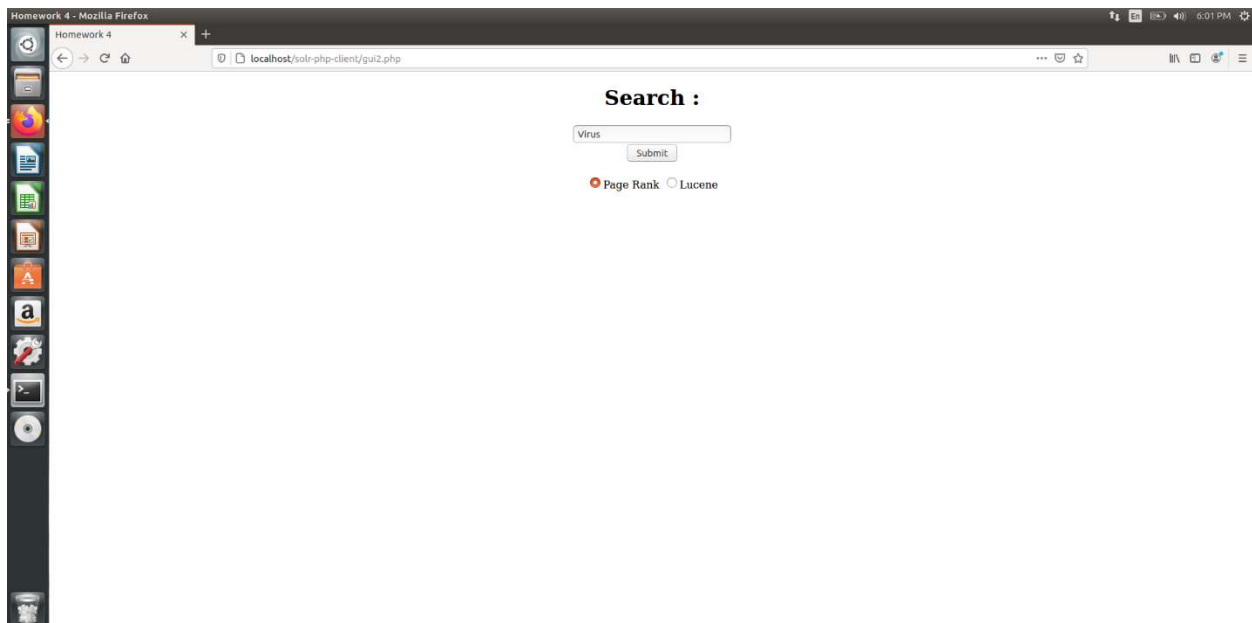
Homework 4   +

localhost/solr-php-client/gui2.php?q=Virus&sort=pagerank   90%   ··· ☆

Search :

Virus

Submit

● Page Rank ○ Lucene

Results 1 - 10 of 141 :

1. Title: The New York Times en Español - The New York Times
   URL: https://www.nytimes.com/es/
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/d4a8397e-dc08-4075-b95c-69f430a3b7e7.html
2. Title: U.S. News - The New York Times
   URL: https://www.nytimes.com/section/us
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/2b6d902b-ab03-4da1-a1ae-467b39b8b8cd.html
3. Title: World News - The New York Times
   URL: https://www.nytimes.com/section/world
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/a47bf605-297e-415f-8ee9-65b0fefb4dd7.html
4. Title: Science - The New York Times
   URL: https://www.nytimes.com/section/science
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/c9629b08-4298-4857-95ed-0fc9b7f32a0c.html
5. Title: Health - The New York Times
   URL: https://www.nytimes.com/section/health
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/1d33f834-e45d-4fa3-8794-3c1eaeb30409.html
6. Title: Climate and Environment - The New York Times
   URL: https://www.nytimes.com/section/climate
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/2f3c5ea6-7f4b-45c7-ab28-2022f611907a.html
7. Title: The Seven Commandments of Fake News - Video - NYTimes.com
   URL: https://www.nytimes.com/video/opinion/100000006188102/what-is-pizzagate.html?playlistId=100000006207751
   Description: The Pizzagate playbook: Same tactics, new technologies. How the seven rules of Soviet disinformation are being used to create today's fake news stories. Pizza anyone?
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/ea605bfe-018c-41a6-a888-2ef13792f34d.html
8. Title: The Worldwide War on Truth - Video - NYTimes.com
   URL: https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html?playlistId=100000006207751
   Description: Governments from Pakistan to Mexico to Washington are woefully unequipped to combat disinformation warfare. Eastern European countries living in Russia's shadow can teach us how to start fighting back, but only if our politicians decide to stop profiting from these tactics and fight them instead.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/fa779dc1-8b6b-4962-9089-057ad549c26b.html
9. Title: Meet the KGB Spies Who Invented Fake News - Video - NYTimes.com
   URL: https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html?playlistId=100000006207751
   Description: We reveal how one of the biggest fake news stories ever concocted — the 1984 AIDS-is-a-biological-weapon hoax — went viral in the pre-Internet era. Meet the KGB cons who invented it, and the "truth squad" that quashed it. For a bit.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/43cd8aa5-bc48-4e4b-a723-e85f6092548a.html
10. Title: The Worldwide War on Truth - Video - NYTimes.com
    URL: https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html

---

Homework 4   +

localhost/solr-php-client/gui2.php?q=Virus&sort=Lucene   90%   ··· ☆

Search :

Virus

Submit

○ Page Rank ● Lucene

Results 1 - 10 of 141 :

1. Title: How Did a Virus From the Atlantic Infect Mammals in the Pacific? - The New York Times
   URL: https://www.nytimes.com/2019/11/09/science/seals-distemper.html
   Description: Thawing sea ice may have opened the door, allowing the infection to cross oceans, a new study suggests.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/73e479de-d240-43e0-8630-e6b8351bce9f.html
2. Title: Kay Hagan, Former North Carolina Senator, Dies at 66 - The New York Times
   URL: https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html
   Description: A Democrat, she served one term after defeating Elizabeth Dole. She was found to have a tick-borne virus in 2016.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/e293ff76-9271-4362-a931-d126403384b2.html
3. Title: Scientists Were Hunting for the Next Ebola. Now the U.S. Has Cut Off Their Funding. - The New York Times
   URL: https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html
   Description: Predict, a government research program, sought to identify animal viruses that might infect humans and to head off new pandemics.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/e0f08feb-8216-4316-a138-bde83e38ecf5.html
4. Title: Trump Plan to Stop Spread of H.I.V. Will Target 'Hot Spot' Areas - The New York Times
   URL: https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html?module=inline
   Description: The plan to end transmission of the virus in the United States by 2030, outlined in the president's State of the Union address, will focus on 48 counties where about half of new infections occur.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/e59cb7d1-fa94-4ae5-88db-600de167adde.html
5. Title: Meet the KGB Spies Who Invented Fake News - Video - NYTimes.com
   URL: https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html
   Description: We reveal how one of the biggest fake news stories ever concocted — the 1984 AIDS-is-a-biological-weapon hoax — went viral in the pre-Internet era. Meet the KGB cons who invented it, and the "truth squad" that quashed it. For a bit.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/d20bb69f-dd88-4b1a-a09e-52a294c7f549.html
6. Title: Meet the KGB Spies Who Invented Fake News - Video - NYTimes.com
   URL: https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html?playlistId=100000006207751
   Description: We reveal how one of the biggest fake news stories ever concocted — the 1984 AIDS-is-a-biological-weapon hoax — went viral in the pre-Internet era. Meet the KGB cons who invented it, and the "truth squad" that quashed it. For a bit.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/43cd8aa5-bc48-4e4b-a723-e85f6092548a.html
7. Title: Donald G. McNeil Jr.
   URL: https://www.nytimes.com/by/donald-g-mcneil-jr
   Description: Array
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/54a0fb20-865c-4522-82d6-802fceeb209e.html
8. Title: When Is the Best Time to Get Your Flu Shot? - The New York Times
   URL: https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html
   Description: Immunity to the flu virus may wane over the course of the flu season, so you don't want to get a flu shot too early, or too late.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/82a5705a-fcb2-4173-a99b-450a4eeab8f7.html
9. Title: Her Daughter, Born With a Dangerous Infection, Needs Close Care - The New York Times
   URL: https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html
   Description: Separated from her husband, with few others to rely on, Luisa Estefany Jimenez struggles to raise three daughters, including one with special needs.
   ID: /home/sajal/Downloads/solr-8.7.0/crawl_data/90cac8ee-3c87-4c33-95a3-8f8fded256e0.html
10. Title: Trump Expected to Announce Plan to Stop Spread of H.I.V. in the U.S. by 2030 - The New York Times
    URL: https://www.nytimes.com/2019/02/04/us/politics/trump-hiv-aids-plan.html?module=inline

IV) Comparison of top 10 results for Default and PageRank Algorithms

1. Cannes

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html?module=inline	https://www.nytimes.com/2019/11/07/movies/nyc-this-weekend-film-series.html
https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html	https://www.nytimes.com/news-event/awards-season
https://www.nytimes.com/2018/04/12/movies/cannes-lineup-netflix-boycott.html?module=inline	https://www.nytimes.com/interactive/2019/10/15/t-magazine/nick-cave-artist.html
https://www.nytimes.com/2019/05/24/movies/cannes-almodovar-kechiche.html?module=inline	https://www.nytimes.com/interactive/2019/10/15/t-magazine/nicolas-ghesquiere-louis-vuitton.html
https://www.nytimes.com/2019/05/16/movies/cannes-film-festival-jim-jarmusch.html?module=inline	https://www.nytimes.com/interactive/2019/10/15/t-magazine/rachel-weisz-acting-movies.html
https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html?module=inline	https://www.nytimes.com/2019/10/30/movies/bong-joon-ho-parasite.html
https://www.nytimes.com/2019/05/27/movies/cannes-oscars.html	https://www.nytimes.com/2019/11/01/movies/taika-waititi.html

https://www.nytimes.com/2019/05/20/arts/cannes-film-festival-black-director.html?module=inline	https://www.nytimes.com/interactive/2019/10/15/t-magazine/shigeru-ban.html
https://www.nytimes.com/2019/05/22/movies/quentin-tarantino-margot-robbie.html?module=inline	https://www.nytimes.com/interactive/2019/10/15/t-magazine/the-greats.html
https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html?module=inline	https://www.nytimes.com/interactive/2017/obituaries/notable-deaths/jack-doroshow

2. Congress

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/interactive/2019/us/politics/trump-impeachment-congress-list.html	https://www.nytimes.com/interactive/2019/admin/100000004045625.embedded.html
https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html	https://www.nytimes.com/
https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html?module=inline	https://www.nytimes.com/section/magazine
https://www.nytimes.com/2019/07/25/us/politics/mueller-impeachment.html	https://www.nytimes.com/section/t-magazine
https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html	https://www.nytimes.com/section/todayspaper
https://www.nytimes.com/2018/11/21/obituaries/james-billington-dead.html	https://www.nytimes.com/section/jobs
https://www.nytimes.com/2019/07/26/us/politics/donald-trump-impeachment.html?rref=collection%2Fbyline%2Fnicholas-fandos&action=click&contentCollection=undefined&region=stream&module=inline&version=latest&contentPlacement=9&pgtype=collection	https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html
https://www.nytimes.com/2018/07/30/obituaries/ron-dellums-forceful-liberal-in-congress-for-27-years-dies-at-82.html	https://www.nytimes.com/video/us/politics/100000005524412/paul-manafort-fraud-laundering-scandal.html
https://www.nytimes.com/2019/08/07/us/politics/foreign-aid-freeze-congress.html?module=inline	https://www.nytimes.com/video/us/politics/100000006639814/presidential-candidates-campaigns.html?playlistId=100000006402034
https://www.nytimes.com/2017/12/21/world/americas/peru-president-impeachment.html?module=inline	https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html?playlistId=100000006207751

3. Democrats

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2018/09/04/us/politics/ayanna-pressley-massachusetts.html?module=inline	https://www.nytimes.com/interactive/2019/admin/100000006811573.embedded.html
https://www.nytimes.com/2018/11/07/us/politics/election-history-firsts-blackburn-pressley.html?module=inline	https://www.nytimes.com/interactive/2019/admin/100000006811054.embedded.html
https://www.nytimes.com/2019/10/16/us/politics/democrats-impeachment.html?module=inline	https://www.nytimes.com/interactive/2019/admin/100000006810726.embedded.html
https://www.nytimes.com/2018/01/23/us/politics/impeach-trump-democrats-tom-steyer.html	https://www.nytimes.com/
https://www.nytimes.com/2019/09/29/us/politics/pelosi-trump-impeachment.html	https://www.nytimes.com/video
https://www.nytimes.com/2019/09/24/us/politics/trump-democrats.html	https://www.nytimes.com/section/magazine
https://www.nytimes.com/2019/10/28/us/politics/trump-impeachment-subpoena-house.html	https://www.nytimes.com/section/politics
https://www.nytimes.com/2019/11/05/opinion/trump-democrats-2020.html	https://www.nytimes.com/section/us
https://www.nytimes.com/2019/11/01/us/politics/house-republican-box-prank.html	https://www.nytimes.com/section/opinion
https://www.nytimes.com/2019/07/20/us/politics/trump-democrats-socialists.html?module=inline	https://www.nytimes.com/section/todayspaper

4. Patriot Movement

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2019/08/14/opinion/joe-walsh-trump-primary.html	https://www.nytimes.com/interactive/2019/admin/100000004045625.embedded.html
https://www.nytimes.com/1976/08/27/archives/lewis-michaux-92-dies-ran-bookstore-in-harlem.html?module=inline	https://www.nytimes.com/
https://www.nytimes.com/2019/05/15/us/politics/us-iraq-embassy-evacuation.html?module=inline	https://www.nytimes.com/video
https://www.nytimes.com/2019/01/07/obituaries/moshe-arens-dead.html	https://www.nytimes.com/video/opinion
https://www.nytimes.com/2018/12/29/obituaries/shehu-shagari-dead.html	https://www.nytimes.com/section/t-magazine
https://www.nytimes.com/2019/10/27/obituaries/john-conyers-jr-dead.html	https://www.nytimes.com/section/opinion

https://www.nytimes.com/2019/10/22/sports/football/jets-patriots.html	https://www.nytimes.com/section/technology
https://www.nytimes.com/2019/10/30/opinion/alexander-vindman-trump-ukraine.html	https://www.nytimes.com/section/todayspaper
https://www.nytimes.com/2019/09/23/opinion/democrats-republicans.html	https://www.nytimes.com/section/lens
https://www.nytimes.com/by/jaclyn-peiser	https://www.nytimes.com/video/opinion/100000006525694/stonewall-the-making-of-a-monument.html?playlistId=100000004867895

5. Republicans

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html	https://www.nytimes.com/interactive/2019/admin/100000004045625.embedded.html
https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html	https://www.nytimes.com/
https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html?module=inline	https://www.nytimes.com/section/politics
https://www.nytimes.com/2019/10/04/us/politics/republicans-trump-defense.html	https://www.nytimes.com/section/us
https://www.nytimes.com/by/catie-edmondson	https://www.nytimes.com/section/todayspaper
https://www.nytimes.com/interactive/2019/08/14/magazine/republicans-racism-african-americans.html	https://www.nytimes.com/section/upshot
https://www.nytimes.com/2019/10/31/us/politics/house-impeachment-partisan-vote.html	https://www.nytimes.com/news-event/2020-election
https://www.nytimes.com/2019/10/08/us/politics/republicans-trump-syria.html	https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html
https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html?module=inline	https://www.nytimes.com/video/us/elections/100000006810114/kentucky-election-bevin-beshear.html?playlistId=100000001444053
https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html	https://www.nytimes.com/video/us/politics/100000006812485/hillary-clinton-medicare.html?playlistId=100000002797598

6. Senate

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2019/10/01/us/politics/senate-trump-impeachment.html	https://www.nytimes.com/section/us
https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html	https://www.nytimes.com/section/todayspaper
https://www.nytimes.com/2019/10/26/us/trump-senate-presidency.html	https://www.nytimes.com/section/climate
https://www.nytimes.com/2019/08/03/us/politics/senate-votes-mcconnell.html?module=inline	https://www.nytimes.com/news-event/2020-election
https://www.nytimes.com/2019/08/22/us/politics/john-hickenlooper-senate-2020.html	https://www.nytimes.com/video/opinion/100000006064751/coal-miner-to-trump-coal-mining-isnt-coming-back.html
https://www.nytimes.com/2019/08/31/us/politics/senate-race-2020.html	https://www.nytimes.com/times-journeys/tour-type/
https://www.nytimes.com/by/carl-hulse	https://www.nytimes.com/times-journeys/tour-type/small-group/
https://www.nytimes.com/2019/11/08/us/impeachment-senate-democrats.html	https://www.nytimes.com/times-journeys/trip-type/politics-and-perspective/
https://www.nytimes.com/2019/10/18/us/politics/mcconnell-impeachment-senate.html	https://www.nytimes.com/video/us/politics/100000006639814/presidential-candidates-campaigns.html?playlistId=100000006402034
https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html?module=inline	https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html?playlistId=100000006207751

7. Olympics 2020

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2019/10/01/travel/what-you-need-to-know-the-tokyo-2020-olympics.html	https://www.nytimes.com/interactive/2019/admin/100000006432889.embedded.html
https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html?module=inline	https://www.nytimes.com/interactive/2019/admin/100000006339727.embedded.html
https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html?rref=collection%2Fcolumn%2Flearning-student-opinion&module=inline	https://www.nytimes.com/interactive/2019/admin/100000006511370.embedded.html
https://www.nytimes.com/2017/12/05/sports/olympics/ioc-russia-winter-olympics.html?module=inline	https://www.nytimes.com/interactive/2019/admin/100000006511368.embedded.html
https://www.nytimes.com/2018/04/23/sports/breakdancing-olympics.html?module=inline	https://www.nytimes.com/

https://www.nytimes.com/2016/08/02/sports/olympics/rare-show-of-discord-between-ioc-and-world-anti-doping-agency-over-russian-scandal.html?module=inline	https://www.nytimes.com/video
https://www.nytimes.com/2016/12/08/sports/great-moments-in-cheerleading-could-the-olympics-be-next.html?_r=0&module=inline	https://www.nytimes.com/section/movies
https://www.nytimes.com/2017/12/11/sports/olympics/fastest-color.html?module=inline	https://www.nytimes.com/section/arts/television
https://www.nytimes.com/2019/09/23/sports/olympics/russia-doping-wada.html?module=inline	https://www.nytimes.com/video/opinion
https://www.nytimes.com/2019/11/01/sports/olympics/enriqueta-basilio-71-dies-first-woman-to-light-olympic-flame.html	https://www.nytimes.com/es/

8. Stock

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/interactive/2017/10/04/us/bump-stock-las-vegas-gun.html?module=inline	https://www.nytimes.com/section/world
https://www.nytimes.com/2018/12/16/technology/tech-workers-company-stock-shareholder-activism.html?module=inline	https://www.nytimes.com/section/business
https://www.nytimes.com/2019/05/02/business/tesla-stock-fundraising.html?module=inline	https://www.nytimes.com/section/todayspaper
https://www.nytimes.com/2019/05/10/technology/uber-stock-price-ipo.html?module=inline	https://www.nytimes.com/section/automobiles
https://www.nytimes.com/2019/11/04/business/under-armour-stock-investigation.html	https://www.nytimes.com/section/upshot
https://www.nytimes.com/2019/10/24/style/rebag-clair-handbag-stock-market.html	https://www.nytimes.com/video/opinion/100000006616246/the-king-of-fish-and-chips.html?playlistId=100000004867895
https://www.nytimes.com/2019/10/28/business/stock-market-record.html?module=inline	https://www.nytimes.com/section/fashion
https://www.nytimes.com/2019/05/30/technology/uber-stock-earnings.html?module=inline	https://www.nytimes.com/times-journeys/travel/egypt/
https://www.nytimes.com/2018/11/12/obituaries/david-pearson-nascar-dead.html	https://www.nytimes.com/2019/11/09/opinion/sunday/stripping-sex-work.html

| https://www.nytimes.com/2019/09/20/tech nology/airbnb-employees-ipo-payouts.html?module=inline | https://www.nytimes.com/video/movies/1000000 06675663/brittany-runs-a-marathon-scene.html?playlistId=100000002420711 |

9. Virus

Default Solr Algorithm	External PageRank Algorithm
https://www.nytimes.com/2019/11/09/scie nce/seals-distemper.html	https://www.nytimes.com/es/
https://www.nytimes.com/2019/10/28/us/p olitics/kay-hagan-dead.html	https://www.nytimes.com/section/us
https://www.nytimes.com/2019/10/25/heal th/predict-usaid-viruses.html	https://www.nytimes.com/section/world
https://www.nytimes.com/2019/02/05/us/p olitics/trump-hiv-aids-plan.html?module=inline	https://www.nytimes.com/section/science
https://www.nytimes.com/video/opinion/1 00000006210828/russia-disinformation-fake-news.html	https://www.nytimes.com/section/health
https://www.nytimes.com/video/opinion/1 00000006210828/russia-disinformation-fake-news.html?playlistId=100000006207751	https://www.nytimes.com/section/climate
https://www.nytimes.com/by/donald-g-mcneil-jr	https://www.nytimes.com/video/opinion/100000 006188102/what-is-pizzagate.html?playlistId=100000006207751
https://www.nytimes.com/2019/09/27/well /live/when-is-the-best-time-to-get-your-flu-shot.html	https://www.nytimes.com/video/opinion/100000 006188105/countering-disinformation-active-measures.html?playlistId=100000006207751
https://www.nytimes.com/2019/01/23/nee diest-cases/daughter-cmv.html	https://www.nytimes.com/video/opinion/100000 006210828/russia-disinformation-fake-news.html?playlistId=100000006207751
https://www.nytimes.com/2019/02/04/us/p olitics/trump-hiv-aids-plan.html?module=inline	https://www.nytimes.com/video/opinion/100000 006188105/countering-disinformation-active-measures.html

V) Overlaps

Query No.	Query	No. of Overlaps
1.	Cannes	0
2.	Congress	0
3.	Democrats	0
4.	Patriot Movement	0
5.	Republicans	0
6.	Senate	0
7.	Olympics 2020	0
8.	Stock	0
9.	Virus	1

Number of overlaps between results