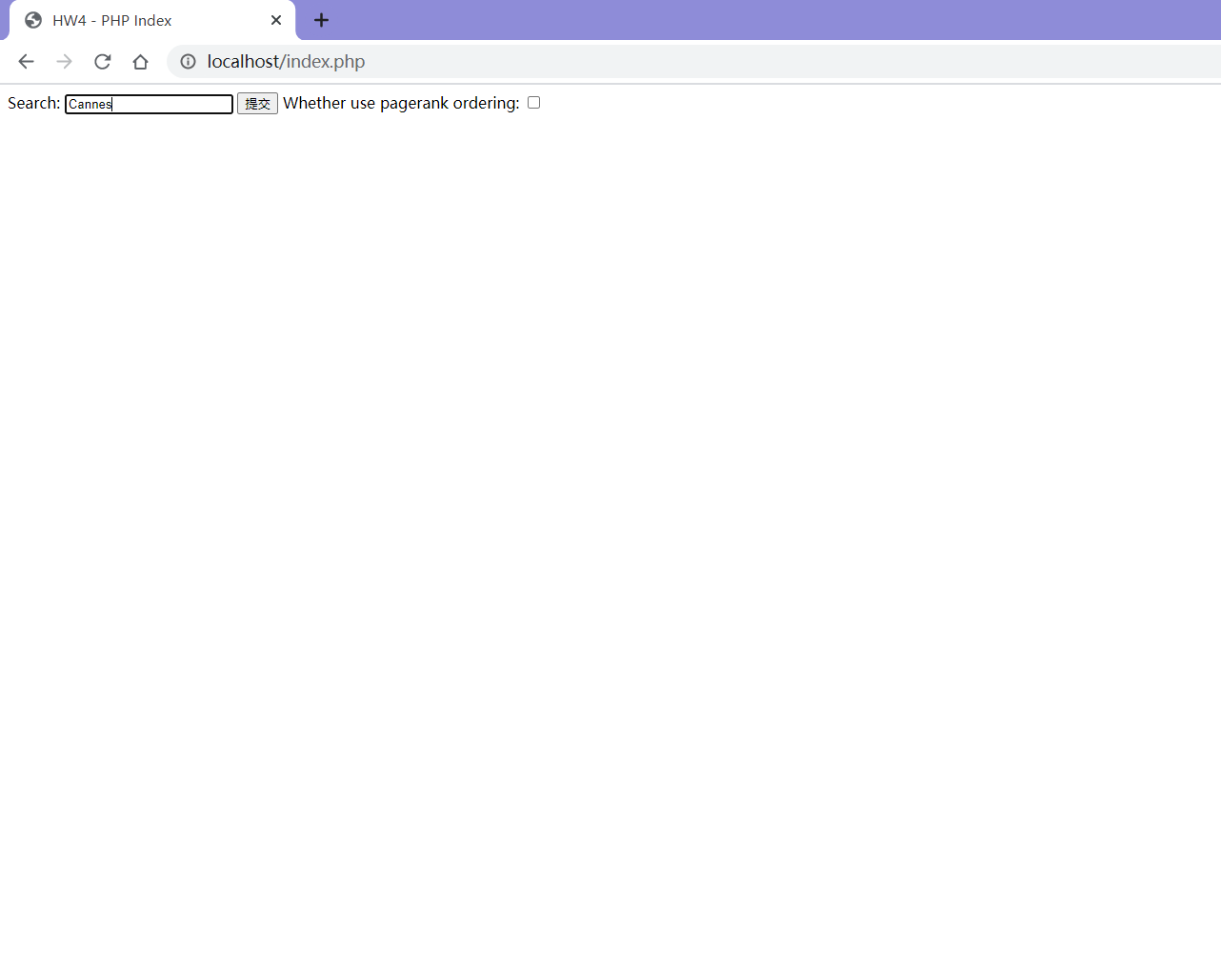
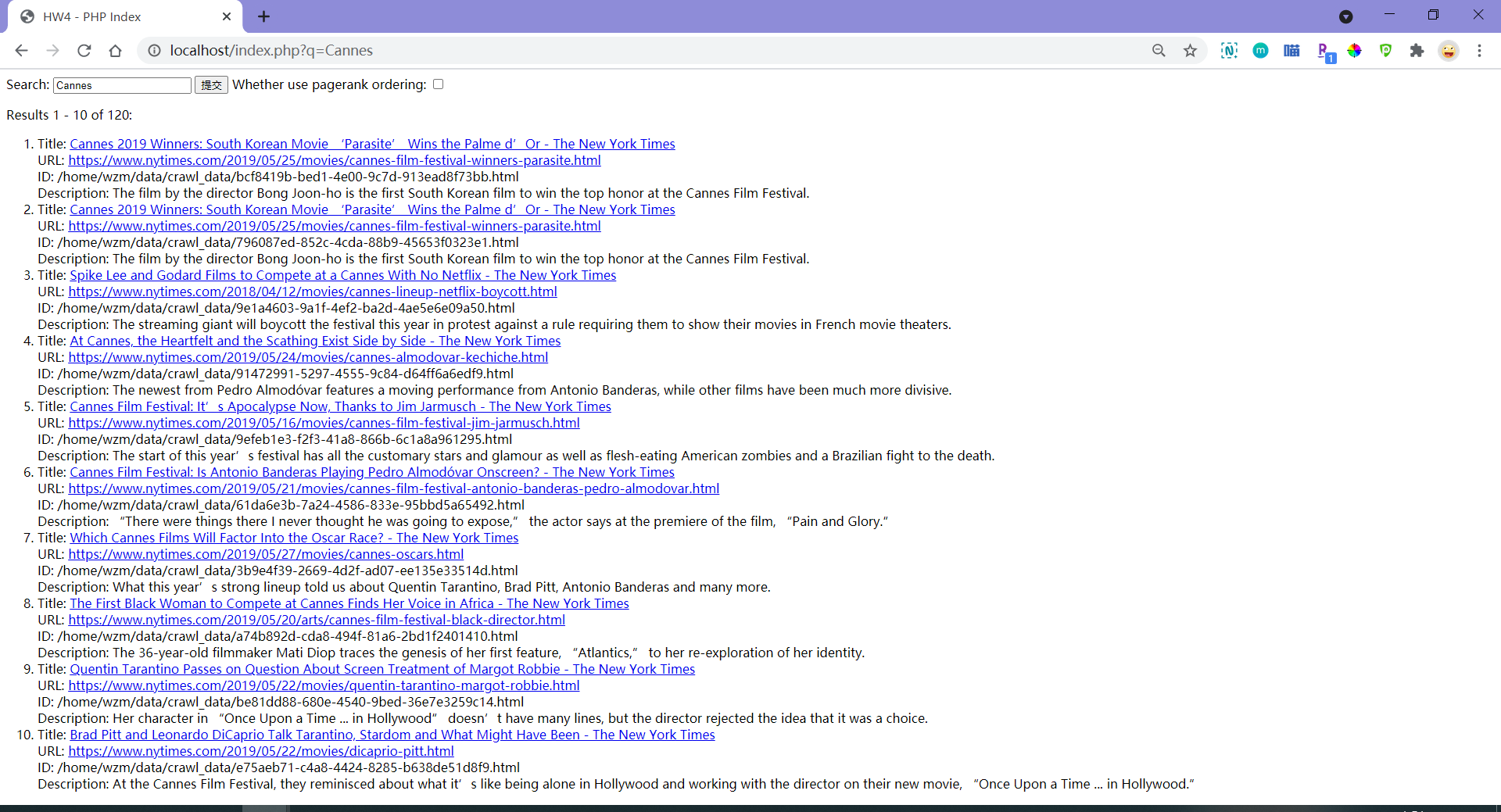
1. **Steps to complete the assignment**
2. Using Ubuntu 16.04 (in VirtualBox) to setup Solr standalone, and maps port between host machine and virtual machine for host to connect Solr.
3. Change the config of my core, provide html files to Solr and index pages.
4. Extract outgoing links and build DiGraph (all by Python3): using BeautifulSoup4 to extract outgoing links of html files; while extracting, using file\_to\_url and url\_to\_file (from URLtoHTML\_nytimes\_news.csv) to filter links, and adding edges to the DiGraph using the NetworkX library. So get the directed graph of files.
5. Calculate page rank: by the DiGraph created before, using pagerank function to calculate PageRank. The parameters include: alpha=0.85, personalization=None, max\_iter=30, tol=1.0e-06, nstart=None, weight='weight', dangling=None.
6. Configure the standalone core to support the external PageRank file and reload it.
7. Build an Apache & PHP server to show the results of the queries.
8. Using Python to compute the similarities of query results.
9. **Explain why some pages have higher PageRank values than others**

Some pages have higher page rank because they have more pages linking into them, so there PageRanks have a lot of PR(A)/C(A) sum up. For example, pages that are front pages of the site or category, or pages that are on recommendation lists.

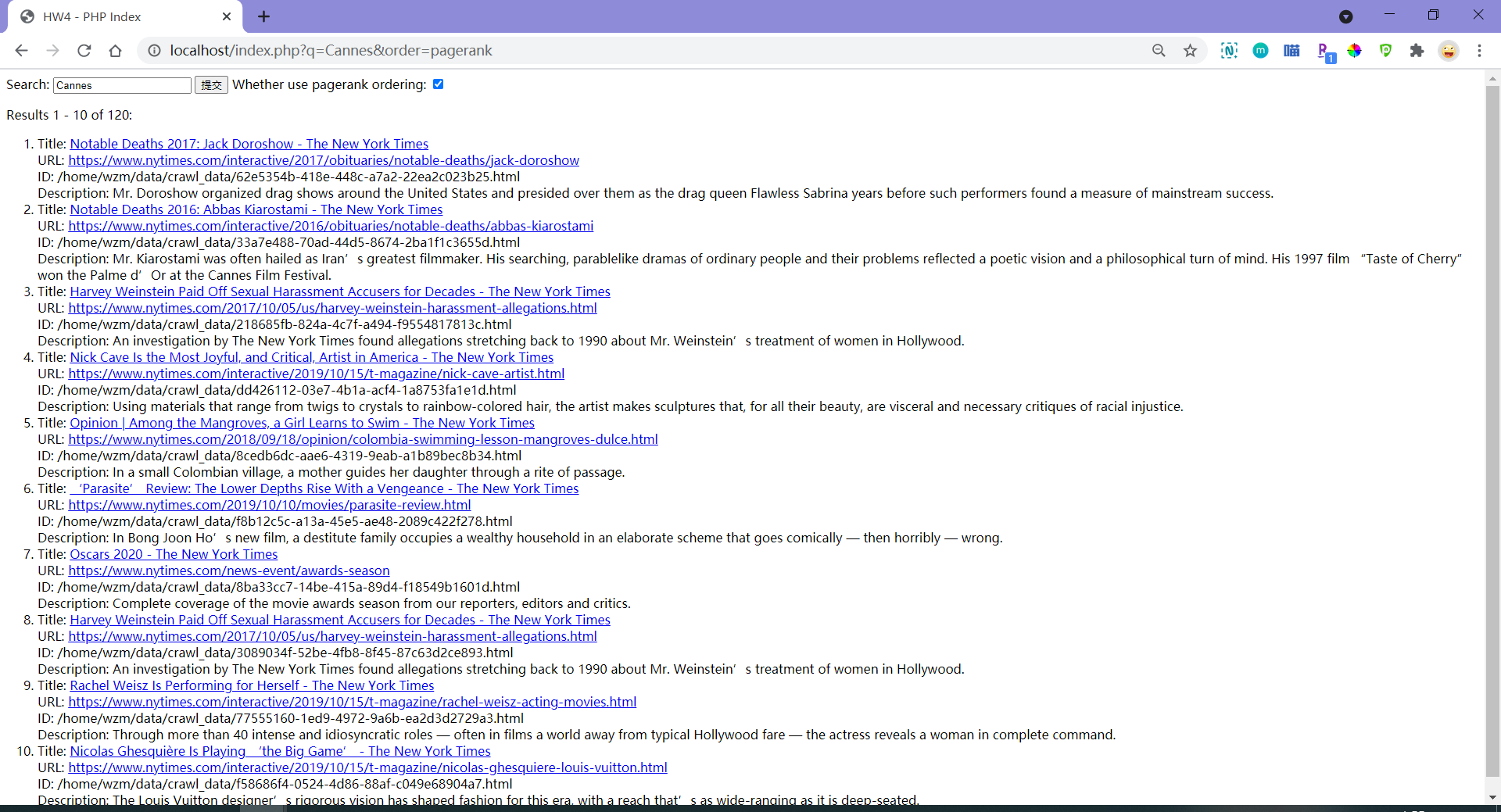
1. **Screenshots**



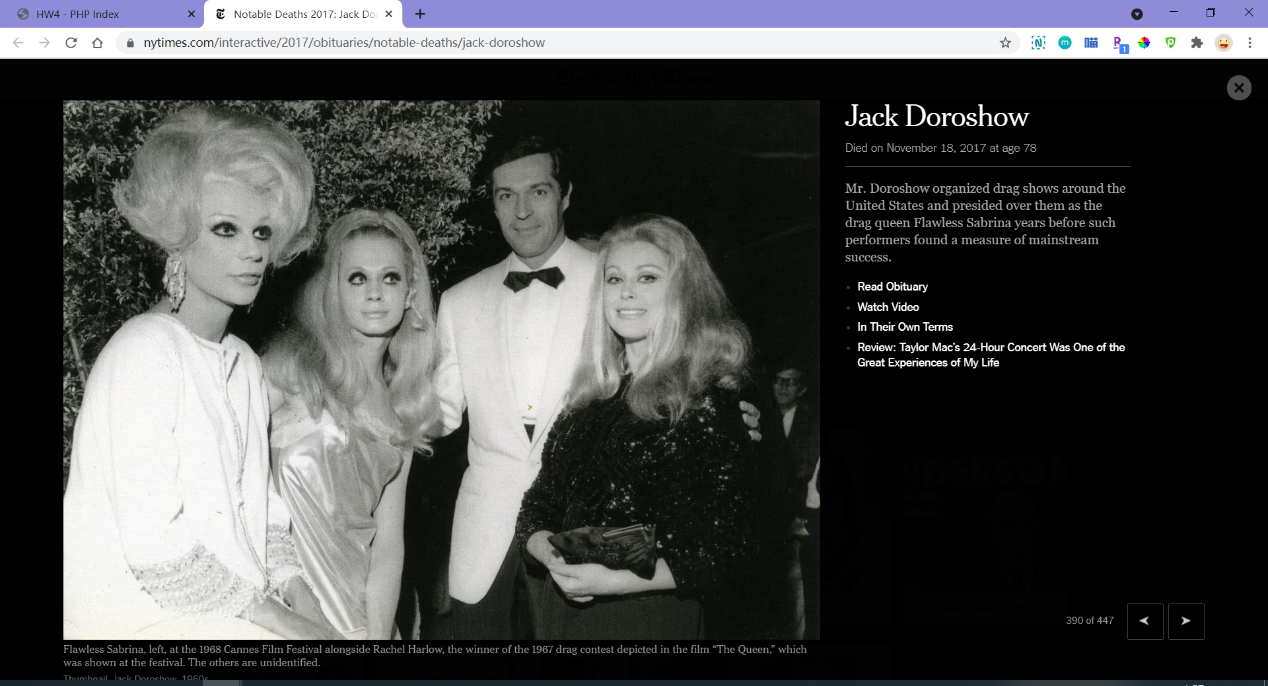
Initial query page (PHP)



Results for Lucene (Default)



Results for PageRank (Default)



When click on a link

1. **A table containing list of URLs of the top ten results**

|  |  |  |
| --- | --- | --- |
| **Cannes** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html | https://www.nytimes.com/interactive/2017/obituaries/notable-deaths/jack-doroshow |
| 1 | https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html | https://www.nytimes.com/interactive/2016/obituaries/notable-deaths/abbas-kiarostami |
| 2 | https://www.nytimes.com/2018/04/12/movies/cannes-lineup-netflix-boycott.html | https://www.nytimes.com/2017/10/05/us/harvey-weinstein-harassment-allegations.html |
| 3 | https://www.nytimes.com/2019/05/24/movies/cannes-almodovar-kechiche.html | https://www.nytimes.com/interactive/2019/10/15/t-magazine/nick-cave-artist.html |
| 4 | https://www.nytimes.com/2019/05/16/movies/cannes-film-festival-jim-jarmusch.html | https://www.nytimes.com/2018/09/18/opinion/colombia-swimming-lesson-mangroves-dulce.html |
| 5 | https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html | https://www.nytimes.com/2019/10/10/movies/parasite-review.html |
| 6 | https://www.nytimes.com/2019/05/27/movies/cannes-oscars.html | https://www.nytimes.com/news-event/awards-season |
| 7 | https://www.nytimes.com/2019/05/20/arts/cannes-film-festival-black-director.html | https://www.nytimes.com/2017/10/05/us/harvey-weinstein-harassment-allegations.html |
| 8 | https://www.nytimes.com/2019/05/22/movies/quentin-tarantino-margot-robbie.html | https://www.nytimes.com/interactive/2019/10/15/t-magazine/rachel-weisz-acting-movies.html |
| 9 | https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html | https://www.nytimes.com/interactive/2019/10/15/t-magazine/nicolas-ghesquiere-louis-vuitton.html |
| **Congress** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/interactive/2019/us/politics/trump-impeachment-congress-list.html | https://www.nytimes.com |
| 1 | https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html | https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html |
| 2 | https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html | https://www.nytimes.com/video/us/politics/100000005524412/paul-manafort-fraud-laundering-scandal.html |
| 3 | https://www.nytimes.com/2019/07/25/us/politics/mueller-impeachment.html | https://www.nytimes.com/section/magazine |
| 4 | https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html | https://www.nytimes.com/section/t-magazine |
| 5 | https://www.nytimes.com/2018/11/21/obituaries/james-billington-dead.html | https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times |
| 6 | https://www.nytimes.com/2019/07/26/us/politics/donald-trump-impeachment.html | https://www.nytimes.com/section/jobs |
| 7 | https://www.nytimes.com/2019/08/07/us/politics/foreign-aid-freeze-congress.html | https://www.nytimes.com/2017/10/30/us/politics/paul-manafort.html |
| 8 | 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/2018/08/22/us/politics/offense-impeachable-constitution.html |
| 9 | https://www.nytimes.com/2017/12/21/world/americas/peru-president-impeachment.html | https://www.nytimes.com/interactive/2017/04/19/crosswords/CROSSWORDS-sports-names.html |
| **Democrats** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2018/09/04/us/politics/ayanna-pressley-massachusetts.html | https://www.nytimes.com/video |
| 1 | https://www.nytimes.com/2018/11/07/us/politics/election-history-firsts-blackburn-pressley.html | https://www.nytimes.com |
| 2 | https://www.nytimes.com/2019/10/16/us/politics/democrats-impeachment.html | https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html |
| 3 | https://www.nytimes.com/2018/01/23/us/politics/impeach-trump-democrats-tom-steyer.html | https://www.nytimes.com/video/opinion/100000006064751/coal-miner-to-trump-coal-mining-isnt-coming-back.html |
| 4 | https://www.nytimes.com/2019/09/29/us/politics/pelosi-trump-impeachment.html | https://www.nytimes.com/section/magazine |
| 5 | https://www.nytimes.com/2019/09/24/us/politics/trump-democrats.html | https://www.nytimes.com/section/us |
| 6 | https://www.nytimes.com/2019/10/28/us/politics/trump-impeachment-subpoena-house.html | https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times |
| 7 | https://www.nytimes.com/2019/11/05/opinion/trump-democrats-2020.html | https://www.nytimes.com/section/opinion/columnists |
| 8 | https://www.nytimes.com/2019/11/01/us/politics/house-republican-box-prank.html | https://www.nytimes.com/section/opinion |
| 9 | https://www.nytimes.com/2019/07/20/us/politics/trump-democrats-socialists.html | https://www.nytimes.com/section/politics |
| **Patriot Movement** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2019/08/14/opinion/joe-walsh-trump-primary.html | https://www.nytimes.com/video |
| 1 | https://www.nytimes.com/1976/08/27/archives/lewis-michaux-92-dies-ran-bookstore-in-harlem.html | https://www.nytimes.com |
| 2 | https://www.nytimes.com/2019/05/15/us/politics/us-iraq-embassy-evacuation.html | https://www.nytimes.com/section/t-magazine |
| 3 | https://www.nytimes.com/2019/01/07/obituaries/moshe-arens-dead.html | https://www.nytimes.com/section/technology |
| 4 | https://www.nytimes.com/2019/10/27/obituaries/john-conyers-jr-dead.html | https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times |
| 5 | https://www.nytimes.com/2018/12/29/obituaries/shehu-shagari-dead.html | https://www.nytimes.com/section/opinion |
| 6 | https://www.nytimes.com/2019/10/22/sports/football/jets-patriots.html | https://www.nytimes.com/section/lens |
| 7 | https://www.nytimes.com/2019/10/30/opinion/alexander-vindman-trump-ukraine.html | https://www.nytimes.com/video/opinion |
| 8 | https://www.nytimes.com/2019/09/23/opinion/democrats-republicans.html | https://www.nytimes.com/2019/11/01/arts/television/new-to-stream-netflix.html |
| 9 | https://www.nytimes.com/by/jaclyn-peiser | https://www.nytimes.com/2018/08/07/opinion/deafness-cochlear-implants.html |
| **Republicans** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html | https://www.nytimes.com |
| 1 | https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html | https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html |
| 2 | https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html | https://www.nytimes.com/section/us |
| 3 | https://www.nytimes.com/2019/10/04/us/politics/republicans-trump-defense.html | https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times |
| 4 | https://www.nytimes.com/by/catie-edmondson | https://www.nytimes.com/section/politics |
| 5 | https://www.nytimes.com/interactive/2019/08/14/magazine/republicans-racism-african-americans.html | https://www.nytimes.com/section/upshot |
| 6 | https://www.nytimes.com/2019/10/08/us/politics/republicans-trump-syria.html | https://www.nytimes.com/news-event/2020-election |
| 7 | https://www.nytimes.com/2019/10/31/us/politics/house-impeachment-partisan-vote.html | https://www.nytimes.com/2017/10/30/us/politics/paul-manafort.html |
| 8 | https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html | https://www.nytimes.com/2018/08/22/us/politics/offense-impeachable-constitution.html |
| 9 | https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html | https://www.nytimes.com/interactive/2019/us/politics/2020-presidential-candidates.html |
| **Senate** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2019/10/01/us/politics/senate-trump-impeachment.html | https://www.nytimes.com/video/opinion/100000006064751/coal-miner-to-trump-coal-mining-isnt-coming-back.html |
| 1 | https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html | https://www.nytimes.com/section/us |
| 2 | https://www.nytimes.com/2019/10/26/us/trump-senate-presidency.html | https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times |
| 3 | https://www.nytimes.com/2019/08/03/us/politics/senate-votes-mcconnell.html | https://www.nytimes.com/section/climate |
| 4 | https://www.nytimes.com/2019/08/22/us/politics/john-hickenlooper-senate-2020.html | https://www.nytimes.com/news-event/2020-election |
| 5 | https://www.nytimes.com/2019/08/31/us/politics/senate-race-2020.html | https://www.nytimes.com/times-journeys/tour-type/small-group/ |
| 6 | https://www.nytimes.com/by/carl-hulse | https://www.nytimes.com/times-journeys/trip-type/politics-and-perspective/ |
| 7 | https://www.nytimes.com/2019/11/08/us/impeachment-senate-democrats.html | https://www.nytimes.com/2018/08/22/us/politics/offense-impeachable-constitution.html |
| 8 | https://www.nytimes.com/2019/10/18/us/politics/mcconnell-impeachment-senate.html | https://timesjourneys.nytimes.com/tour-type/ |
| 9 | https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html | https://www.nytimes.com/interactive/2019/us/politics/2020-presidential-candidates.html |
| **Olympics 2020** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2019/10/01/travel/what-you-need-to-know-the-tokyo-2020-olympics.html | https://www.nytimes.com/video |
| 1 | https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html | https://www.nytimes.com |
| 2 | https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html | https://www.nytimes.com/video/world/middleeast/100000001253159/2011-arab-spring.html |
| 3 | https://www.nytimes.com/2017/12/05/sports/olympics/ioc-russia-winter-olympics.html | https://www.nytimes.com/video/opinion/100000006033259/between-sound-and-silence.html |
| 4 | https://www.nytimes.com/2018/04/23/sports/breakdancing-olympics.html | https://www.nytimes.com/video/opinion/100000005675688/my-dead-dads-porno-tapes.html |
| 5 | https://www.nytimes.com/2016/08/02/sports/olympics/rare-show-of-discord-between-ioc-and-world-anti-doping-agency-over-russian-scandal.html | https://www.nytimes.com/video/opinion/100000006048561/calving-season.html |
| 6 | https://www.nytimes.com/2016/12/08/sports/great-moments-in-cheerleading-could-the-olympics-be-next.html | https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html |
| 7 | https://www.nytimes.com/2019/09/23/sports/olympics/russia-doping-wada.html | https://www.nytimes.com/video/us/politics/100000005524412/paul-manafort-fraud-laundering-scandal.html |
| 8 | https://www.nytimes.com/2017/12/11/sports/olympics/fastest-color.html | https://www.nytimes.com/video/opinion/100000006055559/sanctuary-city-hotline.html |
| 9 | https://www.nytimes.com/2019/11/01/sports/olympics/enriqueta-basilio-71-dies-first-woman-to-light-olympic-flame.html | https://www.nytimes.com/video/opinion/100000006064751/coal-miner-to-trump-coal-mining-isnt-coming-back.html |
| **Stock** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/interactive/2017/10/04/us/bump-stock-las-vegas-gun.html | https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times |
| 1 | https://www.nytimes.com/2018/12/16/technology/tech-workers-company-stock-shareholder-activism.html | https://www.nytimes.com/section/world |
| 2 | https://www.nytimes.com/2019/05/02/business/tesla-stock-fundraising.html | https://www.nytimes.com/section/business |
| 3 | https://www.nytimes.com/2019/05/10/technology/uber-stock-price-ipo.html | https://www.nytimes.com/section/automobiles |
| 4 | https://www.nytimes.com/2019/11/04/business/under-armour-stock-investigation.html | https://www.nytimes.com/section/upshot |
| 5 | https://www.nytimes.com/2019/10/24/style/rebag-clair-handbag-stock-market.html | https://www.nytimes.com/times-journeys/travel/egypt/ |
| 6 | https://www.nytimes.com/2019/10/28/business/stock-market-record.html | https://www.nytimes.com/section/fashion |
| 7 | https://www.nytimes.com/2019/05/30/technology/uber-stock-earnings.html | https://www.nytimes.com/times-journeys/travel-experience/greenwich-village-old-new-and-delicious/ |
| 8 | https://www.nytimes.com/2018/11/12/obituaries/david-pearson-nascar-dead.html | https://www.nytimes.com/interactive/2019/04/10/t-magazine/terrence-mcnally-interview.html |
| 9 | https://www.nytimes.com/2019/09/20/technology/airbnb-employees-ipo-payouts.html | https://www.nytimes.com/section/world/middleeast |
| **Virus** | **default** | **pagerank** |
| 0 | https://www.nytimes.com/2019/11/09/science/seals-distemper.html | https://www.nytimes.com/es/ |
| 1 | https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html | https://www.nytimes.com/section/us |
| 2 | https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html | https://www.nytimes.com/section/science |
| 3 | https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html | https://www.nytimes.com/section/world |
| 4 | https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html | https://www.nytimes.com/section/health |
| 5 | https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html | https://www.nytimes.com/section/climate |
| 6 | https://www.nytimes.com/by/donald-g-mcneil-jr | https://www.nytimes.com/2019/08/01/business/vas-narasimhan-novartis-corner-office.html |
| 7 | https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html | https://www.nytimes.com/interactive/2016/obituaries/notable-deaths/da-henderson |
| 8 | https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html | https://www.nytimes.com/2018/11/12/opinion/russia-meddling-disinformation-fake-news-elections.html |
| 9 | https://www.nytimes.com/2019/02/04/us/politics/trump-hiv-aids-plan.html | https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html |

1. **Overlap**

My result shows that there is no overlap.

This is because that the default (Lucene) algorithm may give more relevant pages compared to PageRank algorithm. PageRank focus more on ‘vote’, that is, it relates to how much pages are related to

other pages, but may not focus on the relevance to the query.