

1. Steps:

As my computer is MacBook, so there was no need to download Ubuntu. First of all, I downloaded the required documents and packages from Internet, including solr-7.5.0, the html files, mapping documents of NY Post and NetworkX library. After installing the solr-7.5.0, I learnt how to use solr and tried to index the required web pages on the UI provided by solr.

Then I tried to finish the front-end php file by modifying the code sample from the pdf. In this php file, I designed the query box and video buttons for choosing the algorithm of search engine. As for some html files, there were no "og_url", in this case, I tried to read given htmlToUrl.csv and fetched the url according from the document. The top 10 results shown after the query should be listed with the order "title", "url", "id" and "description". And the first two "title" and "url" were clickable and able to redirect to the actual website of the url.

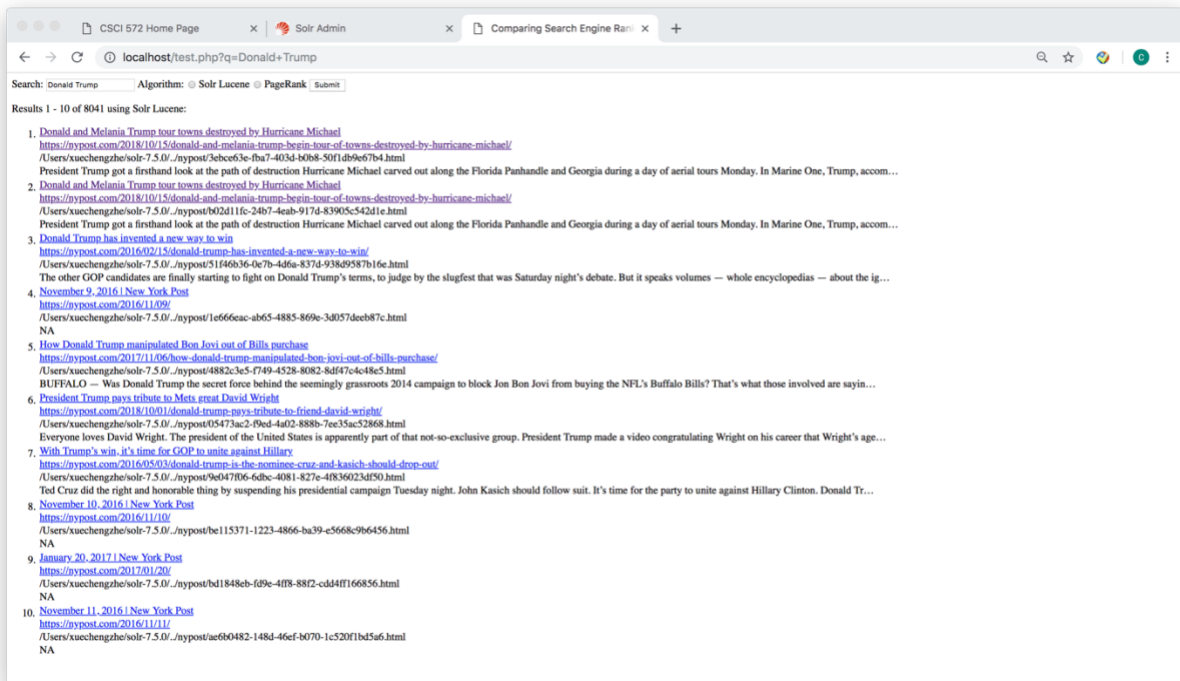
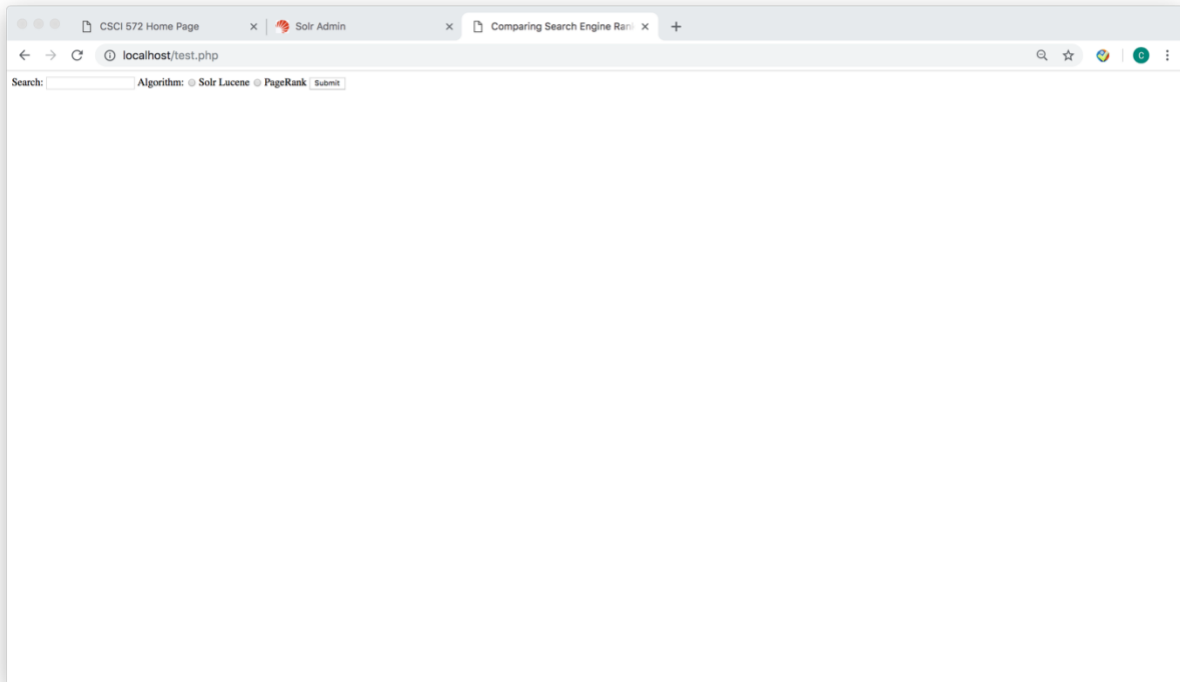
After finishing the front-end implementation, I came to work on PageRank algorithm. Firstly, I wrote a Java code based on the example and constructed a graph with edge list. In this edgeList.txt, each row contains 2 html separated by a space, which means the first html had a outgoing link to the second html. Having this graph, I passed this graph to NetworkX in python, by setting the parameters as "alpha=0.85", "personalization=None", "max_iter=30", "tol=1e-06", "nstart=None", "weight='weight'", "dangling=None", and wrote the resulting dictionary into a new txt file named "external_pageRankFile.txt". This is the file included all the html ("id") and its corresponding PageRank values.

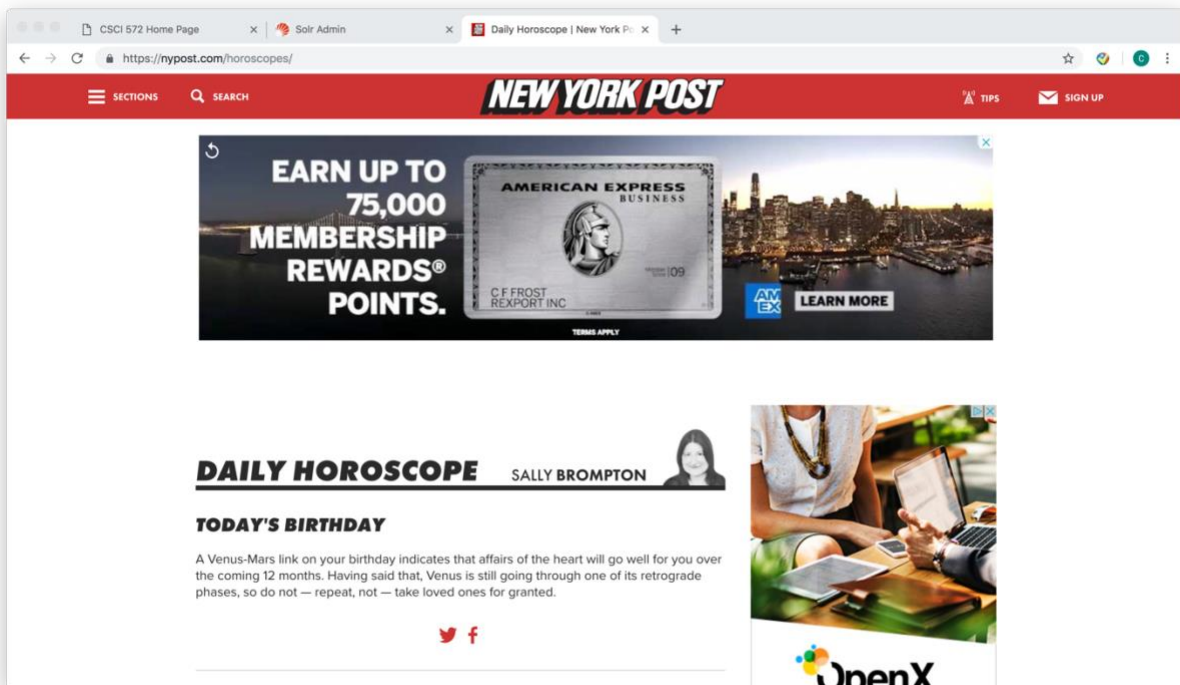
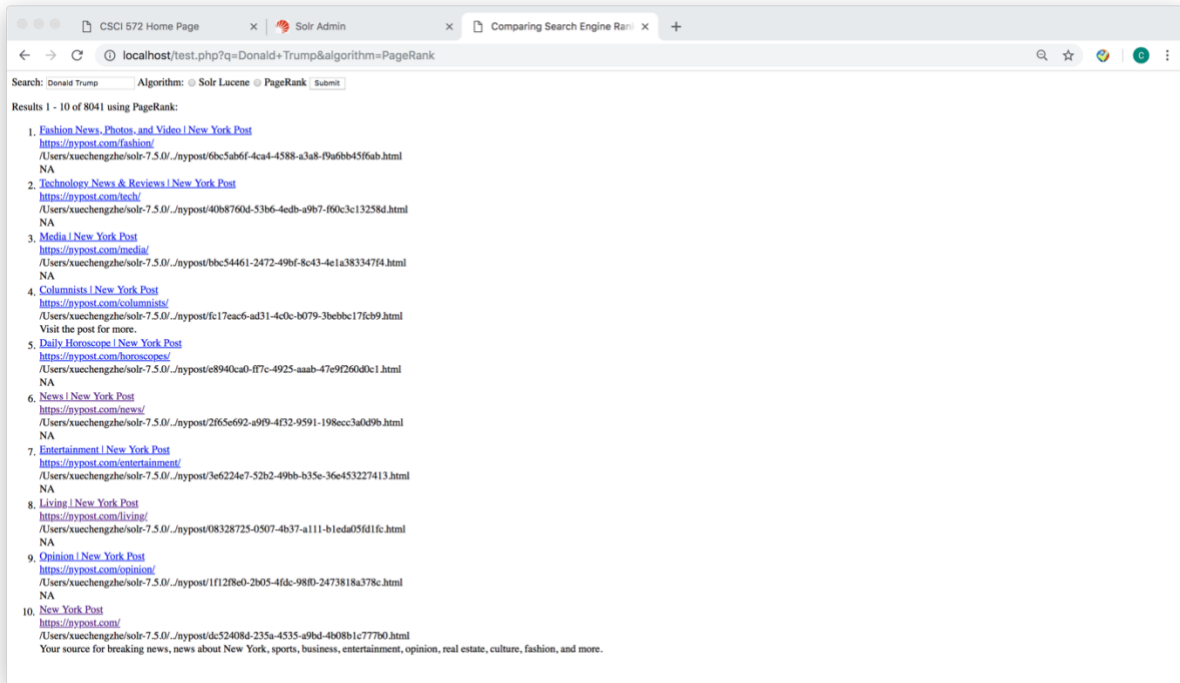
Here comes to the final steps, after adding fields in the managed-schema which refers to this score and defining listeners within the <query> element in the solrconfig.xml file, now we can get the top 10 results based on PageRank on solr UI by typing "pageRankFile desc" in "sort" box. Finally, I added this function into the php front-end source code of searching by \$additionalParameters.

2. Explanation regarding why some pages have higher PageRank values than others:

The pages have higher PageRank values have more indegree links than those with lower PageRank values. For example, for the query "Donald Trump", checking from the edgeList.txt, I find that the html with highest PageRank value (pageRankFile desc) has 41 indegree links, but the html with lowest PageRank value (pageRankFile asc) has 0 indegree links.

3. Document the flow for any single query using screenshots:





4. A table containing list of URLs of the top ten results produced by both ranking methods for each query:

Query	Ranking	Solr Lucene(Default)	PageRank
Donald Trump	1	https://nypost.com/2018/10/15/donald-and-melania-trump-begin-tour-of-towns-destroyed-by-hurricane-michael/	https://nypost.com/fashion/
	2	https://nypost.com/2018/10/15/donald-and-melania-trump-begin-tour-of-towns-destroyed-by-hurricane-michael/	https://nypost.com/tech/
	3	https://nypost.com/2016/02/15/donald-trump-has-invented-a-new-way-to-win/	https://nypost.com/media/
	4	https://nypost.com/2016/11/09/	https://nypost.com/columnists/
	5	https://nypost.com/2017/11/06/how-donald-trump-manipulated-bon-jovi-out-of-bills-purchase/	https://nypost.com/horoscopes/
	6	https://nypost.com/2018/10/01/donald-trump-pays-tribute-to-friend-david-wright/	https://nypost.com/news/
	7	https://nypost.com/2016/05/03/donald-trump-is-the-nominee-cruz-and-kasich-should-drop-out/	https://nypost.com/entertainment/
	8	https://nypost.com/2016/11/10/	https://nypost.com/living/
	9	https://nypost.com/2017/01/20/	https://nypost.com/opinion/
	10	https://nypost.com/2016/11/11/	https://nypost.com/
LA Lakers	1	https://nypost.com/2018/09/25/lebron-james-knows-it-wont-come-easy-with-lakers/	https://nypost.com/sports/
	2	https://nypost.com/2018/09/07/warriors-star-emerges-as-lakers-flashy-plan-b-in-free-agency/	https://nypost.com/2018/10/15/tom-bradys-postgame-words-to-gronk-are-an-nfl-nightmare/
	3	https://nypost.com/2018/10/10/klay-thompson-gushes-over-lebron-as-lakers-pursuit-looms/	https://nypost.com/2018/10/15/nfl-flameout-admits-to-having-sex-with-12-year-old-girl-police/

	4	https://nypost.com/tag/nba/	https://nypost.com/2018/10/15/eli-manning-trade-could-save-him-and-the-giants-espn-qb/
	5	https://nypost.com/tag/los-angeles-lakers/	https://nypost.com/2018/10/14/aaron-hernandez-had-sexual-relationship-with-hs-quarterback/
	6	https://nypost.com/2001/02/14/	https://nypost.com/2018/10/15/mlb-youngsters-bring-the-trash-talk-and-much-needed-fun/
	7	https://nypost.com/2014/05/02/	https://nypost.com/2018/10/15/empty-vows-alderson-back-stab-inside-mets-gm-search-mess/
	8	https://nypost.com/2012/11/12/	https://nypost.com/2018/10/15/angry-about-the-subways-send-cuomo-a-message/
	9	https://nypost.com/2015/06/26/	https://nypost.com/2018/10/15/a-giants-trade-for-ex-patriots-quarterback-would-make-sense/
	10	https://nypost.com/2017/02/21/	https://nypost.com/2018/10/15/how-one-veterans-vietnam-horrors-inspired-this-is-us/
Star Wars	1	https://nypost.com/2015/12/18/	https://nypost.com/real-estate/
	2	https://nypost.com/2015/12/15/	https://nypost.com/fashion/
	3	https://nypost.com/2015/12/17/	https://nypost.com/tech/
	4	https://nypost.com/2015/12/16/	https://nypost.com/metro/
	5	https://nypost.com/2015/12/22/	https://nypost.com/media/
	6	https://nypost.com/2015/12/14/	https://nypost.com/columnists/
	7	https://nypost.com/2015/11/24/	https://nypost.com/sports/
	8	https://nypost.com/2015/04/20/	https://nypost.com/horoscopes/
	9	https://nypost.com/2015/10/20/	https://nypost.com/photos/
	10	https://nypost.com/2015/04/16/	https://nypost.com/news/

Lebron James	1	https://nypost.com/2018/09/25/lebron-james-knows-it-wont-come-easy-with-lakers/	https://nypost.com/metro/
	2	https://nypost.com/2018/04/04/lebron-saban-feud-over-shop-talk-show-in-copyright-dispute/	https://nypost.com/sports/
	3	https://nypost.com/tag/los-angeles-lakers/	https://nypost.com/2018/10/13/100m-cybersecurity-center-slated-to-open-in-soho-next-year/
	4	https://nypost.com/2010/07/09/	https://nypost.com/2018/10/15/judge-tosses-stormy-daniels-defamation-suit-against-trump/
	5	https://nypost.com/2018/07/02/	https://nypost.com/2018/10/15/beach-house-still-standing-after-hurricane-michael-devastates-nearby-homes/
	6	https://nypost.com/2018/10/10/klay-thompson-gushes-over-lebron-as-lakers-pursuit-looms/	https://nypost.com/2018/10/15/ex-nightlife-king-plans-to-open-a-private-club-in-noho/
	7	https://nypost.com/2010/07/08/	https://nypost.com/2018/10/15/nyc-goes-entire-weekend-without-a-shooting-for-first-time-in-14-years/
	8	https://nypost.com/2014/07/11/	https://nypost.com/2018/10/15/new-york-city-restaurant-scene-teeming-with-new-entries/
	9	https://nypost.com/2010/07/01/	https://nypost.com/2018/10/15/knicks-season-preview-what-theyre-really-playing-for/
	10	https://nypost.com/2010/07/07/	https://nypost.com/2018/10/12/little-kids-are-wearing-grown-up-makeup-is-it-ok/

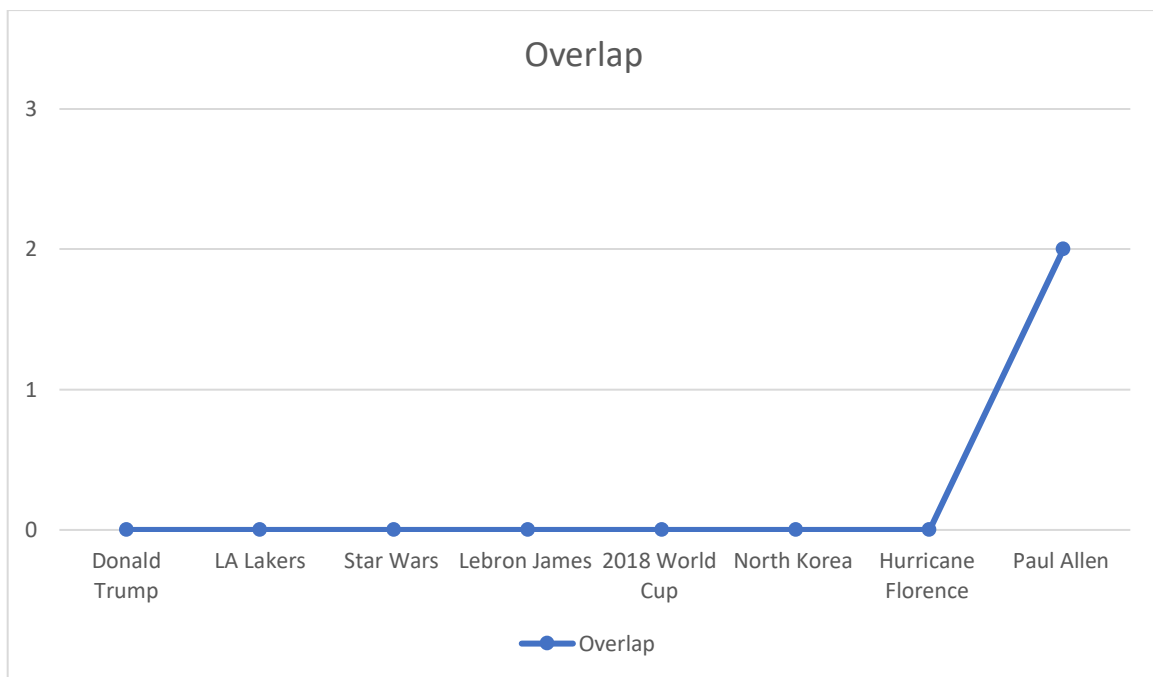
2018 World Cup	1	https://nypost.com/2018/07/12/theres-a-knicks-battle-and-bet-over-the-world-cup-final/	https://nypost.com/mobile-apps/
	2	https://nypost.com/video/this-psycho-pig-just-predicted-the-world-cup/	https://nypost.com/real-estate/
	3	https://nypost.com/2018/06/05/nycfc-cant-wait-to-get-monkey-off-their-back-in-u-s-open-cup/	https://nypost.com/fashion/
	4	https://nypost.com/2018/10/04/stricken-ryder-cup-fan-i-thought-id-be-trampled-by-tiger-woods-fans/	https://nypost.com/video/
	5	https://nypost.com/2018/07/13/son-of-stanley-cup-winning-devil-chasing-nhl-dream/	https://nypost.com/tech/
	6	https://nypost.com/2014/06/12/	https://nypost.com/metro/
	7	https://nypost.com/2014/06/11/	https://nypost.com/business/
	8	https://nypost.com/2018/10/04/capitals-stanley-cup-tribute-was-a-little-bit-petty/	https://nypost.com/contact/
	9	https://nypost.com/2018/10/02/nhl-preview-these-stanley-cup-contenders-are-racing-the-clock/	https://nypost.com/media/
	10	https://nypost.com/2010/06/14/	https://nypost.com/columnists/
North Korea	1	https://nypost.com/2018/08/30/aglimpse-at-life-inside-north-korea/slide-1/	https://nypost.com/real-estate/
	2	https://nypost.com/2018/08/24/state-department-reportedly-blindsided-by-trump-canceling-north-korea-trip/	https://nypost.com/sports/
	3	https://nypost.com/2018/06/07/former-pow-calls-on-trump-to-reclaim-us-ship-captured-by-north-korea/	https://nypost.com/2018/10/15/tom-bradys-postgame-words-to-gronk-are-an-nfl-nightmare/
	4	https://nypost.com/2018/10/07/pompeo-progress-made-with-kim-jong-un-on-north-korea-trip/	https://nypost.com/2018/10/15/nfl-flameout-admits-to-having-sex-with-12-year-old-girl-police/

	5	https://nypost.com/2018/09/10/north-korea-marks-the-hermit-kingdoms-70th-anniversary/slide-1/	https://nypost.com/2018/10/15/eli-manning-trade-could-save-him-and-the-giants-ESPN-QB/
	6	https://nypost.com/2017/09/09/how-democrats-left-america-naked-before-north-koreas-nukes/	https://nypost.com/2018/10/14/aaron-hernandez-had-sexual-relationship-with-his-quarterback/
	7	https://nypost.com/2017/09/03/	https://nypost.com/2018/10/15/mlb-youngsters-bring-the-trash-talk-and-much-needed-fun/
	8	https://nypost.com/2018/10/02/trumps-love-for-kim-is-undercutting-his-own-teams-necessary-toughness/	https://nypost.com/2018/10/15/empty-vows-alderson-backstab-inside-mets-gm-search-mess/
	9	https://nypost.com/2017/09/04/	https://nypost.com/2018/10/15/how-eddie-lampert-led-sears-into-bankruptcy-disaster/
	10	https://nypost.com/2018/10/04/north-korean-hackers-linked-to-global-banking-heist/	https://nypost.com/2018/10/15/the-south-bronx-is-in-the-midst-of-a-development-boom/
Hurricane Florence	1	https://nypost.com/2018/09/14/hurricane-florence-storms-north-carolina/slide-1/	https://nypost.com/real-estate/
	2	https://nypost.com/2018/09/11/the-carolinas-brace-for-hurricane-florence/slide-1/	https://nypost.com/photos/
	3	https://nypost.com/2018/09/13/hurricane-florence-closes-in-on-carolina-coast/	https://nypost.com/news/
	4	https://nypost.com/2018/09/10/hurricane-florence-looks-wild-from-the-international-space-station/	https://nypost.com/2018/10/15/surveillance-video-shows-moment-before-model-shot-in-neck-crashes-car/
	5	https://nypost.com/2018/09/20/unc-football-players-lend-hand-to-hurricane-florence-victims/	https://nypost.com/2018/10/15/beach-house-still-standing-after-hurricane-

			michael-devastates-nearby-homes/
	6	https://nypost.com/2018/09/14/as-tronaut-snaps-picture-of-hurricane-florence-as-it-makes-landfall/	https://nypost.com/2018/10/15/ex-nightlife-king-plans-to-open-a-private-club-in-noho/
	7	https://nypost.com/2018/09/15/kitten-named-survivor-rescued-with-owner-from-hurricane-florence/	https://nypost.com/2018/10/15/sears-death-of-an-american-icon/slide-1/
	8	https://nypost.com/2018/09/24/hurricane-florence-isnt-done-with-the-carolinas-just-yet/	https://nypost.com/2018/10/15/heres-what-elizabeth-warrens-dna-test-shows/
	9	https://nypost.com/2018/09/19/north-carolina-hog-lagoons-are-exploding-poo-after-hurricane-florence/	https://nypost.com/2018/10/11/michael-leaves-florida-panhandle-in-shambles/slide-1/
	10	https://nypost.com/tag/hurricanes/page/2/	https://nypost.com/2018/10/10/hurricane-michael-starts-pounding-the-florida-panhandle/slide-1/
Paul Allen	1	https://nypost.com/2018/10/15/microsoft-co-founder-paul-allen-dead-at-65/	https://nypost.com/tech/
	2	https://nypost.com/2018/10/15/microsoft-co-founder-paul-allen-dead-at-65/	https://nypost.com/columnists/
	3	https://nypost.com/2018/10/15/microsoft-co-founder-paul-allen-dead-at-65/	https://nypost.com/sports/
	4	https://nypost.com/tag/paul-allen/	https://nypost.com/news/
	5	https://nypost.com/2018/03/06/microsoft-co-founder-finds-long-lost-wwii-aircraft-carrier/	https://nypost.com/
	6	https://nypost.com/author/carleton-english/	https://nypost.com/2018/10/15/microsoft-co-founder-paul-allen-dead-at-65/
	7	https://nypost.com/tech/page/3/	https://nypost.com/2018/10/15/worlds-youngest-male-billionaire-is-25-year-old-model/

	8	https://nypost.com/tech/page/2/	https://nypost.com/2018/10/15/sears-death-of-an-american-icon/slide-1/
	9	https://nypost.com/tech/	https://nypost.com/2018/10/15/state-department-sends-out-email-featuring-cookie-monster-cat/
	10	https://nypost.com/2018/03/06/microsoft-co-founder-finds-long-lost-wwii-aircraft-carrier/	https://nypost.com/2018/10/15/de-blasios-nycha-strategy-was-deceptive-from-the-start/

5. Overlap graph



For the eight queries, in NY Post, only the last query "Paul Allen" has 2 overlaps between Lucene and PageRank algorithm, the graph is as the above shows.