Assignment 4 Report

Yuchuan Chen 4096061963

1. Steps to complete the assignment

- Download binary released solr and unzip it from given link and use command 'bin/solr start' to start solr;
- Create a core named myexample on solr using command 'bin/solr create -c myexample';
- 3) Modify the config fille in my solr core folder named managed-schema as below:

```
<field name="_root_" type="string" docValues="false" indexed="true"
stored="false"/>
<field name="_text_" type="text_general" multiValued="true" indexed="true"
stored="false"/>
<field name="_version_" type="plong" indexed="false" stored="false"/>
<copyField source="*" dest=" text "/>
```

- 4) Indexing the HTML files in solr, using command 'bin/post -c myexample filetypes html <path_to_html_files>'. For me, the command is 'bin/post -c myexample -filetypes html /Users/lacticacid/Desktop/2018\ Fall/CSCI\ 572/Homework/4/mercurynews/mercurynews html files';
- 5) Open solr UI on browser via http://localhost:8983, use the query "*:*" and successfully got the expected results;
- 6) Download MAMP for running my PHP webpage;
- Clone the solr-php-client repository on my computer and use it when developing the user interface;
- 8) Create a folder named csci572-hw4 under /Applications/MAMP/htdocs/, copy solr-php-client/Apache/Solr/Service.php to this folder, create a php file named index.php, modify the codes to provide the query box, option radio-buttons and a submit button;

- 9) Write Java codes in Eclipse with JSOUP library to get all the outgoing links in HTML pages and then create a edgelist file. Use this file as the input for calculating the pageranks;
- 10) Write Python codes in Sublime with networkx library to generate the graph using the edgelist file we got from last step and also calculate the pagerank values based on the graph. The configuration of Pagerank algorithm I used in the Python code is: alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None. Save pagerank values in a file named "external_pageRankFile.txt";
- 11) Copy the pagerank file to the data folder of my core—myexample;
- 12) Modify the config file—managed-schema by adding the field as instructed:

 <fieldType name="external" keyField="id" defVal="0"

 class="solr.ExternalFileField"/>

 <field name="pageRankFile" type="external" stored="false"

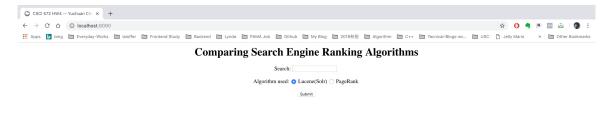
 indexed="false"/>
- 13) Reload Solr in Solr admin;
- 14) Open the query webpage via http://localhost:8090, and successfully get the results;

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

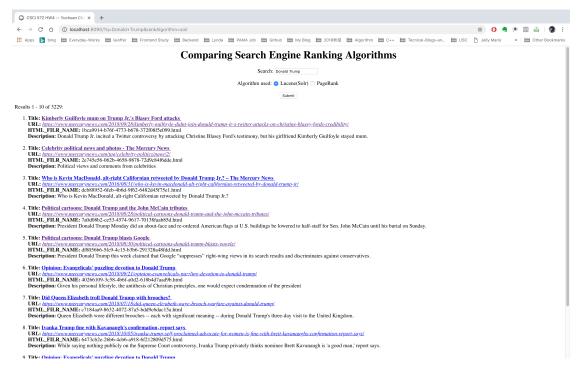
PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites. For example, https://www.mercurynews.com/tag/cartoons/, which is reachable from many other web pages, it has the higher PageRank value. If a page can be reached from more pages, it will have a higher indegrees and thus, it will have a higher PageRank value.

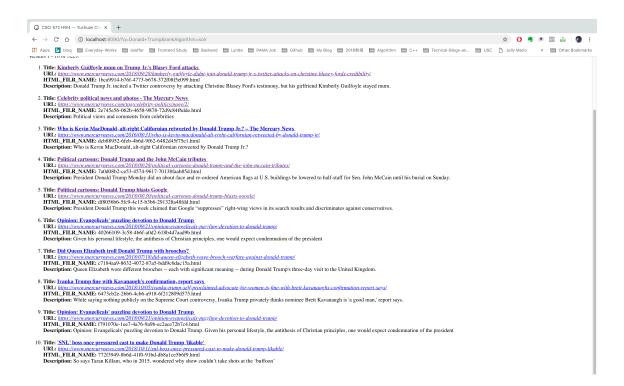
3. Screenshots

• Screenshot of initial page where you enter the query in input box.

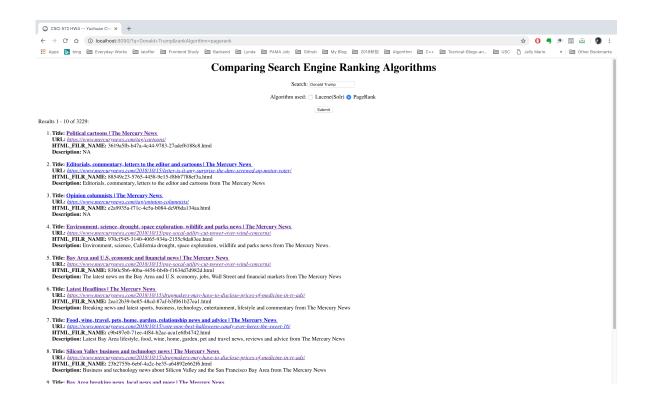


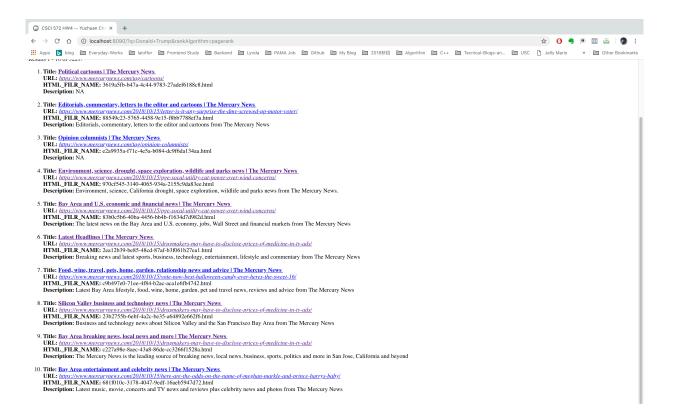
Screenshot of page which shows the results for Lucene(Default).



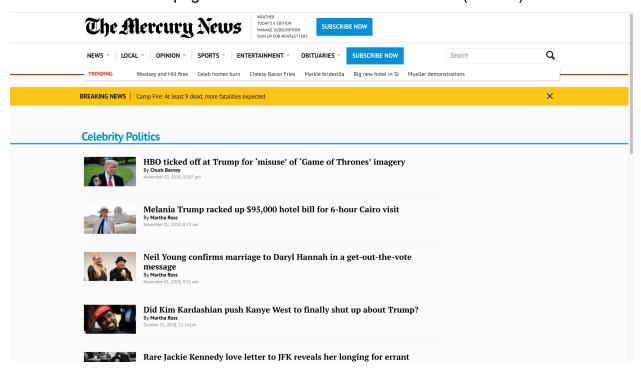


Screenshot of page which shows the results for Lucene(Default).





• Screenshot of page which shows the results for Lucene(Default).



4. URL List Table

• Donald Trump

Luncene	PageRank
https://www.mercurynews.com/2018/09/28/kimb erly-guilfoyle-didnt-join-donald-trump-jr-s-twitter- attacks-on-christine-blasey-fords-credibility/	https://www.mercurynews.com/tag/cartoons/
https://www.mercurynews.com/tag/celebrity-politics/page/2/	https://www.mercurynews.com/2018/10/15/letter-is-it-any-surprise-the-dmv-screwed-up-motor-voter/
https://www.mercurynews.com/2016/08/31/who- is-kevin-macdonald-alt-right-californian- retweeted-by-donald-trump-jr/	https://www.mercurynews.com/tag/opinion- columnists/
https://www.mercurynews.com/2018/08/28/political-cartoons-donald-trump-and-the-john-mccaintributes/	https://www.mercurynews.com/2018/10/15/pge- socal-utility-cut-power-over-wind-concerns/
https://www.mercurynews.com/2018/08/30/political-cartoons-donald-trump-blasts-google/	https://www.mercurynews.com/2018/10/15/pge-socal-utility-cut-power-over-wind-concerns/
https://www.mercurynews.com/2018/09/21/opini on-evangelicals-puzzling-devotion-to-donald- trump/	https://www.mercurynews.com/2018/10/15/drugmak ers-may-have-to-disclose-prices-of-medicine-in-tv- ads/
https://www.mercurynews.com/2018/07/18/did- queen-elizabeth-wage-brooch-warfare-against- donald-trump/	https://www.mercurynews.com/2018/10/15/vote- now-best-halloween-candy-ever-heres-the-sweet- 16/
https://www.mercurynews.com/2018/10/05/ivank a-trump-self-proclaimed-advocate-for-women-is- fine-with-brett-kavanaughs-confirmation-report- says/	https://www.mercurynews.com/2018/10/15/drugmak ers-may-have-to-disclose-prices-of-medicine-in-tv- ads/
https://www.mercurynews.com/2018/09/21/opini on-evangelicals-puzzling-devotion-to-donald- trump/	https://www.mercurynews.com/2018/10/15/drugmak ers-may-have-to-disclose-prices-of-medicine-in-tv- ads/
https://www.mercurynews.com/2018/10/11/snl- boss-once-pressured-cast-to-make-donald- trump-likable/	https://www.mercurynews.com/2018/10/15/here- are-the-odds-on-the-name-of-meghan-markle-and- prince-harrys-baby/

• LA Lakers

Luncene	PageRank

https://www.mercurynews.com/2018/05/06/nba-	https://www.mercurynews.com/2018/10/15/heres-
playoffs-2018-golden-state-warriors-kevin-	how-the-warriors-can-make-opening-night-a-
durant-game-4-highlights-video-new-orleans-	success/
pelicans-stephen-curry-score-schedule-time/	0.00000
https://www.mercurynews.com/2017/12/06/6-	https://www.mercurynews.com/2018/10/15/heralded
acre-brentwood-brush-fire-burning-uphill-near-	-stage-phenomenon-barber-shop-chronicles-
skirball-center/	comes-to-sf-bay-area/
https://www.mercurynews.com/2018/05/05/nba-	https://www.mercurynews.com/2018/10/09/and-
playoffs-2018-golden-state-warriors-new-	kerri-walsh-jennings-new-partner-is-beach-star-
orleans-pelicans-game-schedule-channel-	gets-defensive/
stream-draymond-green-gsw-stephen-curry-	
rajon-rondo/	
https://www.mercurynews.com/2018/05/07/nba-	https://www.mercurynews.com/2018/10/15/heres-
playoffs-2018-golden-state-warriors-new-	how-the-warriors-can-make-opening-night-a-
orleans-pelicans-andre-iguodala-draymond-	success/
green-hamptons-five-5-death-lineup-schedule-	
time-gsw/	
	https://www.mercurynews.com/2018/10/15/prep-
https://www.mercurynews.com/tag/warriors-live/	football-rankings-week-10-bay-area-news-group-
	top-25/
https://www.mercurynews.com/2018/05/08/gold	https://www.mercurynews.com/2018/10/14/oakland-
en-state-warriors-new-orleans-pelicans-	raiders-jon-gruden-derek-carr-news-seattle-
houston-rockets-stream-time-channel-nba-	seahawks-nfl-london-fire-contract-salary-years-
playoffs-2018-western-conference-finals-kevin-	roster-las-vegas-stadium-schedule/
durant/	
https://www.morour.movec.com/2019/01/22/iggs	https://www.mercurynews.com/2018/10/15/here-
https://www.mercurynews.com/2018/01/22/jaso n-kidd-fired-as-milwaukee-bucks-head-coach/	are-the-odds-on-the-name-of-meghan-markle-and-
II-Nuu-IIIeu-as-IIIIWaukee-bucks-IIeau-coacii/	prince-harrys-baby/
https://www.mercurynews.com/2018/10/12/live-	https://www.mercurynews.com/2018/10/15/sponsor
preseason-updates-warriors-vs-lakers-friday-at-	ed-one-of-a-kind-fresh-renovated-orinda-home-with-
730-p-m/	mt-diablo-and-hillside-views/
https://www.mercurynews.com/2018/10/13/gold	https://www.mercurynews.com/2018/10/15/kim-
en-state-warriors-vs-los-angeles-lakers-	kardashian-promoted-mental-health-during-google-
highlights-video-lebron-james-news-contract-	visit-but-what-about-unhinged-kanye-west/
salary-rumors-stephen-curry-twitter-steve-kerr/	

https://www.mercurynews.com/2018/10/15/here	https://www.mercurynews.com/2018/10/15/paul-
s-how-the-warriors-can-make-opening-night-a-	allen-microsoft-co-founder-seahawks-owner-dead-
success/	at-65/

• Star Wars

Luncene	PageRank
https://www.mercurynews.com/tag/star-wars/	https://www.mercurynews.com/2018/10/15/heres-
	how-the-warriors-can-make-opening-night-a-
	success/
https://www.mercurynews.com/2017/12/14/fans-	https://www.mercurynews.com/2018/10/15/top-20-
line-up-for-star-wars-the-last-jedi-at-the-tech-	foodie-cities/
museum/	
https://www.mercurynews.com/2018/10/08/disney	https://www.mercurynews.com/2018/10/15/vote-
land-star-wars-galaxys-edge-preview/	now-best-halloween-candy-ever-heres-the-sweet-
	16/
https://www.mercurynews.com/2018/10/08/disney	https://www.mercurynews.com/2018/10/09/and-
land-star-wars-galaxys-edge-preview/	kerri-walsh-jennings-new-partner-is-beach-star-
	gets-defensive/
http://www.ocregister.com/disneyland-will-sell-	https://www.mercurynews.com/2018/10/15/drugma
alcohol-for-the-first-time-when-star-wars-land-	kers-may-have-to-disclose-prices-of-medicine-in-
opens-in-2019	tv-ads/
http://www.ocregister.com/disneyland-is-quietly-	https://www.mercurynews.com/2018/10/15/vote-
removing-seating-and-planters-for-star-wars-land-	now-best-halloween-candy-ever-heres-the-sweet-
access	16/
https://www.mercurynews.com/2018/08/01/star-	https://www.mercurynews.com/2018/10/15/49ers-
wars-filmmaker-j-j-abrams-sends-his-first-ever-	packers-pregame-te-dwelley-added-green-bay-
tweet-heres-what-it-said/	missing-wrs/
https://www.mercurynews.com/2018/08/22/star-	https://www.mercurynews.com/2018/10/15/heres-
wars-actress-defies-racists-with-return-to-	how-the-warriors-can-make-opening-night-a-
instagram/	success/
https://www.mercurynews.com/2018/10/15/17-	https://www.mercurynews.com/2018/10/08/san-
ways-to-save-money-on-a-trip-to-disneyland/	jose-earthquakes-hire-big-time-coach/
http://www.ocregister.com/disneyland-is-quietly-	https://www.mercurynews.com/2018/10/15/here-
removing-seating-and-planters-for-star-wars-land-	are-the-odds-on-the-name-of-meghan-markle-and-
access	prince-harrys-baby/

• Lebron James

Luncene	PageRank
https://www.mercurynews.com/tag/lebron-	https://www.mercurynews.com/location/california/b
james/page/2/	ay-area/east-bay/alameda-county/
https://www.mercurynews.com/tag/lebron-james/	https://www.mercurynews.com/2018/10/15/heres-
	how-the-warriors-can-make-opening-night-a-
	success/
https://www.mercurynews.com/2018/08/04/presi	https://www.mercurynews.com/2018/10/15/drugma
dent-trump-rips-lebron-james-says-i-like-mike/	kers-may-have-to-disclose-prices-of-medicine-in-tv-
	ads/
https://www.mercurynews.com/2018/08/06/lebro	https://www.mercurynews.com/2018/10/15/heres-
n-james-son-bronny-headed-to-crossroads-	how-the-warriors-can-make-opening-night-a-
reports-say/	success/
https://www.mercurynews.com/tag/cleveland-	https://www.mercurynews.com/2018/10/15/prep-
cavaliers/	football-rankings-week-10-bay-area-news-group-
	top-25/
https://www.mercurynews.com/2018/05/30/nba-	https://www.mercurynews.com/2018/10/14/oakland
finals-2018-golden-sta-warriors-lebron-james-	-raiders-jon-gruden-derek-carr-news-seattle-
cleveland-cavaliers-game-1-time-schedule-	seahawks-nfl-london-fire-contract-salary-years-
roster-stephen-curry/	roster-las-vegas-stadium-schedule/
https://www.mercurynews.com/2018/08/04/presi	https://www.mercurynews.com/2018/10/15/sponsor
dent-trump-rips-lebron-james-says-i-like-mike/	ed-one-of-a-kind-fresh-renovated-orinda-home-
	with-mt-diablo-and-hillside-views/
https://www.mercurynews.com/2018/10/10/report	https://www.mercurynews.com/2018/10/15/drugma
-lebron-james-wont-face-warriors-in-san-jose-	kers-may-have-to-disclose-prices-of-medicine-in-tv-
friday/	ads/
https://www.mercurynews.com/2018/10/01/did-	https://www.mercurynews.com/2018/10/15/paul-
javale-mcgee-upstage-lebron-james-in-their-	allen-microsoft-co-founder-seahawks-owner-dead-
laker-debuts/	at-65/
https://www.mercurynews.com/2018/06/12/lebro	https://www.mercurynews.com/2018/10/15/raiders-
n-james-son-california-high-school/	report-card-jon-grudens-1-5-start-is-very-ugly/

• 2018 World Cup

Luncene	PageRank

pol-baseball-monta-vista-beats-lynbrook-3-1/	
s://www.mercurynews.com/tag/world-	https://www.mercurynews.com/2018/10/15/chp-
/page/2/ fa	fatal-crash-involving-caltrans-truck-on-highway-
1	101-in-san-jose/
s://www.mercurynews.com/tag/soccer/page/ h	https://www.mercurynews.com/tag/personal-
te	technology/
s://www.mercurynews.com/tag/world-cup/	https://www.mercurynews.com/location/california/b
а	ay-area/east-bay/alameda-county/
s://www.mercurynews.com/2018/09/30/us-	https://www.mercurynews.com/2018/10/15/letter-is-
ing-for-another-historic-comeback-at-the-	it-any-surprise-the-dmv-screwed-up-motor-voter/
er-cup/	
s://www.mercurynews.com/2018/07/13/world h	https://www.mercurynews.com/2018/10/15/heres-
o-final-what-time-do-france-and-croatia-play-	how-the-warriors-can-make-opening-night-a-
soccer-title/	success/
s://www.mercurynews.com/2018/05/11/startu h	https://www.mercurynews.com/tag/opinion-
orld-cup-competition-decides-best-of-28-	columnists/
tups-from-around-the-globe/	
s://www.mercurynews.com/2018/07/16/world h	https://www.mercurynews.com/2018/10/13/oakland
o-rewind-why-this-brit-and-many-like-him-are-	-as-95-7-the-game-splitting-ways-on-a-very-bitter-
ody-angry-and-still-hurt/	note/
s://www.mercurynews.com/2018/09/30/us-	https://www.mercurynews.com/2018/10/15/pge-
ing-for-another-historic-comeback-at-the-	socal-utility-cut-power-over-wind-concerns/
er-cup/	
s://www.mercurynews.com/2018/07/15/franc h	https://www.mercurynews.com/2018/10/15/heralde
ns-2nd-world-cup-title-beats-croatia-4-2/	d-stage-phenomenon-barber-shop-chronicles-
C	comes-to-sf-bay-area/

North Korea

Luncene	PageRank
https://www.mercurynews.com/2018/10/02/seou	https://www.mercurynews.com/2018/10/15/pge-
I-north-korea-estimated-to-have-20-60-nuclear-	socal-utility-cut-power-over-wind-concerns/
weapons/	
http://wirehub-digitalfirstmedia-com.go-	https://www.mercurynews.com/2018/10/15/drugmak
vip.co/two-koreas-agree-to-end-war-this-year-	ers-may-have-to-disclose-prices-of-medicine-in-tv-
pursue-denuclearization	ads/

https://www.mercurynews.com/2018/09/26/rubin	https://www.mercurynews.com/2018/10/15/sponsore
-four-reasons-why-declaring-peace-on-korean-	d-one-of-a-kind-fresh-renovated-orinda-home-with-
peninsula-is-a-bad-idea/	mt-diablo-and-hillside-views/
https://www.mercurynews.com/2017/08/29/tivo-	https://www.mercurynews.com/tag/pac-12-hotline/
apple-among-gainers-as-investors-shake-off-	
north-korea-fears/	
https://www.mercurynews.com/2018/05/14/trudy	https://www.mercurynews.com/2018/10/15/paul-
-rubin-hold-the-trump-nobel-korea-talks-will-be-	allen-microsoft-co-founder-seahawks-owner-dead-
a-long-haul/	at-65/
https://www.mercurynews.com/2018/04/19/hans	https://www.mercurynews.com/location/california/ba
on-will-syrian-airstrikes-take-us-down-slippery-	y-area/south-bay/santa-clara-county/san-jose/
slope-of-more-involvement-in-the-middle-east/	
https://www.mercurynews.com/2018/08/01/penc	https://www.mercurynews.com/location/california/ba
e-welcomes-return-of-presumed-korean-war-	y-area/
dead/	
https://www.mercurynews.com/2018/06/18/trudy	https://www.mercurynews.com/2018/10/14/woman-
-rubin-the-nauseating-flattery-trump-lavished-	dies-in-solo-i-280-rollover-collision-in-redwood-city/
on-this-mass-murderer/	
https://www.mercurynews.com/2018/06/24/the-	https://www.mercurynews.com/2018/10/15/us-eyes-
southern-poverty-law-center-has-lost-all-	west-coast-bases-for-coal-exports/
credibility/	
https://www.mercurynews.com/2018/09/26/rubin	https://www.mercurynews.com/2018/10/15/pge-
-four-reasons-why-declaring-peace-on-korean-	socal-utility-cut-power-over-wind-concerns/
peninsula-is-a-bad-idea/	
l	

• Hurricane Florence

Luncene	PageRank
https://www.mercurynews.com/2018/09/14/how-	https://www.mercurynews.com/2018/10/15/pge-
hurricane-florence-is-affecting-bay-area-	socal-utility-cut-power-over-wind-concerns/
travelers-and-what-they-need-to-know/	
https://www.mercurynews.com/tag/travel/	https://www.mercurynews.com/2018/10/15/vote-
	now-best-halloween-candy-ever-heres-the-sweet-
	16/
https://www.mercurynews.com/2018/09/27/beer-	https://www.mercurynews.com/2018/10/15/at-40-
trending-on-social-media-after-brett-	san-francisco-girls-chorus-sounds-as-young-as-
kavanaughs-references-during-testimony/	ever/

https://www.mercurynews.com/2018/09/27/beer-	https://www.mercurynews.com/tag/celebrities/
trending-on-social-media-after-brett-	
kavanaughs-references-during-testimony/	
https://www.mercurynews.com/tag/weather/pag	https://www.mercurynews.com/2018/10/14/these-
e/3/	actors-favorite-co-stars-their-horses/
https://www.mercurynews.com/2018/10/11/florid	https://www.mercurynews.com/2018/10/15/meghan
a-surveys-damage-as-michael-inundates-the-	-duchess-of-sussex-and-prince-harry-are-
carolinas/	expecting-says-kensington-palace/
https://www.mercurynews.com/2018/10/10/mich	https://www.mercurynews.com/2018/10/12/bradley-
ael-still-offshore-begins-battering-florida/	cooper-reportedly-miserable-with-model-girlfriend-
	after-bonding-with-lady-gaga/
https://www.mercurynews.com/2018/10/12/the-	https://www.mercurynews.com/2018/10/15/paul-
latest-official-michael-responsible-for-georgia-	allen-microsoft-co-founder-seahawks-owner-dead-
death/	at-65/
https://www.mercurynews.com/tag/pets/page/2/	https://www.mercurynews.com/2018/10/15/kim-
	kardashian-promoted-mental-health-during-google-
	visit-but-what-about-unhinged-kanye-west/
https://www.mercurynews.com/2018/09/14/live-	https://www.mercurynews.com/2018/10/15/washing
radar-images-of-hurricane-florences-path/	ton-archdiocese-releases-id-of-28-accused-priests/

• Paul Allen

Luncene	PageRank
https://www.mercurynews.com/2018/10/15/paul-	https://www.mercurynews.com/2018/10/15/letter-is-
allen-microsoft-co-founder-seahawks-owner-	it-any-surprise-the-dmv-screwed-up-motor-voter/
dead-at-65/	
https://www.mercurynews.com/2018/10/15/paul-	https://www.mercurynews.com/2018/10/15/letter-is-
allen-microsoft-co-founder-seahawks-owner-	it-any-surprise-the-dmv-screwed-up-motor-voter/
dead-at-65/	
https://www.mercurynews.com/2018/01/02/youtu	https://www.mercurynews.com/2018/10/15/drugma
ber-logan-paul-apologizes-for-showing-body-in-	kers-may-have-to-disclose-prices-of-medicine-in-tv-
japans-suicide-forest/	ads/
https://www.mercurynews.com/2018/09/24/vinta	https://www.mercurynews.com/2018/10/15/drugma
ge-typewriters-are-booming-at-manhattan-shop/	kers-may-have-to-disclose-prices-of-medicine-in-tv-
	ads/

https://www.mercurynews.com/2018/08/24/drug-	https://www.mercurynews.com/2018/10/15/san-
tunnel-to-mexico-found-under-abandoned-	jose-inmate-found-dead-in-cell-at-main-jail/
arizona-kfc/	
https://www.mercurynews.com/2017/09/05/car-	https://www.mercurynews.com/2018/10/15/sponsor
navigation-tech-brings-new-twists-and-turns-to-	ed-one-of-a-kind-fresh-renovated-orinda-home-
driving/	with-mt-diablo-and-hillside-views/
https://www.mercurynews.com/2018/09/11/renes	https://www.mercurynews.com/2018/10/15/drugma
as-to-acquire-u-s-chipmaker-idt-for-6-7-billion-3/	kers-may-have-to-disclose-prices-of-medicine-in-tv-
	ads/
https://www.mercurynews.com/2018/08/06/tensi	https://www.mercurynews.com/location/california/b
ons-flare-in-brazil-as-it-considers-loosening-	ay-area/south-bay/santa-cruz-county/
abortion-laws/	
https://www.mercurynews.com/2018/09/14/misc	https://www.mercurynews.com/2018/10/15/paul-
hel-creator-of-the-marshmallow-test-dies-at-88/	allen-microsoft-co-founder-seahawks-owner-dead-
	at-65/
https://www.mercurynews.com/2018/07/12/looki	https://www.mercurynews.com/2018/10/15/paul-
ng-up-pc-sales-rise-for-first-time-in-six-years/	allen-microsoft-co-founder-seahawks-owner-dead-
	at-65/

5. Overlap graph

Query	Overlap
Donald Trump	0
LA Lakers	1
Star Wars	0
Lebron James	0
2018 World Cup	0
North Korea	0
Hurricane Florence	0
Paul Allen	2

