HW - 4

Step 1 Installing Solr and Running Apache:

- Installed Solr on mac and verified the connection, since i am working on mac, Apache runs by default on it.
- Started Solr using bin/solr start command.
- Followed the pdf to index the html files from google drive by creating the core **myexample** using **bin/solr create –c myexample**
- In the managed-schema folder, solrconfig.xml is edited accordingly.
- · Indexing of the HTML files are performed using given command in the document.
- Made sure apache is running on mac by typing sudo apachectl restart and running the given php file.

Step 2 Generating EdgeList:

- A java program is written using by Jsoup jar file as external resources.
- The program creates mapping files which are csv in format and fileUrlMap and urlFileMap are generated using these csv.
- All the outgoing links that are traversed from the folder which are saved in a file called edgeList.txt where the webpages acts as nodes.

Step 3 Computing Page Rank:

 The following commands are used in python along with networkx to compute the page rank of edgeList.txt by parsing the entire file.

G = nx.read edgelist("/Users/siri/IR/edgeList.txt", create using=nx.DiGraph())

pr = nx.pagerank(G,alpha=0.85, personalization=None, max_iter=30,
tol=1e-06, nstart=None, weight='weight',dangling=None)

```
import sys
>>> f = open("external_pageRankFile.txt",'w')
>>> sys.stdout = f
>>> for k, v in pr.items():
... print(str(k)+'='+str(v))
```

Step 4 Changes to the schema:

- Changes are made to the managed schema as given such as the fieldname as 'pageRankFile' and type as 'external'.
- Defined listeners within the <query> element in the solrconfig.xml file as mentioned.
- Placed the files in appropriate folders and restarted solr and loaded the core.
- Verified whether solr is running by executing the query with and without the sort as "pageRankFile desc"

Step 5 Changes to PHP file:

- Additional parameters are added as sort => pageRankFile desc.
- Every query is mapped from a URLtoHTML_mercury.csv file to make sure that URL exists.
- Title, id, description, URL are filtered out of every output.
- Radio buttons are added for the users to choose to execute Licene or PageRank.

NOTE:

Two parameters influence the page rank value

• If a page exists with many number of links or if a page is linked to another page which has got higher rank value, then the current page will have high page rank value.

Comparison Results:

1.Donald Trump

Solr Lucene	Evtornal Paga Pank
	External Page Rank
https://www.mercurynews.com/2018/09/28/kimberly-guilfoyle-didnt-join-donald-trump-jr-s-twitter-attacks-on-christine-blasey-fords-credibility/	https://www.mercurynews.com/ 2018/06/06/multi-million-dollar- prostitution-ring-stretched-from- kansas-to-bay-area-feds-say/
NA	https://www.mercurynews.com/ 2018/10/08/where-theres-smoke- solano-county-fire-draws- shutterbugs-to-capture-smoke-in- bay-area/
https://www.mercurynews.com/ 2016/08/31/who-is-kevin- macdonald-alt-right-californian- retweeted-by-donald-trump-jr/	https://www.mercurynews.com/ 2018/05/18/bay-area-economy- roars-life-may-adds-11000-jobs- google-apple-facebook/
https://www.mercurynews.com/ 2018/10/05/ivanka-trump-self- proclaimed-advocate-for-women-is- fine-with-brett-kavanaughs- confirmation-report-says/	https://www.mercurynews.com/ 2018/09/12/great-design-takes- center-court-at-the-u-s-open/
https://www.mercurynews.com/ 2018/07/18/did-queen-elizabeth- wage-brooch-warfare-against- donald-trump/	https://www.mercurynews.com/ 2018/10/03/raiders-report-five- things-we-learned-wednesday-3/

https://www.mercurynews.com/ 2018/09/24/more-ivanka-trump- complicity-first-daughter-stays- silent-on-brett-kavanaugh- controversy/	https://www.mercurynews.com/ 2018/07/27/justices-suit-by-trump- backers-against-san-jose-police- can-proceed/
https://www.mercurynews.com/ 2018/08/28/political-cartoons- donald-trump-and-the-john-mccain- tributes/	https://www.mercurynews.com/ 2018/05/14/trudy-rubin-hold-the- trump-nobel-korea-talks-will-be-a- long-haul/
https://www.mercurynews.com/ 2018/08/30/political-cartoons- donald-trump-blasts-google/	https://www.mercurynews.com/ 2018/09/05/real-reason-why- kaepernicks-nike-ad-troubles- military-veterans/
https://www.mercurynews.com/ 2018/09/21/opinion-evangelicals- puzzling-devotion-to-donald-trump/	https://www.mercurynews.com/ 2018/10/15/what-readers-had-to- say-about-revisiting-columbus-day/
https://www.mercurynews.com/ 2018/09/21/opinion-evangelicals- puzzling-devotion-to-donald-trump/	NA

2. LA Lakers

Solr Lucene	external Page Rank
https://www.mercurynews.com/ 2018/05/06/nba-playoffs-2018- golden-state-warriors-kevin-durant- game-4-highlights-video-new- orleans-pelicans-stephen-curry- score-schedule-time/	https://www.mercurynews.com/2018/10/15/what-readers-had-to-say-about-revisiting-columbus-day/

https://www.mercurynews.com/2017/12/06/6-acre- brentwood-brush-fire-burning-uphill-near-skirball- center/	NA
https://www.mercurynews.com/2018/05/05/nba-playoffs-2018-golden-state-warriors-new-orleans-pelicans-game-schedule-channel-stream-draymond-green-gsw-stephen-curry-rajon-rondo/	https://www.mercurynews.com/2018/10/10/live-preseason-updates-warriors-vs-lakers-wednesday-at-730-p-m/
https://www.mercurynews.com/2018/05/07/nba-playoffs-2018-golden-state-warriors-new-orleans-pelicans-andre-iguodala-draymond-green-hamptons-five-5-death-lineup-schedule-time-gsw/	https://www.mercurynews.com/2018/06/06/highlight-reel-weeks-high-school-standouts-36/
https://www.mercurynews.com/2018/05/08/golden- state-warriors-new-orleans-pelicans-houston- rockets-stream-time-channel-nba-playoffs-2018- western-conference-finals-kevin-durant/	https://www.mercurynews.com/2018/03/09/warriors-tonight-curry-out-against-blazing-hot-portland/
NA	https://www.mercurynews.com/2018/10/15/warriors-report-how-will-golden-state-fare-on-ring-night/
https://www.mercurynews.com/2018/10/12/live- preseason-updates-warriors-vs-lakers-friday-at-730- p-m/	https://www.mercurynews.com/2018/10/06/inside-kevin-durants-return-to-seattle-in-warriors-preseason-win-over-kings/
https://www.mercurynews.com/2018/01/22/jason-kidd-fired-as-milwaukee-bucks-head-coach/	https://www.mercurynews.com/2018/07/27/why-warriors-biggest-threat-vows-well-be-right-there/
https://www.mercurynews.com/2018/10/13/golden- state-warriors-vs-los-angeles-lakers-highlights- video-lebron-james-news-contract-salary-rumors- stephen-curry-twitter-steve-kerr/	https://www.mercurynews.com/2018/02/22/planet-dubs-podcast-exclusive-how-jerry-west-helped-build-two-dynasties/
https://www.mercurynews.com/2018/10/15/heres- how-the-warriors-can-make-opening-night-a- success/	https://www.mercurynews.com/2018/05/21/hit-and-run-victims-should-not-chase-the-bad-guys-roadshow/

3. Star Wars

Solr Lucene	External Page Rank
NA	https://www.mercurynews.com/2018/02/17/photos- olympics-mens-large-hill-individual-ski-jumping/
https://www.mercurynews.com/2017/12/14/fans-line-up-for-star-wars-the-last-jedi-at-the-tech-museum/	https://www.mercurynews.com/2018/01/07/golden-globes-dressed-in-black-who-stood-out-the-good-and-the-bad/

https://www.mercurynews.com/2018/10/08/disneyland-star-wars-galaxys-edge-preview/	https://www.mercurynews.com/2018/10/10/live- preseason-updates-warriors-vs-lakers-wednesday- at-730-p-m/
https://www.mercurynews.com/2018/10/08/disneyland-star-wars-galaxys-edge-preview/	https://www.mercurynews.com/2017/05/12/bay-area-greats-steve-nash-jason-kidd-entering-the-hall-of-fame-this-year-but-not-that-one/
http://www.ocregister.com/disneyland-will-sell-alcohol-for-the-first-time-when-star-wars-land-opens-in-2019	https://www.mercurynews.com/2018/10/10/picks-of-the-week-edge-to-washington-advantage-usc-comfortable-win-for-utah-and-no-idea-what-to-make-of-cal-ucla/
https://www.mercurynews.com/2018/08/01/star- wars-filmmaker-j-j-abrams-sends-his-first-ever- tweet-heres-what-it-said/	https://www.mercurynews.com/2018/10/13/peek-inside-a-seattle-cannabis-tour/
http://www.ocregister.com/disneyland-is-quietly-removing-seating-and-planters-for-star-wars-land-access	https://www.mercurynews.com/2018/10/05/49ers-practice-top-things-we-learned-friday/
https://www.mercurynews.com/2018/08/22/star- wars-actress-defies-racists-with-return-to-instagram/	https://www.mercurynews.com/2018/06/06/highlight-reel-weeks-high-school-standouts-36/
https://www.mercurynews.com/2018/10/15/17-ways-to-save-money-on-a-trip-to-disneyland/	https://www.mercurynews.com/2018/03/09/warriors-tonight-curry-out-against-blazing-hot-portland/
http://www.ocregister.com/disneyland-is-quietly-removing-seating-and-planters-for-star-wars-land-access	https://www.mercurynews.com/2018/02/01/meet-the-cast-of-survivor-ghost-island/

4. Lebron James

Solr Lucene	External page rank
NA	https://www.mercurynews.com/2018/06/06/multi-million-dollar-prostitution-ring-stretched-from-kansas-to-bay-area-feds-say/
NA	https://www.mercurynews.com/2018/10/03/raiders-report-five-things-we-learned-wednesday-3/
https://www.mercurynews.com/2018/08/06/lebron- james-son-bronny-headed-to-crossroads-reports- say/	https://www.mercurynews.com/2018/10/15/what-readers-had-to-say-about-revisiting-columbus-day/
NA	https://www.mercurynews.com/2018/10/10/live-preseason-updates-warriors-vs-lakers-wednesday-at-730-p-m/

https://www.mercurynews.com/2018/08/04/ president-trump-rips-lebron-james-says-i-like-mike/	https://www.mercurynews.com/2018/09/14/campbell-city-council-candidates-discuss-housing-traffic-measure-o/
https://www.mercurynews.com/2018/05/30/nba-finals-2018-golden-sta-warriors-lebron-james-cleveland-cavaliers-game-1-time-schedule-roster-stephen-curry/	https://www.mercurynews.com/2018/06/28/klay-thompson-outraged-over-hard-foul-in-pickup-game-on-china-trip/
https://www.mercurynews.com/2018/08/04/ president-trump-rips-lebron-james-says-i-like-mike/	https://www.mercurynews.com/2018/09/24/san-jose-state-football-meets-sjpd-for-bridge-building/
https://www.mercurynews.com/2018/10/01/did- javale-mcgee-upstage-lebron-james-in-their-laker- debuts/	https://www.mercurynews.com/2018/08/27/source-toyota-to-invest-500m-in-uber/
https://www.mercurynews.com/2018/10/10/report-lebron-james-wont-face-warriors-in-san-jose-friday/	https://www.mercurynews.com/2018/06/06/ southern-california-lawmaker-recalled-stops- democratic-supermajority/
https://www.mercurynews.com/2018/08/29/lebron-talks-about-fatherhood-double-standards-and-the-nword-in-hbo-premiere/	NA

5. 2018 World Cup

Solr Lucene	External Page Rank
https://www.mercurynews.com/2018/04/26/high-school-baseball-monta-vista-beats-lynbrook-3-1/	https://www.mercurynews.com/2018/06/06/multi-million-dollar-prostitution-ring-stretched-from-kansas-to-bay-area-feds-say/
NA	https://www.mercurynews.com/2018/10/08/where-theres-smoke-solano-county-fire-draws-shutterbugs-to-capture-smoke-in-bay-area/
NA	https://www.mercurynews.com/2018/05/18/bay-area-economy-roars-life-may-adds-11000-jobs-google-apple-facebook/
NA	https://www.mercurynews.com/2018/09/12/great-design-takes-center-court-at-the-u-s-open/
https://www.mercurynews.com/2018/09/30/us-looking-for-another-historic-comeback-at-the-ryder-cup/	NA
https://www.mercurynews.com/2018/07/13/world- cup-final-what-time-do-france-and-croatia-play-for- soccer-title/	https://www.mercurynews.com/2017/03/28/last-piece-of-workhorse-bridge-floats-away/

https://www.mercurynews.com/2018/07/16/world- cup-rewind-why-this-brit-and-many-like-him-are- bloody-angry-and-still-hurt/	https://www.mercurynews.com/2018/10/03/pac-12-volleyball-libero-from-shp-expands-social-media-brand/
https://www.mercurynews.com/2018/05/11/startup-world-cup-competition-decides-best-of-28-startups-from-around-the-globe/	https://www.mercurynews.com/2018/02/17/photos- olympics-mens-large-hill-individual-ski-jumping/
NA	https://www.mercurynews.com/2018/10/03/raiders-report-five-things-we-learned-wednesday-3/
https://www.mercurynews.com/2018/09/30/us-looking-for-another-historic-comeback-at-the-ryder-cup/	https://www.mercurynews.com/2018/05/24/ european-privacy-law-to-shake-up-entire-tech- industry-and-impact-teens/

6. North Korea

Solr lucene	External Page Rank
http://wirehub-digitalfirstmedia-com.go-vip.co/two-koreas-agree-to-end-war-this-year-pursue-denuclearization	https://www.mercurynews.com/2018/06/06/multi-million-dollar-prostitution-ring-stretched-from-kansas-to-bay-area-feds-say/
https://www.mercurynews.com/2018/10/02/seoul- north-korea-estimated-to-have-20-60-nuclear- weapons/	https://www.mercurynews.com/2018/10/03/pac-12-volleyball-libero-from-shp-expands-social-media-brand/
https://www.mercurynews.com/2018/09/26/rubin- four-reasons-why-declaring-peace-on-korean- peninsula-is-a-bad-idea/	https://www.mercurynews.com/2018/02/17/photos- olympics-mens-large-hill-individual-ski-jumping/
https://www.mercurynews.com/2018/05/14/trudy-rubin-hold-the-trump-nobel-korea-talks-will-be-a-long-haul/	https://www.mercurynews.com/2018/05/14/trudy-rubin-hold-the-trump-nobel-korea-talks-will-be-a-long-haul/
https://www.mercurynews.com/2017/08/29/tivo-apple-among-gainers-as-investors-shake-off-north-korea-fears/	https://www.mercurynews.com/2018/10/15/what-readers-had-to-say-about-revisiting-columbus-day/
https://www.mercurynews.com/2018/04/19/hanson-will-syrian-airstrikes-take-us-down-slippery-slope-of-more-involvement-in-the-middle-east/	https://www.mercurynews.com/2018/10/10/picks-of-the-week-edge-to-washington-advantage-usc-comfortable-win-for-utah-and-no-idea-what-to-make-of-cal-ucla/
https://www.mercurynews.com/2018/08/01/pence-welcomes-return-of-presumed-korean-war-dead/	https://www.mercurynews.com/2018/10/13/peek-inside-a-seattle-cannabis-tour/

https://www.mercurynews.com/2018/09/26/rubin- four-reasons-why-declaring-peace-on-korean- peninsula-is-a-bad-idea/	https://www.mercurynews.com/2018/10/03/big-downtown-san-jose-residential-complex-sprouts-land-deal-google-adobe-amazon/
https://www.mercurynews.com/2018/06/18/trudy-rubin-the-nauseating-flattery-trump-lavished-on-this-mass-murderer/	https://www.mercurynews.com/2017/04/24/three- men-arrested-in-gang-related-stabbing-in-san- mateo/
https://www.mercurynews.com/2018/06/24/the-southern-poverty-law-center-has-lost-all-credibility/	https://www.mercurynews.com/2018/10/02/expect-madison-bumgarner-to-be-a-topic-of-discussion-ingiants-gm-interviews/

7. Hurricane Florence

Solr Lucene	External Page Rank
https://www.mercurynews.com/2018/09/14/how-hurricane-florence-is-affecting-bay-area-travelers-and-what-they-need-to-know/	NA
https://www.mercurynews.com/2018/09/27/beer-trending-on-social-media-after-brett-kavanaughs-references-during-testimony/	https://www.mercurynews.com/2018/09/17/will-in- perhaps-the-years-most-instructive-house-race- democrats-seem-serious-about-winning/
https://www.mercurynews.com/2018/09/27/beer-trending-on-social-media-after-brett-kavanaughs-references-during-testimony/	NA
NA	https://www.mercurynews.com/2018/02/01/meet-the-cast-of-survivor-ghost-island/
NA	https://www.mercurynews.com/2017/11/09/alec-baldwin-in-full-blown-trump-mode-is-coming-to-san-franciscos-curran-theater/
https://www.mercurynews.com/2018/10/11/florida- surveys-damage-as-michael-inundates-the- carolinas/	NA
https://www.mercurynews.com/2018/10/10/michael-still-offshore-begins-battering-florida/	https://www.mercurynews.com/2018/06/09/airbnb-to-launch-new-disaster-relief-registry-in-san-jose/
https://www.mercurynews.com/2018/10/12/the- latest-official-michael-responsible-for-georgia-death/	https://www.mercurynews.com/2018/10/08/so-is-dakota-johnson-having-chris-martins-baby/
NA	https://www.mercurynews.com/2018/10/02/drunken-film-festival-pub-crawls-its-way-to-oakland/

https://www.mercurynews.com/2018/09/14/live-
radar-images-of-hurricane-florences-path/

http://www.eastbaytimes.com/family-aghast-after-funeral-homes-premature-cremation-but-jury-rules-they-didnt-suffer-serious-emotional-distress

8. Paul Allen

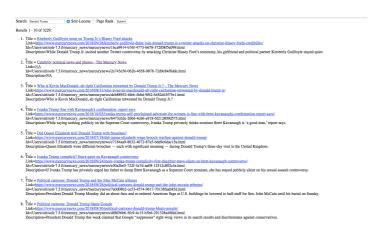
Solr lucene	external Page Rank
https://www.mercurynews.com/2018/10/15/paul- allen-microsoft-co-founder-seahawks-owner-dead- at-65/	https://www.mercurynews.com/2018/10/03/raiders-report-five-things-we-learned-wednesday-3/
https://www.mercurynews.com/2018/10/15/paul- allen-microsoft-co-founder-seahawks-owner-dead- at-65/	https://www.mercurynews.com/2018/10/15/watch-what-happens-when-you-fly-a-drone-is-lightning-storm/
https://www.mercurynews.com/2018/01/02/ youtuber-logan-paul-apologizes-for-showing-body- in-japans-suicide-forest/	https://www.mercurynews.com/2018/05/14/trudy-rubin-hold-the-trump-nobel-korea-talks-will-be-a-long-haul/
https://www.mercurynews.com/2018/09/11/renesas- to-acquire-u-s-chipmaker-idt-for-6-7-billion-3/	https://www.mercurynews.com/2018/10/15/what-readers-had-to-say-about-revisiting-columbus-day/
https://www.mercurynews.com/2018/07/12/looking-up-pc-sales-rise-for-first-time-in-six-years/	https://www.mercurynews.com/2017/11/10/ibm-ups-pressure-on-rivals-with-quantum-computer-prototype-2/
https://www.mercurynews.com/2018/07/20/deadly-gaza-strike-follows-death-of-israeli-soldier/	https://www.mercurynews.com/2018/09/14/ campbell-city-council-candidates-discuss-housing- traffic-measure-o/
https://www.mercurynews.com/2017/10/02/roku- cuts-price-on-top-streaming-player-to-counter-apple- tv/	https://www.mercurynews.com/2018/07/27/foothills-sound-of-music-as-big-and-boisterous-as-original-classic/
https://www.mercurynews.com/2018/09/24/vintage- typewriters-are-booming-at-manhattan-shop/	https://www.mercurynews.com/2018/08/27/source-toyota-to-invest-500m-in-uber/
https://www.mercurynews.com/2017/09/05/car-navigation-tech-brings-new-twists-and-turns-to-driving/	https://www.mercurynews.com/2018/06/24/discontent-lurks-in-vietnam-despite-economic-success/
https://www.mercurynews.com/2018/08/06/tensions-flare-in-brazil-as-it-considers-loosening-abortion-laws/	https://www.mercurynews.com/2018/08/21/h-1b-visa-reliant-cisco-secured-visas-for-foreign-workers-instead-of-hiring-u-s-citizens-report/

Flow for Donald Trump

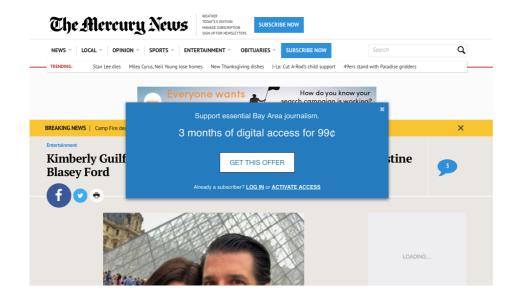
Solr Lucene



When searched



When link is clicked



When searched using page rank

O Solr-Lucene O Page Rank Submit

1. Title = Suspects in large prostitution ring arrested in Bay Area
Link=https://www.mercurynews.com/2018/06/06/multi-million-dollar-prostitution-ring-stretched-from-kansas-to-bay-area-feds-say/
Id=/Users/Siris/ols-75.0/mercurynews/0234421d-2455-476a-84b4-16e2c1c55da5.html
Description=A federal indictment alleges that a San Mateo couple built up a multimillion-dollar prostitution business out of massage parlors in Kansas and used the money to buy real estate in the Bay Area.

2. Title = Health, disease, medicine and fitness news I The Mercury News
Link=https://www.mercurynews.com/2018/10/08/where-theres-smoke-solano-county-fire-draws-shutterbugs-to-capture-smoke-in-bay-area/
Id=/Users/sir/solar-7.5/0/mercury_news/mercurynews/40d50be1-3419-404e-9840-1ae629dd9da5.html
Description=NA

3. Title = Bay Area economy roars to life in April, adds 11.000 jobs
Link=https://www.mercurynews.com/2018/05/18/bay-area-economy-roars-life-may-adds-11000-jobs-google-apple-facebook/
Id=/Users/kir/solv-75.0/mercuryn.ews/fems/ruynews/8618/34-ac5-401b-a637-ae2e5d648f0a.html
Description=Santa Clara County led the charge in the April employment gains and added 7,000 jobs, while the East Bay gained 2,000.

4. Title = New courtside design at the U.S Open a grand slam
Link=https://www.mercurynews.com/2018/09/12/great-design-takes-center-court-at-the-u-s-open/
Id=//Users/sir/solr-7.5.0/mercury_news/mercurynews/69/1dds44-c8d5-4368-a07a-07cb48c8a1f2.html
Description=A redesigned "courtscape" to celebrate the U.S. Open's 50th anniversary includes a new umpire tower, modernized player and line umpire chairs, and many improvements to reduce visual clutter and imp...

Title = Raiders report: Five things we learned Wednesday Link=https://www.mercurynews.com/2018/10/03/raiders-

Title = Randers Teport, Pres titings we related rectorestary.

Link-altips://www.mercurpnews.com/2018/1009/aiders-report-five-things-we-learned-wednesday-3/
Lia-/Users/siris/oi-7-3/bimercury_news/mercurpnews/9e44/72a-0206-0404-97e1-2fa4e3/1521-8.html

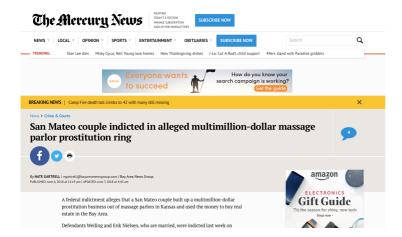
Descriptions-fixed Daryl Worley returns to practice, the offensive line is wounded and other tidbits from Wednesday's practice session.

Link=https://www.mercurynews.com/2018/07/27/justices-suit-by-trump-backers-against-san-jose-police-can-proceed/
Id=/Users/siri/soli-7.5.0/mercury_news/mercurynews/40ed5449-a564-4219-9572-7d5b81286515.html
Description=A federal appeals court on Friday allowed supporters of then-presidential candidate Donald Trump to proceed with a lawsuit alleging they were beaten after San Jose police steered them into a crowd ...

7. Title = Rubin: Hold the Trump Nobel! Korea talks will be a long haul — The Mercury News
Link=https://www.mercurynews.com/2018/05/14/rudy-rubin-hold-the-trump-nobel-korea-talks-will-be-a-long-haul/
ld=//Users/kir/sb-17-5/Johrecury_news/edriv58-8694-46de-kao-fe/0435956 html
Description=If Trump falls for his own hype he may be outwitted by a showman who can best him at this game.

8. Title = Real reason why Kaepernick's Nike ad troubles military veterans
Link=https://www.mercurynews.com/2018/09/05/real-reason-why-kaepernicks-nike-ad-troubles-military-veterans/
Id=/Users/sin/solr-7.5.0/mercury_news/mercurynews/d35daa54-0095-4049-8390-dec504d87d14.html

When link is clicked



Overlap table:

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

Paul Allen	0
------------	---

Overlap graph

