**CSCI572\_Hw4\_Report\_Haoyu Jiang**

**Steps:**

**Step1.**

I followed the instructions at link to install Solr directly in Mac and learned how to create core and index file following INIndexingwithTIKAV3.pdf. Then I downloaded LA Time which contains a set of html files and the mapping file contains mappings between filenames and urls. The latimes file and mapping file were putted in the solr folder. I created new core called “csci572hw4” in solr and index the html files using bash sentences as followed:

bin/solr create -c csci572hw4

bin/post -c csci572hw4 -filetypes html /Users/daisy/solr-7.5.0/latimes/

At this time, I could go to <http://localhost:8983/solr/#/csci572hw4> and query with different condition to get feedback sorted based on Lucene.

Then I tried to create a php client called client.php. Before writing client.php, I git clone solr-php-client in /Library/WebServer/Documents/ and put the client.php in the same path. At the same time, I activated OS X’s builtin Apache Web Server and PHP following the link: https://cachecrew.com/activating-os-xs-builtin-apache-web-server-php/.

My client.php realized three steps:

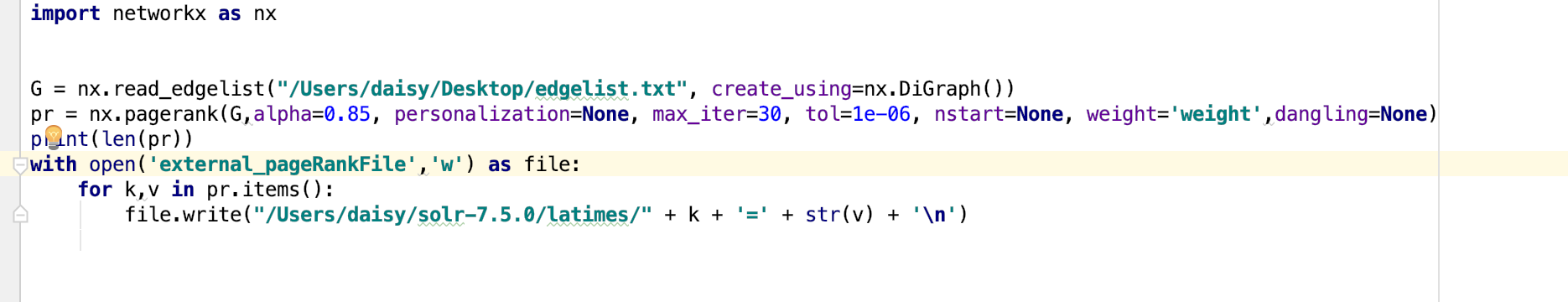
1. Accepted input from the user in a HTML form and sent the request to the Solr server.
2. After Solr server processing the query, it returns the results parsed by the PHP program and formatted as clickable title, clickable url, ID and description to display.
3. Realized previous page and next page function in order to check all results.

During the writing period, I continuously changed and improved the code with checking the results on the url: http://localhost/client.php.

**Step2**

For the edgelist part, I detected edge between two files representing a link between two and used these mappings to filter out urls that are not present in the mapping file by java as followed:

For the pagerank part, I implemented networkx library in python to get a dictionary of graph nodes with corresponding PageRank scores with parameters: “alpha=0.85, personalization=**None**, max\_iter=30, tol=1e-06, nstart=**None**, weight=**'weight'**,dangling=**None”.**



Then I put the pagerankfile without “.txt” in solr/data. Change managed-schema file and solrconfig.xml by adding parameters following the instructions.

After reloading the index, I could choose pagerankfile desc/asc sorting with query to get different feedback in solr UI. Then I revised the client.php and added pagerank desc sorting so that I could choose this kind of sorting at <http://localhost/client.php> and get answers with the ranking based on PageRank scores that were defined in the external file.

Step3

Compared the results of the two ranking algorithms under different queries.

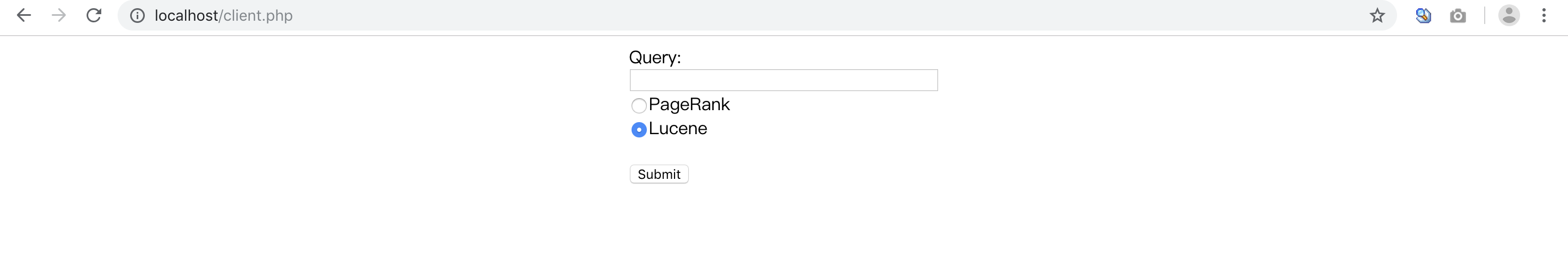
**Explanation**

**Explanation for higher PageRank values pages:**

Pages with higher PageRank values tend to be linked (or have many incoming links) by many other pages. Especially a page that is linked to by many pages with high PageRank receives a high rank itself.

**Screenshots**

**Initial page**

****

**The results for Lucene(Default)**

**图片包含 屏幕截图



自动生成的说明**

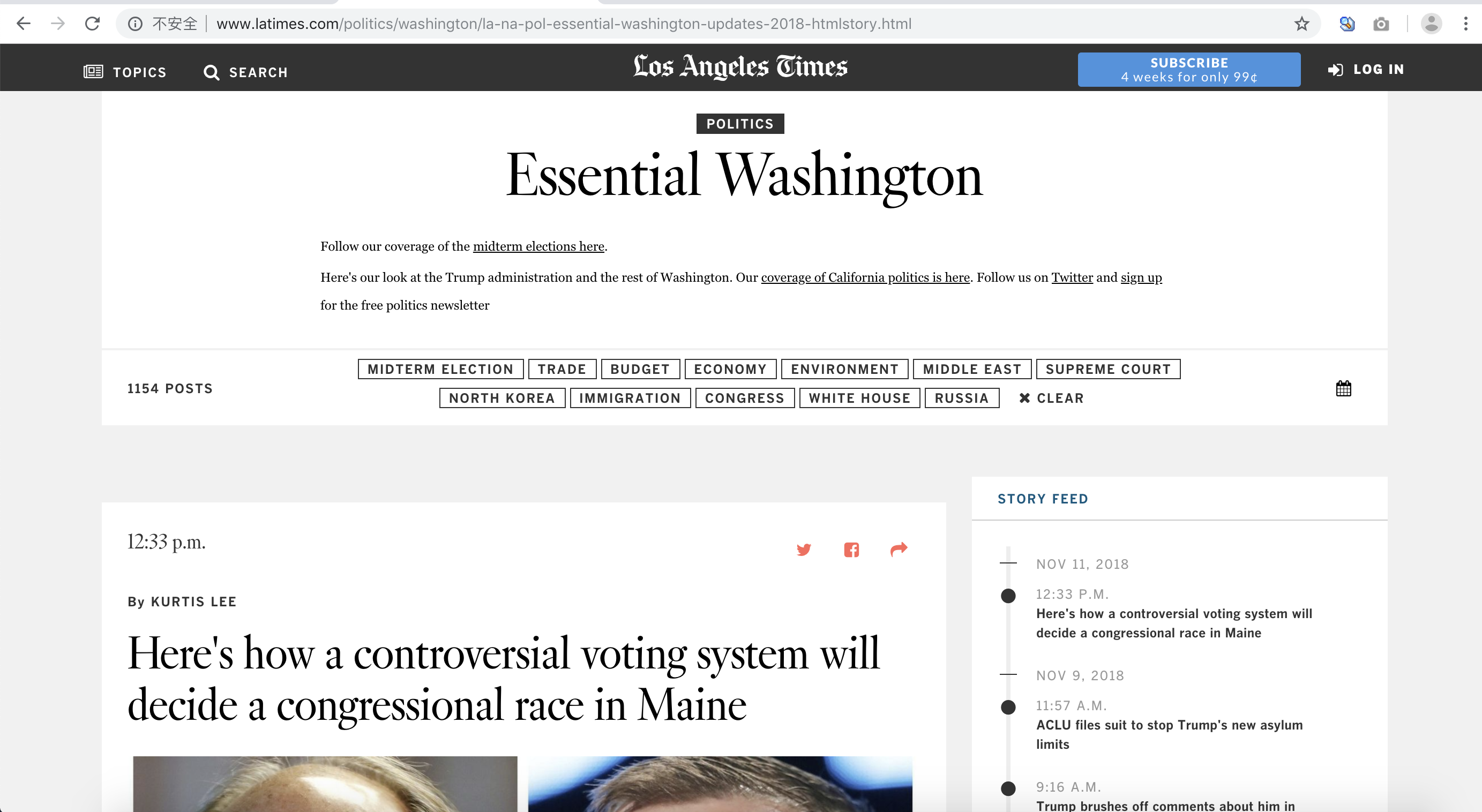
**The results for PageRank图片包含 屏幕截图



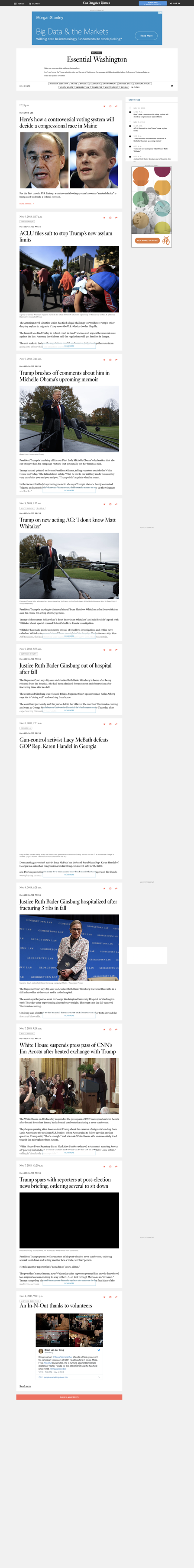
自动生成的说明**

**Actual webpage after clicking on one result link**

Part of this page:

****

Whole page:



**Tables**

**Donald Trump**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com/nation/politics/trailguide/la-na-trailguide-updates-trump-release-new-doctor-s-letter-1473951652-htmlstory.html> | <http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html> |
| <http://www.latimes.com/nation/politics/trailguide/la-na-trailguide-updates-donald-trump-open-to-allowing-those-in-1473296491-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/politics/la-pol-ca-donald-trump-california-races-downticket-20160527-snap-htmlstory.html> | <http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html> |
| <http://www.latimes.com/opinion/op-ed/la-oe-johnston-trump-cons-and-cheats-20181004-story.html> | <http://www.latimes.com/politics/la-na-pol-kavanaugh-ford-hearing-htmlstory.html> |
| <http://www.latimes.com/sports/olympics/la-sp-olympics-live-updates-the-moment-when-the-donald-trump-and-kim-1518187562-htmlstory.html> | <http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html> |
| <http://www.latimes.com/opinion/la-ol-enter-the-fray-whom-to-believe-michael-cohen-or-donald-1532702692-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/espanol/politica/la-es-dos-funerales-y-una-boda-el-rechazo-a-donald-trump-20180828-story.html> | <http://www.latimes.com/entertainment/la-et-entertainment-news-updates-2018-htmlstory.html> |
| <http://www.latimes.com/nation/politics/trailguide/la-na-live-updates-trailguide-1475872277-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/opinion/la-ol-enter-the-fray-donald-trump-might-actually-have-to-1535040294-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/entertainment/tv/la-et-st-donald-trump-60-minutes-20161110-story.html> | <http://www.latimes.com/politics/essential/la-pol-ca-essential-politics-may-2018-htmlstory.html> |

**LA Lakers**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com> | <http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-southern-california-2017-htmlstory.html> |
| <http://www.latimes.com> | <http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-plaschke-20180924-story.html> | <http://www.latimes.com/sports/dodgers/la-sp-live-dodgers-brewers-nlcs-htmlstory.html> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-plaschke-20180924-story.html> | <http://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html> |
| <http://www.latimes.com> | <https://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html> |
| <http://www.latimes.com> | http://www.latimes.com |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-media-day-20180924-story.html> | <http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html> |
| <http://www.latimes.com> | <http://www.latimes.com/politics/la-na-pol-kavanaugh-ford-hearing-htmlstory.html> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-sidebar-20180924-story.html> | <http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-5-questions-20181013-story.html> | <http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-get-the-eric-sondheimer-s-latest-1490657918-htmlstory.html> |

**Star Wars**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com/entertainment/movies/la-ca-mn-star-wars-timeline-20180516-story.html> | <http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html> |
| <http://www.latimes.com/entertainment/movies/la-et-mn-star-wars-last-jedi-characters-20171214-htmlstory.html> | <http://www.latimes.com/entertainment/la-et-entertainment-news-updates-2018-htmlstory.html> |
| <http://www.latimes.com/travel/themeparks/la-trb-star-wars-land-disneyland-20150817-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/entertainment/movies/la-et-mn-ron-howard-star-wars-han-solo-20170622-story.html> | <http://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html> |
| <https://www.latimes.com/fashion/la-ig-wwd-rag-and-bones-star-wars-20171106-story.html> | <http://www.latimes.com/about/la-privacypolicy-20180703-story.html> |
| <http://www.latimes.com/fashion/la-ig-wwd-rag-and-bones-star-wars-20171106-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/entertainment/herocomplex/la-et-hc-jon-favreau-star-wars-mandalorian-20181003-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/entertainment/herocomplex/la-et-hc-comic-con-star-wars-fandom-20180723-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/brandpublishing/travelplus/lasvegasguide/features/la-ss-vegas-the-hot-list-burlesque-star-wars-speakeasies-20180620dto-story.html> | <http://www.latimes.com/travel/la-tr-california-bucket-list-updates-2017-htmlstory.html> |
| <http://www.latimes.com/entertainment/music/la-et-ms-star-wars-secrets-empire-20180116-story.html> | <http://www.latimes.com> |

**Lebron James**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-diet-20181004-story.html?outputType=amp> | <http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-20181013-story.html?outputType=amp> | <http://www.latimes.com> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-nuggets-20181002-story.html> | <http://www.latimes.com> |
| <https://www.latimes.com/sports/lakers/la-sp-lakers-nuggets-20181002-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-rondo-defense-20180930-story.html?outputType=amp> | <http://www.latimes.com> |
| <http://www.latimes.com/sports/nba/la-sp-lebron-james-20160620-snap-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-20181013-story.html> | <http://www.latimes.com/entertainment/la-et-tiff-2018-toronto-film-festival-updates-htmlstory.html> |
| <https://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-20181013-story.html> | <http://www.latimes.com> |
| [https://www.latimes.com](https://www.latimes.com/la-bio-tania-ganguli-staff.html) | <http://www.latimes.com> |
| http://www.latimes.com | <http://www.latimes.com> |

**2018 World Cup**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com/sports/la-sp-42nd-ryder-cup-pictures-2018-photogallery.html> | <http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-southern-california-2017-htmlstory.html> |
| <https://www.latimes.com/sports/la-lb-816-45174-la-me-ln-mexico-korea-soccer-viewing-20180622-htmlstory.html> | <http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html> |
| <http://www.latimes.com/sports/la-lb-816-45174-la-me-ln-mexico-korea-soccer-viewing-20180622-htmlstory.html> | <http://www.latimes.com/sports/dodgers/la-sp-live-dodgers-brewers-nlcs-htmlstory.html> |
| <http://www.latimes.com/sports/soccer/worldcup/la-sp-mexico-korea-1998-world-cup-2018-world-cup-20180622-story.html> | <http://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html> |
| <http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-b-lgica-detiene-en-seco-a-brasil-1530907910-htmlstory.html> | <https://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html> |
| <http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-empresario-que-fue-secuestrado-en-m-xico-1530960855-htmlstory.html> | [http://www.latimes.com](http://www.latimes.com/la-bio-david-lauter-staff.html) |
| <http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-video-brasil-anota-un-autogol-y-b-lgica-1530901739-htmlstory.html> | <http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html> |
| <http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-video-grave-error-de-muslera-le-regala-1530891077-htmlstory.html> | http://www.latimes.com/politics/la-na-pol-kavanaugh-ford-hearing-htmlstory.html |
| <http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-video-francia-acaricia-las-semifinales-1530889005-htmlstory.html> | <http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html> |
| <http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-aficionados-de-rusia-agradecen-a-1530473387-htmlstory.html> | <http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-get-the-eric-sondheimer-s-latest-1490657918-htmlstory.html> |

**North Korea**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-will-president-trump-bring-up-north-1528706649-htmlstory.html> | <http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html> |
| <http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-trump-becomes-first-u-s-president-to-1528766878-htmlstory.html> | <http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html> |
| <http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-kimchi-flavored-ice-cream-a-small-1528697996-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-swift-north-korean-coverage-of-kim-jong-1528782028-htmlstory.html> | <http://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html> |
| <http://www.latimes.com/world/la-fg-north-korea-nuclear-20180420-story.html> | <http://www.latimes.com/about/la-privacypolicy-20180703-story.html> |
| <http://www.latimes.com/nation/la-na-pol-pompeo-north-korea-20181007-story.html?outputType=amp> | <http://www.latimes.com> |
| <http://www.latimes.com/world/la-fg-north-korea-environment-20171006-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/world/asia/la-fg-north-korea-war-threat-20170925-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-fg-north-korea-foreign-minister-20170923-story.html> | <http://www.latimes.com/travel/la-tr-california-bucket-list-updates-2017-htmlstory.html> |
| <http://www.latimes.com/world/asia/la-fg-trump-north-korea-analysis-20180309-story.html> | <http://www.latimes.com/entertainment/la-et-tiff-2018-toronto-film-festival-updates-htmlstory.html> |

**Hurricane Florence**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <https://www.latimes.com/nation/la-na-hurricane-florence-20180913-story.html> | <http://www.latimes.com/about/la-privacypolicy-20180703-story.html> |
| <http://www.latimes.com/nation/la-na-hurricane-florence-20180913-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-na-hurricane-florence-carolinas-virginia-20180912-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-na-hurricane-florence-live-updates-hurricane-florence-meanders-through-1536975893-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-na-hurricane-florence-20180914-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-lb-838-47191-la-na-pol-trump-hurricane-florence-20180915-htmlstory.html> | <http://www.latimes.com/opinion/editorials/la-ed-scotus-wedding-cake-20180605-story.html> |
| <https://www.latimes.com/nation/la-na-hurricane-florence-live-updates-htmlstory.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-na-hurricane-florence-20180915-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/nation/la-lb-838-47190-la-na-hurricane-florence-20180915-htmlstory.html> | <http://www.latimes.com/entertainment/la-et-tony-awards-2018-winners-htmlstory.html> |
| <http://www.latimes.com/nation/la-lb-838-47192-la-na-florence-new-bern-20180915-htmlstory.html> | <http://www.latimes.com> |

**Paul Allen**

|  |  |
| --- | --- |
| Lucene | PageRank. |
| <http://www.latimes.com/local/obituaries/la-fi-paul-allen-pictures-20181015-photogallery.html> | <http://www.latimes.com> |
| <http://www.latimes.com/local/obituaries/la-fi-tn-paul-allen-obituary-20181015-story.html> | <http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html> |
| <https://www.latimes.com/local/obituaries/la-fi-tn-paul-allen-obituary-20181015-story.html> | <http://www.latimes.com> |
| <https://www.latimes.com/business/realestate/hot-property/la-fi-hotprop-paul-allen-beverly-hills-20180710-story.html> | <http://www.latimes.com/entertainment/la-et-2018-emmys-70th-emmy-awards-live-updates-htmlstory.html> |
| <http://www.latimes.com/business/realestate/hot-property/la-fi-hotprop-paul-allen-beverly-hills-20180710-story.html> | <http://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html> |
| <http://www.latimes.com/local/obituaries/la-fi-tn-paul-allen-obituary-20181015-story.html?outputType=amp> | <http://www.latimes.com> |
| <http://www.latimes.com> | <http://www.latimes.com> |
| <http://www.latimes.com> | <http://www.latimes.com> |
| <http://www.latimes.com/local/obituaries/la-me-maureen-paul-turlish-20180802-story.html> | <http://www.latimes.com> |
| <http://www.latimes.com/local/obituaries/la-na-paul-laxalt-20180806-story.html> | <http://www.latimes.com/projects/la-fi-disney-anaheim-city-council/> |

**overlap graph**

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