

# Indexing the Web Using Solr

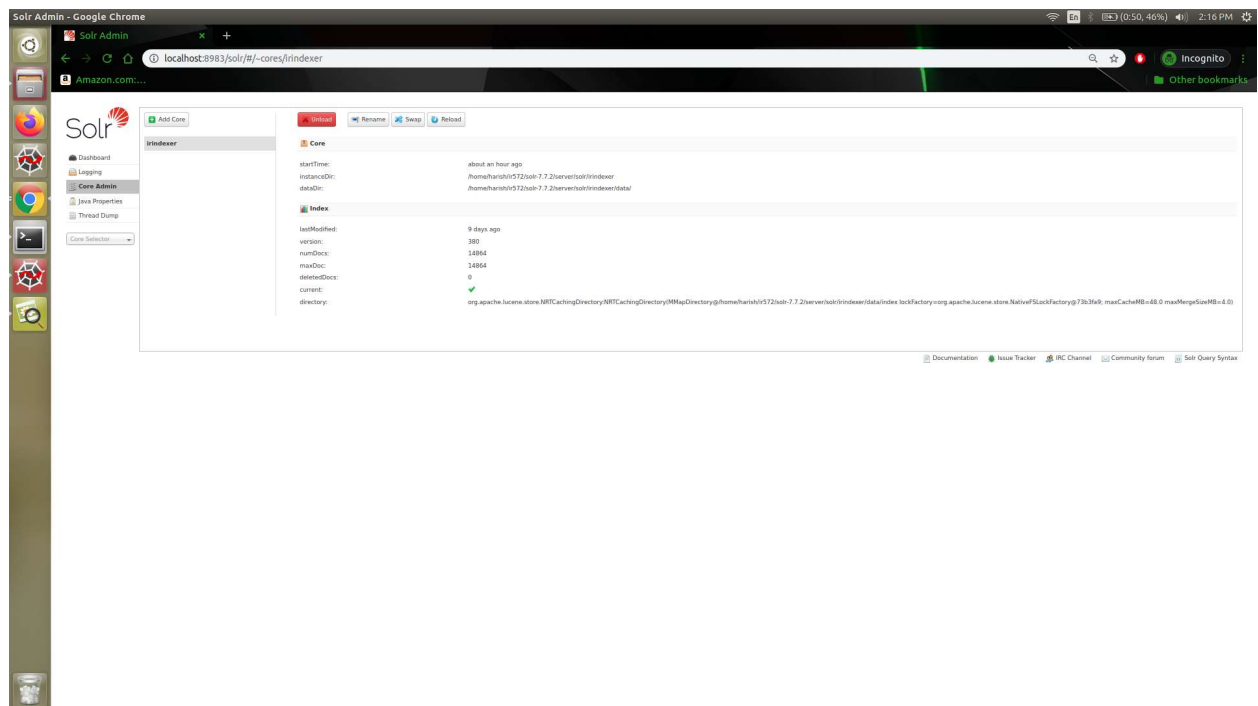
## Steps performed

### Step 1: Creating Apache Solr Core

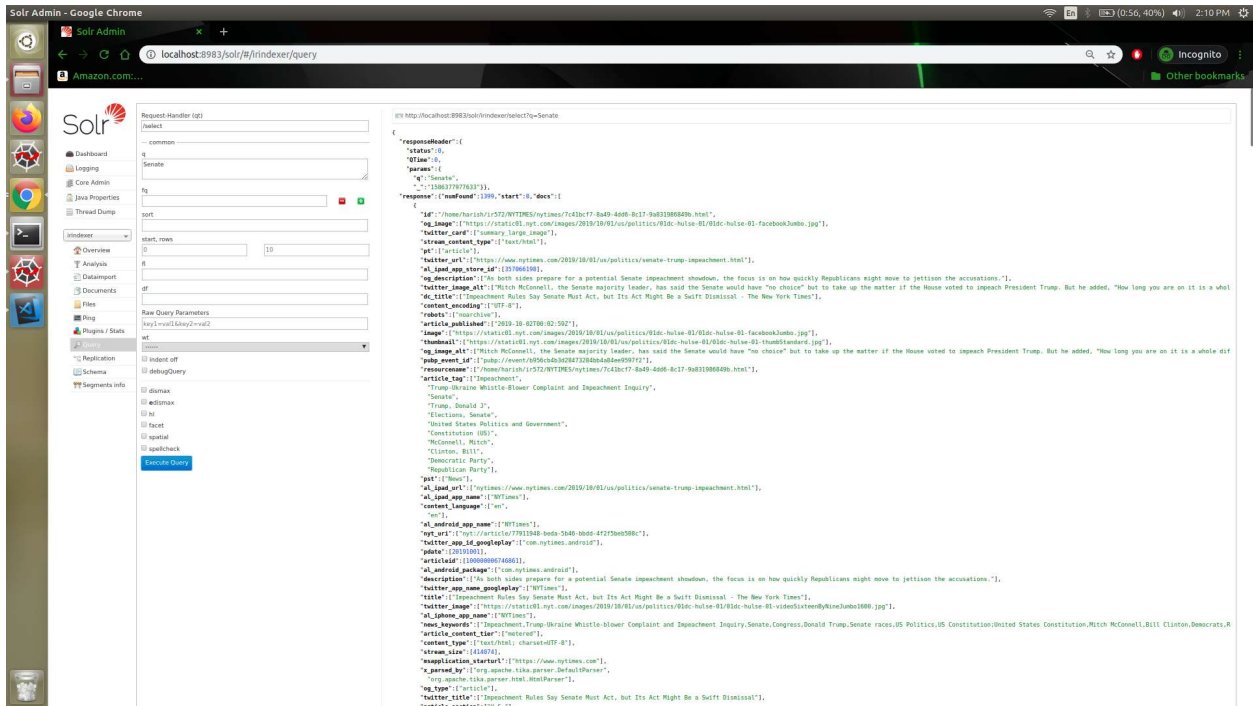
1. Downloaded Apache Solr 7.7.2 and unzipped it.
2. Downloaded the NYTimes folder and unzipped it.
3. Started the Solr Server by running the command **bin/solr start** within the solr-7.7.2 folder.
4. Created a Solr core named **irindexer** by running the command **bin/solr create -c irindexer**
5. Examined the directory structure of the created core by cd'ing to the folder **server/solr**.
6. Edited the **managed-schema** file within the **conf** folder to include the **copyField** element that combines all the fields into **\_text** field.

### Step 2: Indexing HTML pages to Solr

1. Indexed the HTML pages downloaded by running the command **bin/post -c irindexer -filetypes html crawl\_data/** [Here **crawl\_data/** contains the downloaded HTML pages].
2. After completing indexing visited Solr Admin and ran a sample query by selecting **irindexer** core.



3. Ran a sample query using Lucene ranking method in the Solr UI to check if the pages are indexed properly.



### Step 3: Generating External Pagerank values using networkx

1. Downloaded the **Jsoup** library JAR to extract outgoing links from HTML pages in Java.
2. Created two Hashmaps to map FileID->URL and URL->FileID for generating edge lists.
3. For each HTML file in the **crawl\_data/** folder extracted all the outgoing URLs using **Jsoup** library and found the corresponding FileID using Hashmap and added it to a set of tuples.
4. Finally stored the list of edges in a file named **edgeLists.txt**.
5. Created a networkx graph from the edgeLists.txt using the python library **networkx**.
6. Ran the networkx pagerank function over the graph by setting the following parameters: **[alpha=0.85, personalization=None, max\_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None]**
7. Stored the pagerank scores generated for all the files as **external\_PageRankFile.txt**.
8. Added the field in managed-schema indicating this file as external and fieldname as **pageRankFile**.
9. In order to access the file, custom listener tags are set up and the core is reloaded.
10. While querying using pagerank, the parameter **sort = pageRankFile desc** is included.
11. Ran a sample query using PageRank ranking method in the Solr UI to check if the pages are indexed properly.

1. Designed the Client side UI using HTML, CSS, Bootstrap and PHP.
2. For the Query box, included a text box for query, dropdown list for choosing the ranking method (Lucene/PageRank) and a button to submit the query.
3. For the search results, included a table to display the results and added href links to Title and URL using the og\_url data if exists or the sitemap.
4. Upon submitting the query, the client sends a request to the PHP server with the query and additional parameters based on the ranking algorithm specified.
5. The PHP server sends the request using Apache\_Solr\_Service to the Solr Server and retrieves the result and sends it to the client for parsing and displaying the results which includes the fields: id, title, og\_description, og\_url.
6. Additional edge cases for PHP client are included such as handling empty og\_description, og\_url.

1. The given set of queries are run using both the ranking methods and results returned are recorded.
2. UI output flow for single query is captured for documentation and overlap among results are found.

PageRank of a page indicates the significance of a page in terms of the number of incoming links and their relevance to the target page. An incoming link from a page can be interpreted as vote of confidence of that page to the target page that receives the incoming link. Also, if a page P has a high pagerank (based on its incoming links) and it points to page Q, then the pagerank of Q is also higher, thus the pagerank among pages is transmitted in a recursive manner. In this assignment, the incoming links from a set of pages to a page is indicated by a list of edges. We use this edge lists to create a networkx graph and run the pagerank

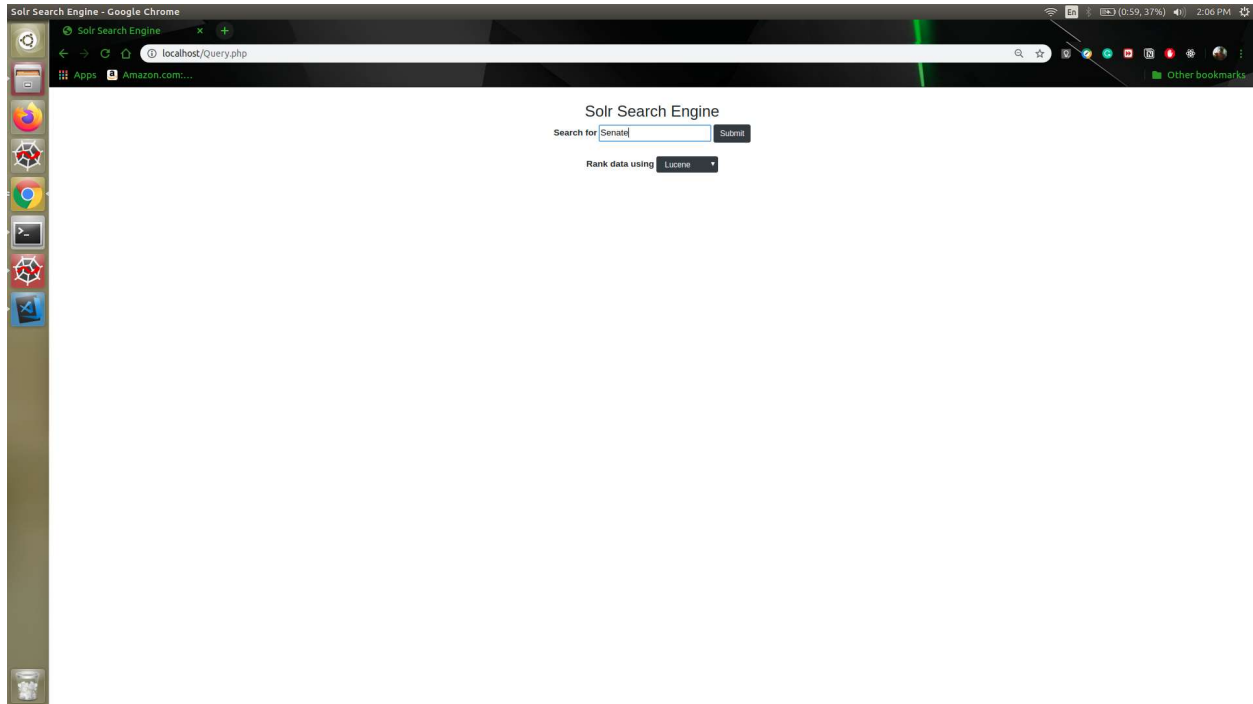
Name: Harish Mohan

USC ID: 1938834211

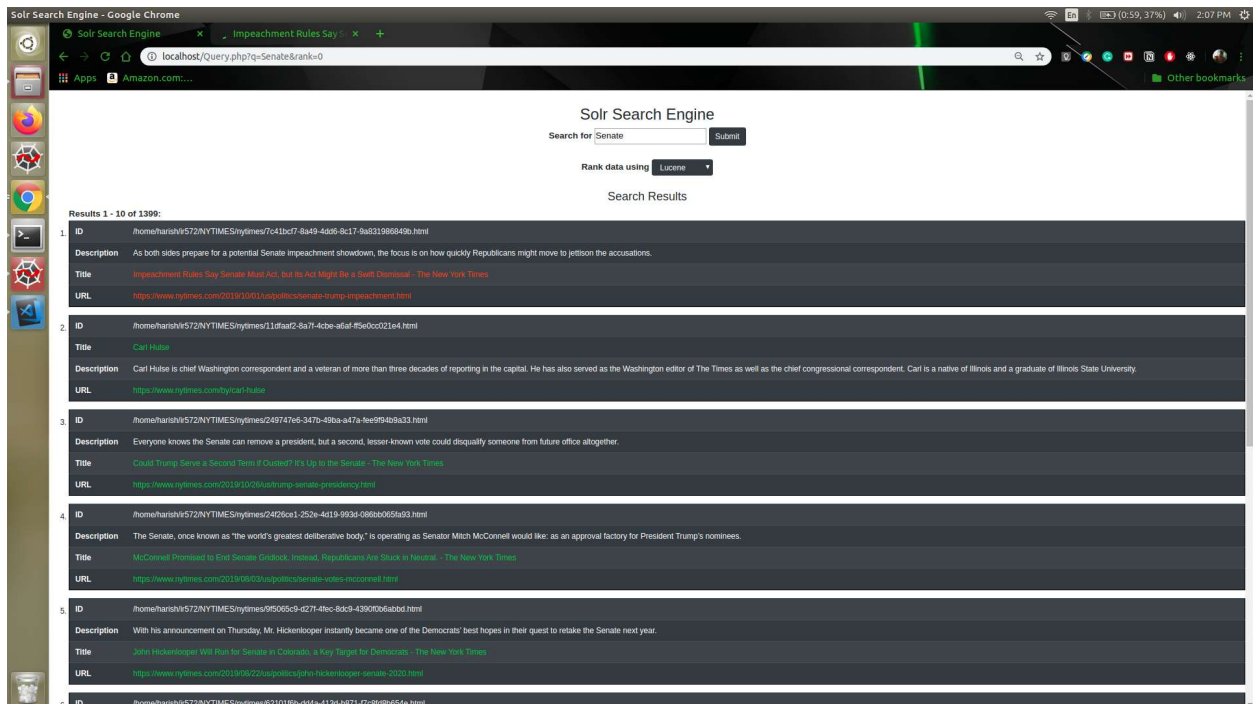
News Site: NY Times

algorithm. Conclusively, if a page has many incoming links or is being pointed to by a page having high pagerank, the target page will have high pagerank. Hence some pages have higher pagerank than the other.

## Query Flow



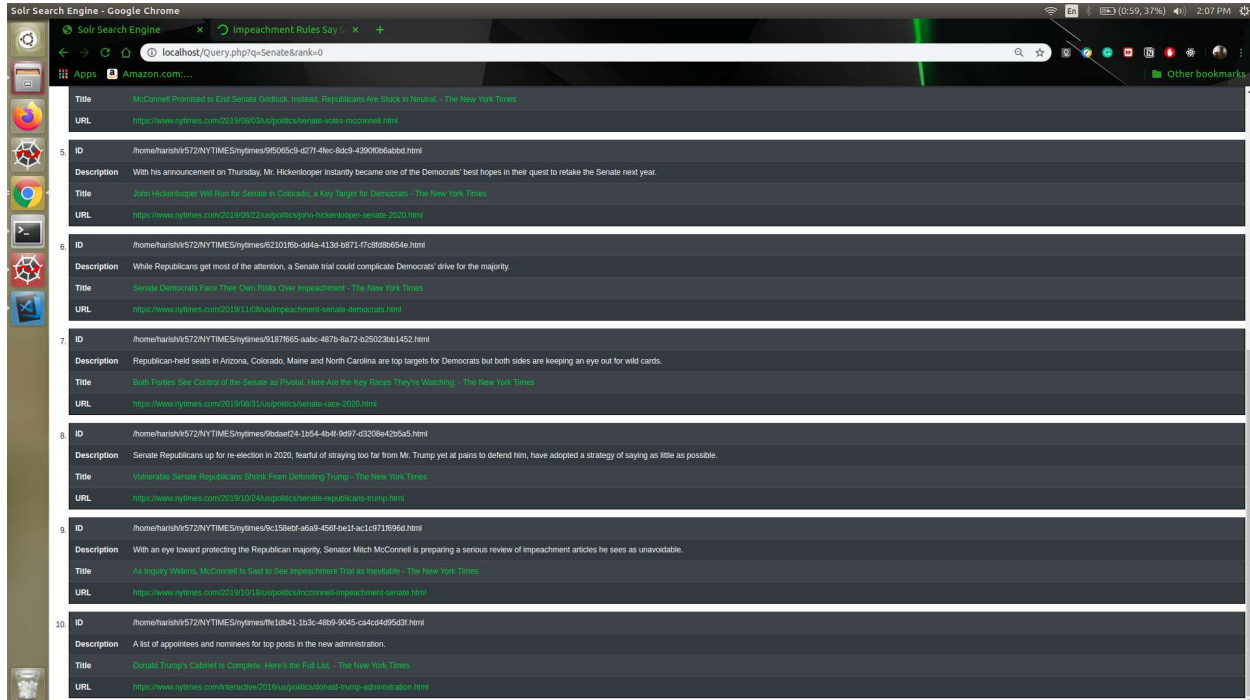
The Search query box is displayed with default ranking method as Lucene.



Name: Harish Mohan

USC ID: 1938834211

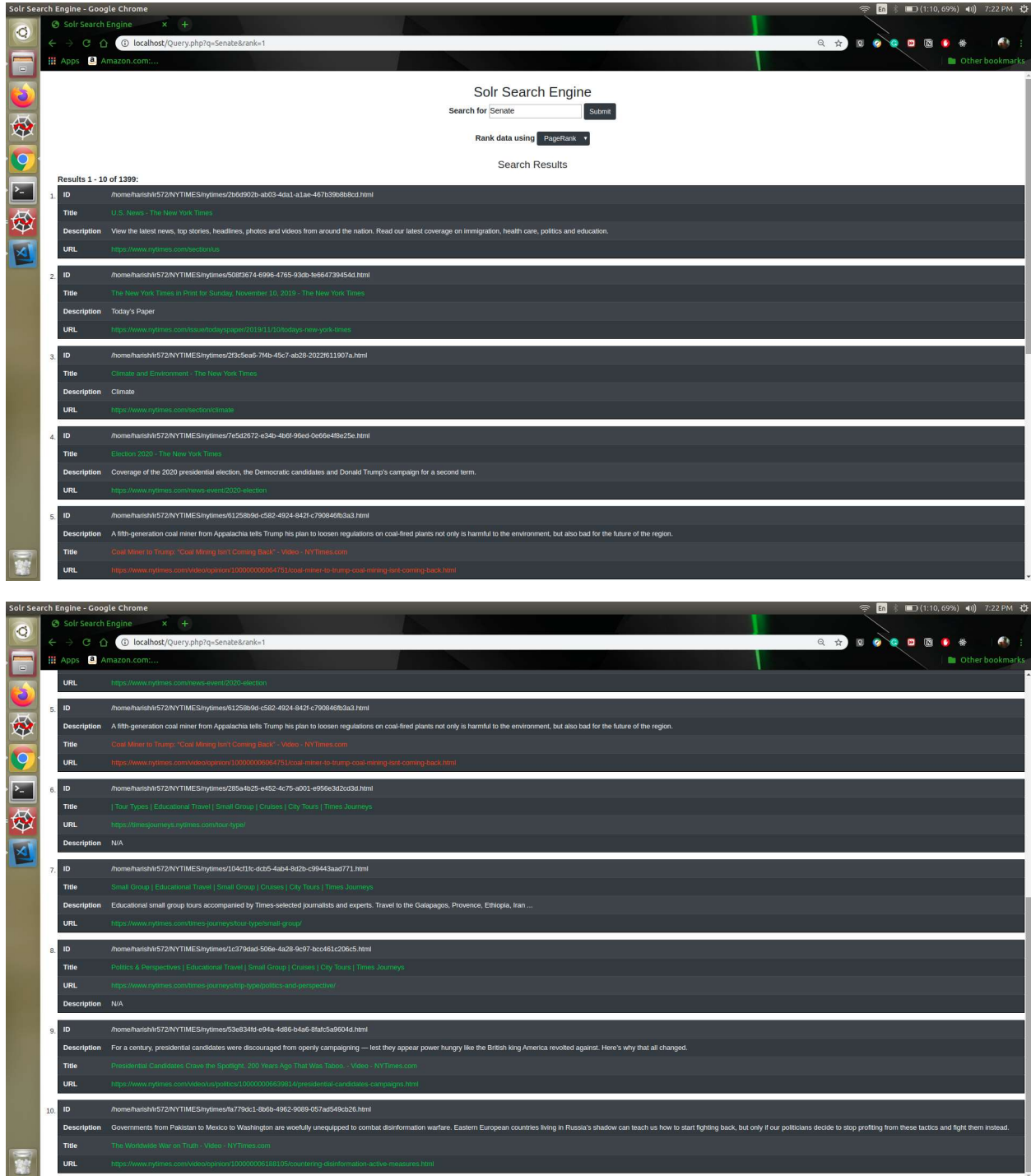
News Site: NY Times



The above 2 images indicate the first ten search results for the query “Senate” ranked by Lucene ranking.

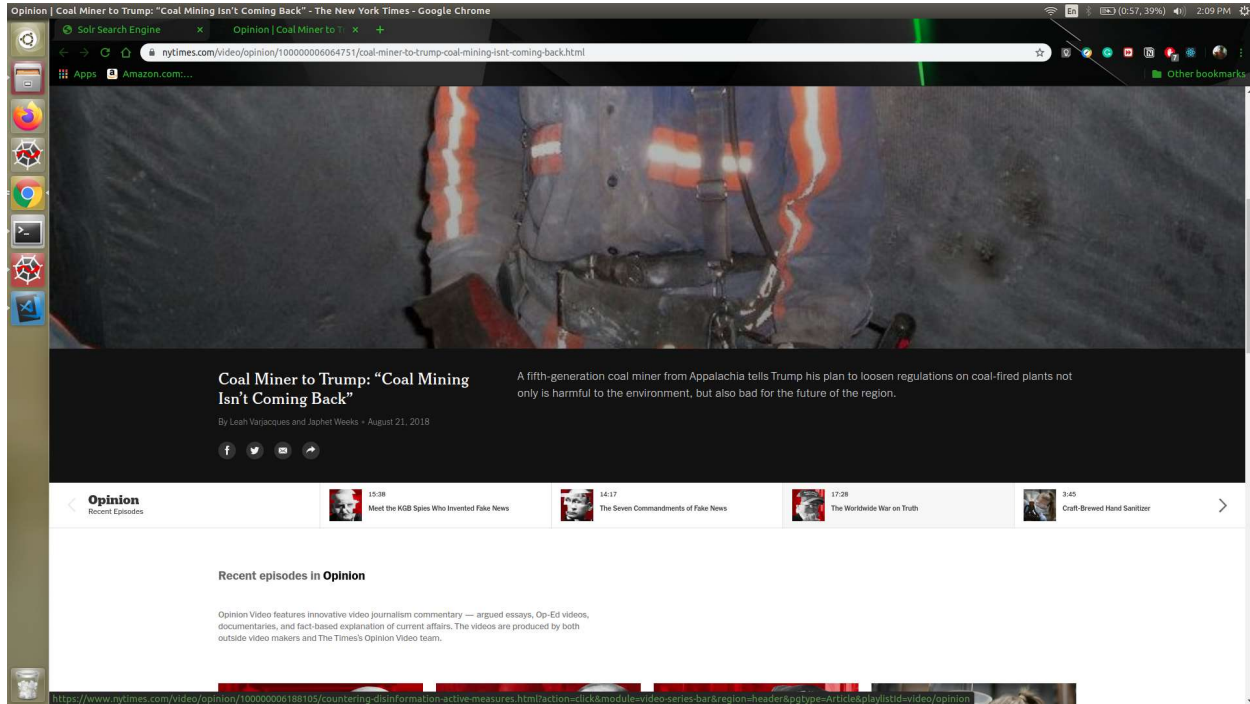


This is the resulting page that loads when we click the first result of the above results.



The above 2 images indicate the first ten search results for the query “Senate” ranked by PageRank ranking.





This is the resulting page that loads when we click the fifth result of the above results.

## URL Results

### Query – Cannes

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html">https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html</a>	<a href="https://www.nytimes.com/2019/11/07/movies/nyc-this-weekend-film-series.html">https://www.nytimes.com/2019/11/07/movies/nyc-this-weekend-film-series.html</a>
<a href="https://www.nytimes.com/2018/04/12/movies/cannes-lineup-netflix-boycott.html">https://www.nytimes.com/2018/04/12/movies/cannes-lineup-netflix-boycott.html</a>	<a href="https://www.nytimes.com/news-event/awards-season">https://www.nytimes.com/news-event/awards-season</a>
<a href="https://www.nytimes.com/2019/05/24/movies/cannes-almodovar-kechiche.html">https://www.nytimes.com/2019/05/24/movies/cannes-almodovar-kechiche.html</a>	<a href="https://www.nytimes.com/interactive/2019/10/15/t-magazine/nick-cave-artist.html">https://www.nytimes.com/interactive/2019/10/15/t-magazine/nick-cave-artist.html</a>
<a href="https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html">https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html</a>	<a href="https://www.nytimes.com/interactive/2019/10/15/t-magazine/nicolas-ghesquiere-louis-vuitton.html">https://www.nytimes.com/interactive/2019/10/15/t-magazine/nicolas-ghesquiere-louis-vuitton.html</a>
<a href="https://www.nytimes.com/2019/05/16/movies/cannes-film-festival-jim-jarmusch.html">https://www.nytimes.com/2019/05/16/movies/cannes-film-festival-jim-jarmusch.html</a>	<a href="https://www.nytimes.com/interactive/2019/10/15/t-magazine/rachel-weisz-acting-movies.html">https://www.nytimes.com/interactive/2019/10/15/t-magazine/rachel-weisz-acting-movies.html</a>
<a href="https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html">https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html</a>	<a href="https://www.nytimes.com/2019/10/30/movies/bong-joon-ho-parasite.html">https://www.nytimes.com/2019/10/30/movies/bong-joon-ho-parasite.html</a>
<a href="https://www.nytimes.com/2019/05/27/movies/cannes-oscars.html">https://www.nytimes.com/2019/05/27/movies/cannes-oscars.html</a>	<a href="https://www.nytimes.com/2019/11/01/movies/taika-waititi.html">https://www.nytimes.com/2019/11/01/movies/taika-waititi.html</a>
<a href="https://www.nytimes.com/2019/05/20/arts/cannes-film-festival-black-director.html">https://www.nytimes.com/2019/05/20/arts/cannes-film-festival-black-director.html</a>	<a href="https://www.nytimes.com/interactive/2019/10/15/t-magazine/shigeru-ban.html">https://www.nytimes.com/interactive/2019/10/15/t-magazine/shigeru-ban.html</a>
<a href="https://www.nytimes.com/2019/05/22/movies/quentin-tarantino-margot-robbie.html">https://www.nytimes.com/2019/05/22/movies/quentin-tarantino-margot-robbie.html</a>	<a href="https://www.nytimes.com/interactive/2019/10/15/t-magazine/the-greats.html">https://www.nytimes.com/interactive/2019/10/15/t-magazine/the-greats.html</a>
<a href="https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html">https://www.nytimes.com/2019/05/22/movies/dicaprio-pitt.html</a>	<a href="https://www.nytimes.com/2019/10/10/movies/parasite-review.html">https://www.nytimes.com/2019/10/10/movies/parasite-review.html</a>

## Query – Congress

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/interactive/2019/us/politics/trump-impeachment-congress-list.html">https://www.nytimes.com/interactive/2019/us/politics/trump-impeachment-congress-list.html</a>	<a href="https://www.nytimes.com">https://www.nytimes.com</a>
<a href="https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html">https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html</a>	<a href="https://www.nytimes.com/section/magazine">https://www.nytimes.com/section/magazine</a>
<a href="https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html">https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html</a>	<a href="https://www.nytimes.com/section/t-magazine">https://www.nytimes.com/section/t-magazine</a>
<a href="https://www.nytimes.com/2019/07/25/us/politics/mueller-impeachment.html">https://www.nytimes.com/2019/07/25/us/politics/mueller-impeachment.html</a>	<a href="https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times">https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times</a>
<a href="https://www.nytimes.com/2019/07/26/us/politics/donald-trump-impeachment.html">https://www.nytimes.com/2019/07/26/us/politics/donald-trump-impeachment.html</a>	<a href="https://www.nytimes.com/section/jobs">https://www.nytimes.com/section/jobs</a>
<a href="https://www.nytimes.com/2018/09/17/us/politics/bloomberg-president-2020-democrat.html">https://www.nytimes.com/2018/09/17/us/politics/bloomberg-president-2020-democrat.html</a>	<a href="https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html">https://www.nytimes.com/video/us/politics/100000006067682/president-trump-impeach-questions-answers-michael-cohen.html</a>
<a href="https://www.nytimes.com/2018/11/21/obituaries/james-billington-dead.html">https://www.nytimes.com/2018/11/21/obituaries/james-billington-dead.html</a>	<a href="https://www.nytimes.com/video/us/politics/100000005524412/paul-manafort-fraud-launders-scandal.html">https://www.nytimes.com/video/us/politics/100000005524412/paul-manafort-fraud-launders-scandal.html</a>
<a href="https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html">https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html</a>	<a href="https://www.nytimes.com/video/us/politics/100000006639814/presidential-candidates-campaigns.html">https://www.nytimes.com/video/us/politics/100000006639814/presidential-candidates-campaigns.html</a>
<a href="https://www.nytimes.com/2019/08/07/us/politics/foreign-aid-freeze-congress.html">https://www.nytimes.com/2019/08/07/us/politics/foreign-aid-freeze-congress.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html">https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html</a>
<a href="https://www.nytimes.com/1973/05/13/archives/constitutional-erwin-ah-thi-the-senate-is-determined-to-recover.html">https://www.nytimes.com/1973/05/13/archives/constitutional-erwin-ah-thi-the-senate-is-determined-to-recover.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html">https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html</a>

## Query – Democrats

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/2018/09/04/us/politics/ayanna-pressley-massachusetts.html">https://www.nytimes.com/2018/09/04/us/politics/ayanna-pressley-massachusetts.html</a>	<a href="https://www.nytimes.com">https://www.nytimes.com</a>
<a href="https://www.nytimes.com/2018/11/07/us/politics/election-history-firsts-blackburn-pressley.html">https://www.nytimes.com/2018/11/07/us/politics/election-history-firsts-blackburn-pressley.html</a>	<a href="https://www.nytimes.com/video">https://www.nytimes.com/video</a>
<a href="https://www.nytimes.com/2019/10/16/us/politics/democrats-impeachment.html">https://www.nytimes.com/2019/10/16/us/politics/democrats-impeachment.html</a>	<a href="https://www.nytimes.com/section/magazine">https://www.nytimes.com/section/magazine</a>
<a href="https://www.nytimes.com/2019/09/29/us/politics/pelosi-trump-impeachment.html">https://www.nytimes.com/2019/09/29/us/politics/pelosi-trump-impeachment.html</a>	<a href="https://www.nytimes.com/section/politics">https://www.nytimes.com/section/politics</a>
<a href="https://www.nytimes.com/2018/01/23/us/politics/impeach-trump-democrats-tom-steyer.html">https://www.nytimes.com/2018/01/23/us/politics/impeach-trump-democrats-tom-steyer.html</a>	<a href="https://www.nytimes.com/section/us">https://www.nytimes.com/section/us</a>
<a href="https://www.nytimes.com/2019/09/24/us/politics/trump-democrats.html">https://www.nytimes.com/2019/09/24/us/politics/trump-democrats.html</a>	<a href="https://www.nytimes.com/section/opinion">https://www.nytimes.com/section/opinion</a>
<a href="https://www.nytimes.com/2019/10/28/us/politics/trump-impeachment-subpoena-house.html">https://www.nytimes.com/2019/10/28/us/politics/trump-impeachment-subpoena-house.html</a>	<a href="https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times">https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times</a>
<a href="https://www.nytimes.com/issue/magazine/2019/06/25/the-61619-issue">https://www.nytimes.com/issue/magazine/2019/06/25/the-61619-issue</a>	<a href="https://www.nytimes.com/section/opinion/columnists">https://www.nytimes.com/section/opinion/columnists</a>
<a href="https://www.nytimes.com/2019/11/05/opinion/trump-democrats-2020.html">https://www.nytimes.com/2019/11/05/opinion/trump-democrats-2020.html</a>	<a href="https://www.nytimes.com/section/upshot">https://www.nytimes.com/section/upshot</a>
<a href="https://www.nytimes.com/2019/09/11/opinion/black-voters-democrats-2020-election.html">https://www.nytimes.com/2019/09/11/opinion/black-voters-democrats-2020-election.html</a>	<a href="https://www.nytimes.com/news-event/2020-election">https://www.nytimes.com/news-event/2020-election</a>

## Query – Patriot Movement

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/2019/08/14/opinion/joe-walsh-trump-primary.html">https://www.nytimes.com/2019/08/14/opinion/joe-walsh-trump-primary.html</a>	<a href="https://www.nytimes.com">https://www.nytimes.com</a>
<a href="https://www.nytimes.com/1976/08/27/archives/lewis-michaux-92-dies-ran-bookstore-in-harlem.html">https://www.nytimes.com/1976/08/27/archives/lewis-michaux-92-dies-ran-bookstore-in-harlem.html</a>	<a href="https://www.nytimes.com/video">https://www.nytimes.com/video</a>



<a href="https://www.nytimes.com/2019/05/15/us/politics/us-iraq-embassy-evacuation.html">https://www.nytimes.com/2019/05/15/us/politics/us-iraq-embassy-evacuation.html</a>	<a href="https://www.nytimes.com/video/opinion">https://www.nytimes.com/video/opinion</a>
<a href="https://www.nytimes.com/2019/01/07/obituaries/moshe-arens-dead.html">https://www.nytimes.com/2019/01/07/obituaries/moshe-arens-dead.html</a>	<a href="https://www.nytimes.com/section/t-magazine">https://www.nytimes.com/section/t-magazine</a>
<a href="https://www.nytimes.com/2018/12/29/obituaries/shehu-shagari-dead.html">https://www.nytimes.com/2018/12/29/obituaries/shehu-shagari-dead.html</a>	<a href="https://www.nytimes.com/section/opinion">https://www.nytimes.com/section/opinion</a>
<a href="https://www.nytimes.com/2019/10/27/obituaries/john-conyers-jr-dead.html">https://www.nytimes.com/2019/10/27/obituaries/john-conyers-jr-dead.html</a>	<a href="https://www.nytimes.com/section/technology">https://www.nytimes.com/section/technology</a>
<a href="https://www.nytimes.com/2019/10/22/sports/football/jets-patriots.html">https://www.nytimes.com/2019/10/22/sports/football/jets-patriots.html</a>	<a href="https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times">https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times</a>
<a href="https://www.nytimes.com/by/jaclyn-peiser">https://www.nytimes.com/by/jaclyn-peiser</a>	<a href="https://www.nytimes.com/section/lens">https://www.nytimes.com/section/lens</a>
<a href="https://www.nytimes.com/2019/09/23/opinion/democrats-republicans.html">https://www.nytimes.com/2019/09/23/opinion/democrats-republicans.html</a>	<a href="https://www.nytimes.com/video/opinion/10000006525694/stonewall-the-making-of-a-monument.html">https://www.nytimes.com/video/opinion/10000006525694/stonewall-the-making-of-a-monument.html</a>
<a href="https://www.nytimes.com/2019/10/30/opinion/alexander-vindman-trump-ukraine.html">https://www.nytimes.com/2019/10/30/opinion/alexander-vindman-trump-ukraine.html</a>	<a href="https://www.nytimes.com/2019/11/01/arts/television/new-to-stream-netflix.html">https://www.nytimes.com/2019/11/01/arts/television/new-to-stream-netflix.html</a>

## Query – Republicans

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/by/catie-edmondson">https://www.nytimes.com/by/catie-edmondson</a>	<a href="https://www.nytimes.com">https://www.nytimes.com</a>
<a href="https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html">https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html</a>	<a href="https://www.nytimes.com/section/politics">https://www.nytimes.com/section/politics</a>
<a href="https://www.nytimes.com/interactive/2019/08/14/magazine/republicans-racism-african-americans.html">https://www.nytimes.com/interactive/2019/08/14/magazine/republicans-racism-african-americans.html</a>	<a href="https://www.nytimes.com/section/us">https://www.nytimes.com/section/us</a>
<a href="https://www.nytimes.com/2019/10/31/us/politics/house-impeachment-partisan-vote.html">https://www.nytimes.com/2019/10/31/us/politics/house-impeachment-partisan-vote.html</a>	<a href="https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times">https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times</a>
<a href="https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html">https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html</a>	<a href="https://www.nytimes.com/section/upshot">https://www.nytimes.com/section/upshot</a>
<a href="https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html">https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html</a>	<a href="https://www.nytimes.com/news-event/2020-election">https://www.nytimes.com/news-event/2020-election</a>
<a href="https://www.nytimes.com/2019/10/04/us/politics/republicans-trump-defense.html">https://www.nytimes.com/2019/10/04/us/politics/republicans-trump-defense.html</a>	<a href="https://www.nytimes.com/video/us/politics/10000006067682/president-trump-impeach-questions-answers-michael-cohen.html">https://www.nytimes.com/video/us/politics/10000006067682/president-trump-impeach-questions-answers-michael-cohen.html</a>
<a href="https://www.nytimes.com/2019/11/09/us/politics/republican-strategy-impeachment-trump.html">https://www.nytimes.com/2019/11/09/us/politics/republican-strategy-impeachment-trump.html</a>	<a href="https://www.nytimes.com/video/us/elections/10000006810114/kentucky-election-bevin-beshear.html">https://www.nytimes.com/video/us/elections/10000006810114/kentucky-election-bevin-beshear.html</a>
<a href="https://www.nytimes.com/2019/09/30/opinion/republicans-trump-impeachment.html">https://www.nytimes.com/2019/09/30/opinion/republicans-trump-impeachment.html</a>	<a href="https://www.nytimes.com/video/us/politics/10000006812485/hillary-clinton-medicare.html">https://www.nytimes.com/video/us/politics/10000006812485/hillary-clinton-medicare.html</a>
<a href="https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html">https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html</a>	<a href="https://www.nytimes.com/2019/11/09/us/politics/virginia-elections-democrats-republicans.html">https://www.nytimes.com/2019/11/09/us/politics/virginia-elections-democrats-republicans.html</a>

## Query – Senate

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/2019/10/01/us/politics/senate-trump-impeachment.html">https://www.nytimes.com/2019/10/01/us/politics/senate-trump-impeachment.html</a>	<a href="https://www.nytimes.com/section/us">https://www.nytimes.com/section/us</a>
<a href="https://www.nytimes.com/by/carl-hulse">https://www.nytimes.com/by/carl-hulse</a>	<a href="https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times">https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times</a>
<a href="https://www.nytimes.com/2019/10/26/us/trump-senate-presidency.html">https://www.nytimes.com/2019/10/26/us/trump-senate-presidency.html</a>	<a href="https://www.nytimes.com/section/climate">https://www.nytimes.com/section/climate</a>
<a href="https://www.nytimes.com/2019/08/03/us/politics/senate-votes-mcconnell.html">https://www.nytimes.com/2019/08/03/us/politics/senate-votes-mcconnell.html</a>	<a href="https://www.nytimes.com/news-event/2020-election">https://www.nytimes.com/news-event/2020-election</a>
<a href="https://www.nytimes.com/2019/08/22/us/politics/john-hickenlooper-senate-2020.html">https://www.nytimes.com/2019/08/22/us/politics/john-hickenlooper-senate-2020.html</a>	<a href="https://www.nytimes.com/video/opinion/10000006064751/coal-miner-to-trump-coal-mining-isnt-coming-back.html">https://www.nytimes.com/video/opinion/10000006064751/coal-miner-to-trump-coal-mining-isnt-coming-back.html</a>
<a href="https://www.nytimes.com/2019/11/08/us/impeachment-senate-democrats.html">https://www.nytimes.com/2019/11/08/us/impeachment-senate-democrats.html</a>	<a href="https://timesjourneys.nytimes.com/tour-type/">https://timesjourneys.nytimes.com/tour-type/</a>
<a href="https://www.nytimes.com/2019/08/31/us/politics/senate-race-2020.html">https://www.nytimes.com/2019/08/31/us/politics/senate-race-2020.html</a>	<a href="https://www.nytimes.com/times-journeys/tour-type/small-group/">https://www.nytimes.com/times-journeys/tour-type/small-group/</a>

<a href="https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html">https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html</a>	<a href="https://www.nytimes.com/times-journeys/trip-type/politics-and-perspective/">https://www.nytimes.com/times-journeys/trip-type/politics-and-perspective/</a>
<a href="https://www.nytimes.com/2019/10/18/us/politics/mcconnell-impeachment-senate.html">https://www.nytimes.com/2019/10/18/us/politics/mcconnell-impeachment-senate.html</a>	<a href="https://www.nytimes.com/video/us/politics/100000006639814/presidential-candidates-campaigns.html">https://www.nytimes.com/video/us/politics/100000006639814/presidential-candidates-campaigns.html</a>
<a href="https://www.nytimes.com/interactive/2016/us/politics/donald-trump-administration.html">https://www.nytimes.com/interactive/2016/us/politics/donald-trump-administration.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html">https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html</a>

## Query – Olympics 2020

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/2019/10/01/travel/what-you-need-to-know-the-tokyo-2020-olympics.html">https://www.nytimes.com/2019/10/01/travel/what-you-need-to-know-the-tokyo-2020-olympics.html</a>	<a href="https://www.nytimes.com">https://www.nytimes.com</a>
<a href="https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html">https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html</a>	<a href="https://www.nytimes.com/video">https://www.nytimes.com/video</a>
<a href="https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html">https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html</a>	<a href="https://www.nytimes.com/section/movies">https://www.nytimes.com/section/movies</a>
<a href="https://www.nytimes.com/2017/12/05/sports/olympics/ioc-russia-winter-olympics.html">https://www.nytimes.com/2017/12/05/sports/olympics/ioc-russia-winter-olympics.html</a>	<a href="https://www.nytimes.com/section/arts/television">https://www.nytimes.com/section/arts/television</a>
<a href="https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html">https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html</a>	<a href="https://www.nytimes.com/video/opinion">https://www.nytimes.com/video/opinion</a>
<a href="https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html">https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html</a>	<a href="https://www.nytimes.com/es/">https://www.nytimes.com/es/</a>
<a href="https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html">https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html</a>	<a href="https://www.nytimes.com/section/magazine">https://www.nytimes.com/section/magazine</a>
<a href="https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html">https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html</a>	<a href="https://www.nytimes.com/section/realestate">https://www.nytimes.com/section/realestate</a>
<a href="https://www.nytimes.com/2018/04/23/sports/breakdancing-olympics.html">https://www.nytimes.com/2018/04/23/sports/breakdancing-olympics.html</a>	<a href="https://www.nytimes.com/section/politics">https://www.nytimes.com/section/politics</a>
<a href="https://www.nytimes.com/video/sports/football/100000005827420/nfl-cheerleaders-and-the-gender-gap.html">https://www.nytimes.com/video/sports/football/100000005827420/nfl-cheerleaders-and-the-gender-gap.html</a>	<a href="https://www.nytimes.com/section/obituaries">https://www.nytimes.com/section/obituaries</a>

## Query – Stock

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/interactive/2017/10/04/us/bump-stock-las-vegas-gun.html">https://www.nytimes.com/interactive/2017/10/04/us/bump-stock-las-vegas-gun.html</a>	<a href="https://www.nytimes.com/section/world">https://www.nytimes.com/section/world</a>
<a href="https://www.nytimes.com/2019/05/02/business/tesla-stock-fundraising.html">https://www.nytimes.com/2019/05/02/business/tesla-stock-fundraising.html</a>	<a href="https://www.nytimes.com/section/business">https://www.nytimes.com/section/business</a>
<a href="https://www.nytimes.com/2018/12/16/technology/tech-workers-company-stock-shareholder-activism.html">https://www.nytimes.com/2018/12/16/technology/tech-workers-company-stock-shareholder-activism.html</a>	<a href="https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times">https://www.nytimes.com/issue/todayspaper/2019/11/10/todays-new-york-times</a>
<a href="https://www.nytimes.com/2019/05/10/technology/uber-stock-price-ipo.html">https://www.nytimes.com/2019/05/10/technology/uber-stock-price-ipo.html</a>	<a href="https://www.nytimes.com/section/automobiles">https://www.nytimes.com/section/automobiles</a>
<a href="https://www.nytimes.com/2019/11/04/business/under-armour-stock-investigation.html">https://www.nytimes.com/2019/11/04/business/under-armour-stock-investigation.html</a>	<a href="https://www.nytimes.com/section/upshot">https://www.nytimes.com/section/upshot</a>
<a href="https://www.nytimes.com/2019/09/20/technology/airbnb-employees-ipo-payouts.html">https://www.nytimes.com/2019/09/20/technology/airbnb-employees-ipo-payouts.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006616246/the-king-of-fish-and-chips.html">https://www.nytimes.com/video/opinion/100000006616246/the-king-of-fish-and-chips.html</a>
<a href="https://www.nytimes.com/2019/10/24/style/rebag-clair-handbag-stock-market.html">https://www.nytimes.com/2019/10/24/style/rebag-clair-handbag-stock-market.html</a>	<a href="https://www.nytimes.com/section/fashion">https://www.nytimes.com/section/fashion</a>
<a href="https://www.nytimes.com/2019/05/30/technology/uber-stock-earnings.html">https://www.nytimes.com/2019/05/30/technology/uber-stock-earnings.html</a>	<a href="https://www.nytimes.com/times-journeys/travel/egypt/">https://www.nytimes.com/times-journeys/travel/egypt/</a>
<a href="https://www.nytimes.com/2018/11/12/obituaries/david-pearson-nascar-dead.html">https://www.nytimes.com/2018/11/12/obituaries/david-pearson-nascar-dead.html</a>	<a href="https://www.nytimes.com/2019/11/09/opinion/sunday/stripping-sex-work.html">https://www.nytimes.com/2019/11/09/opinion/sunday/stripping-sex-work.html</a>
<a href="https://www.nytimes.com/2019/04/03/business/tesla-sales-elon-musk.html">https://www.nytimes.com/2019/04/03/business/tesla-sales-elon-musk.html</a>	<a href="https://www.nytimes.com/video/movies/100000006675663/brittany-runs-a-marathon-scene.html">https://www.nytimes.com/video/movies/100000006675663/brittany-runs-a-marathon-scene.html</a>

## Query – Virus

Lucene Results	Pagerank Results
<a href="https://www.nytimes.com/2019/11/09/science/seals-distemper.html">https://www.nytimes.com/2019/11/09/science/seals-distemper.html</a>	<a href="https://www.nytimes.com/es/">https://www.nytimes.com/es/</a>
<a href="https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html">https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html</a>	<a href="https://www.nytimes.com/section/us">https://www.nytimes.com/section/us</a>
<a href="https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html">https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html</a>	<a href="https://www.nytimes.com/section/world">https://www.nytimes.com/section/world</a>
<a href="https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html">https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html</a>	<a href="https://www.nytimes.com/section/science">https://www.nytimes.com/section/science</a>
<a href="https://www.nytimes.com/by/donald-g-mcneil-jr">https://www.nytimes.com/by/donald-g-mcneil-jr</a>	<a href="https://www.nytimes.com/section/health">https://www.nytimes.com/section/health</a>
<a href="https://www.nytimes.com/by/tariro-mzezewa">https://www.nytimes.com/by/tariro-mzezewa</a>	<a href="https://www.nytimes.com/section/climate">https://www.nytimes.com/section/climate</a>
<a href="https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html">https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html">https://www.nytimes.com/video/opinion/100000006210828/russia-disinformation-fake-news.html</a>
<a href="https://www.nytimes.com/by/denise-grady">https://www.nytimes.com/by/denise-grady</a>	<a href="https://www.nytimes.com/video/opinion/100000006188102/what-is-pizzagate.html">https://www.nytimes.com/video/opinion/100000006188102/what-is-pizzagate.html</a>
<a href="https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html">https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html">https://www.nytimes.com/video/opinion/100000006188105/countering-disinformation-active-measures.html</a>
<a href="https://www.nytimes.com/2019/11/06/health/crispr-cancer-leukemia.html">https://www.nytimes.com/2019/11/06/health/crispr-cancer-leukemia.html</a>	<a href="https://www.nytimes.com/video/opinion/100000006188102/what-is-pizzagate.html">https://www.nytimes.com/video/opinion/100000006188102/what-is-pizzagate.html</a>

## Overlap

There are no overlaps for any of the queries for the first 10 search results.