

CSCI 572 Information Retrieval and Web Search Engines

Homework 4: Indexing the Web using Solar

Name : KE LI
USC ID : 3444143924

● Steps:

1. Install Solr:

a. Solr installation(macOS):

i. Downloaded the Solr Binary release as a zip file from

<https://www.apache.org/dyn/closer.lua/lucene/solr/7.7.2/solr-7.7.2.zip>

ii. Begin by unzipping the Solr release and changing your working directory to the subdirectory where Solr was installed.

```
~$ unzip -q solr-8.5.0.zip
```

b. Downloaded the NY times directory obtained from a Google drive

c. Indexing HTML files in Solr:

i. Start the Solr server, cd into solr-7.x.x folder and enter the command: `bin/solr start`

Waiting up to 180 seconds to see Solr running on port 8983 [\
Started Solr server on port 8983 (pid=30591). Happy searching!

ii. Create a new core, in this tutorial the core name is myexample using the command:

```
bin/solr create -c myexample.
```

iii. Go to the conf folder of the core “myexample”. Set the managed-schema file.

Uncomment the following line - `<copyField source="*" dest="_text_"/>`

iv. Executed `bin/post -c myexample -filetypes html nytimes/` to index html files in

the nytimes folder

POSTing file a1020f1f6-6e85-4b6a-851e-b2f5a4/e291b.html (text/html) to [base]/ext

POSTing file 834b2c6c-c175-4e6a-b1b1-99c13c3fccac.html (text/html) to [base]/ext

16040 files indexed.

COMMITting Solr index changes to <http://localhost:8983/solr/myexample/update...>

Time spent: 0:06:41.731

v. Accessed <http://localhost:8983/> and select the core “myexample” from the dropdown

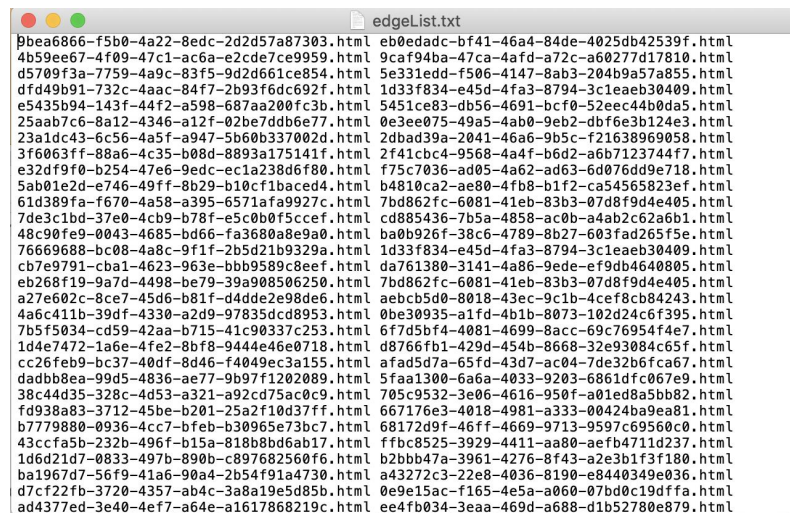
The screenshot shows the Solr Admin web interface in a browser. The address bar indicates the URL is `localhost:8983/solr/#/myexample/core-overview`. The interface includes a sidebar with navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, myexample (selected), Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query, Replication, Schema, and Segments info. The main content area displays statistics for the 'myexample' core, including 'Last Modified: about 20 hours ago', 'Num Docs: 12858', 'Max Doc: 12858', 'Heap Memory Usage: -1', 'Deleted Docs: 0', 'Version: 170', 'Segment Count: 17', and 'Current: ✓'. It also shows a table for 'Replication (Master)' with columns for Version, Gen, and Size, listing 'Master (Searching)' and 'Master (Replicable)'. On the right, the 'Instance' section shows configuration details like CWD, Instance, Data, Index, and Impl. A 'Healthcheck' section at the bottom right indicates 'Ping request handler is not configured with a healthcheck file.' At the very bottom, there are links to Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax.

d. Set up Apache web server, php environment:

- i. Macbook has pre-installed Apache server, so start service of Apache server : `sudo apachectl start`
- ii. open the httpd.conf file: Delete the # from the line:
`#LoadModule php7_module libexec/apache2/libphp7.so`
- iii. Now type the command below to restart your apache server and press enter:
`sudo apachectl restart`
- iv. Download solr-php-client, and put the Apache directory to current directory(same as php file: /Library/WebServer/Documents/). Change the unix include path in php.ini to “.”
- v. Write a php program and you can use <http://localhost/compare.php> to see.

2. Create edgeList.txt and external_pageRankFile.txt:

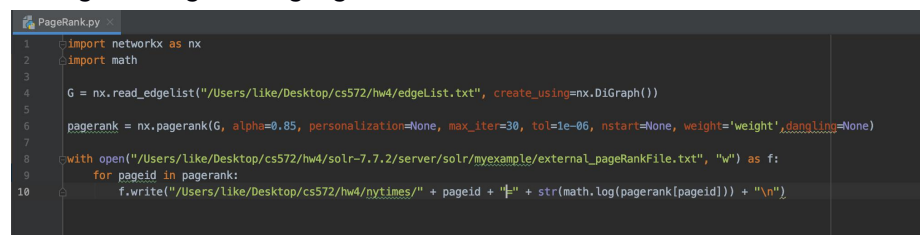
- a. Write a program in Java using **JSoup Library** and HashMaps to compute the EdgeList (edgeList.txt) from the provided html pages.



```
edgeList.txt
b9ea6866-f5b0-4a22-8edc-2d2d57a87303.html
4b59ee67-4f09-47c1-ac6a-e2cde7ce9959.html
d5709f3a-7759-4a9c-83f5-9d2d661ce854.html
dfd49b91-732c-4aac-84f7-2b93f6dc692f.html
e5435b94-143f-44f2-a598-687aa200fc3b.html
25aab7c6-8a12-4346-a12f-02be7ddb6e77.html
23a1dc43-6c56-4a5f-a947-5b60b337002d.html
3f6063ff-88a6-4c35-b08d-8893a175141f.html
e32df9f0-b254-47e6-9edc-ec1a238d6f80.html
5ab01e2d-e746-49ff-8b29-b10cf1baced4.html
61d389fa-f670-4a58-a395-6571afa9927c.html
7de3c1bd-37e0-4cb9-b78f-e5c0b0f5ccef.html
48c90fe9-0043-4685-bd66-fa3680a8e9a0.html
76669688-bc08-4a8c-9f1f-2b5d21b9329a.html
cb7e9791-cba1-4623-963e-bbb9589c8eef.html
eb268f19-9a7d-4498-be79-39a908506250.html
a27e602c-8ce7-45d6-b81f-d4dde2e98de6.html
4a6c411b-39df-4330-a209-97835dcd8953.html
7b5f5034-cd59-42aa-b715-41c90337c253.html
1d4e7472-1a6e-4fe2-8bf8-9444e46e0718.html
cc26feb9-bc37-40df-8d46-f4049ec3a155.html
dadbb8ea-99d5-4836-ae77-9b97f1202089.html
38c44d35-328c-4d53-a321-a92cd75ac0c9.html
fd938a83-3712-45be-b201-25a2f10d37ff.html
b7779880-0936-4cc7-bfeb-b30965e73bc7.html
43ccfa5b-232b-496f-b15a-818b8bd6ab17.html
1d6d21d7-0833-497b-890b-c897682560f6.html
ba1967f7-56f9-41a6-90a4-2b54f91a4730.html
d7cf22fb-3720-4357-ab4c-3a8a19e5d85b.html
ad4377ed-3e40-4ef7-a64e-a1617868219c.html
eb0edadc-bf41-46a4-84de-4025db42539f.html
9caf94ba-47ca-4afd-a72c-a60277d17810.html
5e331edd-f506-4147-8ab3-204b9a57a855.html
1d33f834-e45d-4fa3-8794-3c1eae30409.html
5451ce83-d5b6-4691-bcf0-52ecc44b0da5.html
0e3ee075-49a5-4ab0-9eb2-dbf6e3b124e3.html
2dbad39a-2041-46a6-9b5c-f21638969058.html
2f41cbc4-9568-4a4f-b6d2-a6b7123744f7.html
f75c7036-ad05-4a62-ad63-6d076dd9e718.html
b4810ca2-ae80-4fb8-b1f2-ca54565823ef.html
7bd862fc-6081-41eb-83b3-07d8f9d4e405.html
cd885436-7b5a-4858-ac0b-a4ab2c62a6b1.html
ba0b926f-38c6-4789-8b27-603fad265f5e.html
1d33f834-e45d-4fa3-8794-3c1eae30409.html
da761380-3141-4a86-9ede-ef9db4640805.html
7bd862fc-6081-41eb-83b3-07d8f9d4e405.html
aebcb5d0-8018-43ec-9c1b-4cef8cb84243.html
0be30935-a1fd-4b1b-8073-102d24c6f395.html
6f7d5bf4-4081-4699-8acc-69c76954f4e7.html
d8766fb1-429d-454b-8668-32e93084c65f.html
afad5d7a-65fd-43d7-ac04-7de32b6fca67.html
5faa1300-6a6a-4033-9203-6861dfc067e9.html
705c9532-3e06-4616-950f-a01ed8a5bb82.html
667176e3-4018-4981-a333-00424ba9ea81.html
68172d9f-46ff-4669-9713-9597c69560c0.html
ffbc8525-3929-4411-aa80-aeafb4711d237.html
b2bbb47a-3961-4276-8f43-a2e3b1f3f180.html
a43272c3-22e8-4036-8190-e8440349e036.html
0e9e15ac-f165-4e5a-a060-07bd0c19dfffa.html
ee4fb034-3eaa-469d-a688-d1b52780e879.html
```

b. Write a python code using NetworkX library to compute page rank: pageRankFile.txt

- i. `pip install networkx`
- ii. Compute the page rank given a DiGraph(edgeList.txt), set the parameters: alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None



```
PageRank.py
1 import networkx as nx
2 import math
3
4 G = nx.read_edgelist("/Users/like/Desktop/cs572/hw4/edgeList.txt", create_using=nx.DiGraph())
5
6 pagerank = nx.pagerank(G, alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None)
7
8 with open("/Users/like/Desktop/cs572/hw4/solr-7.7.2/server/solr/myexample/external_pageRankFile.txt", "w") as f:
9     for pageid in pagerank:
10         f.write("/Users/like/Desktop/cs572/hw4/nytimes/" + pageid + ".html" + str(math.log(pagerank[pageid])) + "\n")
```

- iii. The results of the script were stored in external_pageRankFile.txt(in the same directory of html files “nytimes/”) for further use.

```
external_pageRankFile.txt
/Users/like/Desktop/cs572/hw4/nytimes/12e2ab5e-7903-48a7-8ae9-f311d64cde88.html=-9.14658845111
/Users/like/Desktop/cs572/hw4/nytimes/c99dad98-ac90-49ab-9a71-986c063999d6.html=-11.4495482835
/Users/like/Desktop/cs572/hw4/nytimes/7ccdd6c9-caa7-42f9-ac92-9053fc19e4a9.html=-7.71430932329
/Users/like/Desktop/cs572/hw4/nytimes/6f752553-5d83-4fa1-87f7-9206f8ba3c49.html=-11.5048739446
/Users/like/Desktop/cs572/hw4/nytimes/05aa1314-523b-43c4-98a3-04f73a5e55ae.html=-11.518619907
/Users/like/Desktop/cs572/hw4/nytimes/4fb0d554-b2cf-4a97-899d-28bef603472d.html=-9.66589343859
/Users/like/Desktop/cs572/hw4/nytimes/b5dc9299-b54f-4e69-bb7b-9aaaad408a36.html=-11.4742141441
/Users/like/Desktop/cs572/hw4/nytimes/45d55f46-f9e4-4f26-ad61-c0be9c0d32ae.html=-11.5163278221
/Users/like/Desktop/cs572/hw4/nytimes/37015b82-c070-412c-b957-38904eff5728.html=-10.5729704931
/Users/like/Desktop/cs572/hw4/nytimes/ec6bfe3e-508a-4290-863b-860bb56e54d3.html=-11.5160320871
/Users/like/Desktop/cs572/hw4/nytimes/2f27eb8d-e757-4c97-a70f-a97ecd3fedb5.html=-9.40063752528
/Users/like/Desktop/cs572/hw4/nytimes/180d0c68-f578-4578-9b78-825d9a26c69c.html=-10.9630292903
/Users/like/Desktop/cs572/hw4/nytimes/d8f9e55f-45d8-41eb-af76-6aa1fc0f8e45.html=-11.4077808569
/Users/like/Desktop/cs572/hw4/nytimes/ae8fd9cd-0f68-4494-a282-326d233ee8ce.html=-4.96048499598
/Users/like/Desktop/cs572/hw4/nytimes/38af7f50-be03-4aa4-a125-5f70359466cb.html=-11.4652933442
/Users/like/Desktop/cs572/hw4/nytimes/bde1a098-7137-40d5-ad68-f8040a70cdef.html=-11.4272424417
/Users/like/Desktop/cs572/hw4/nytimes/12eca33f-6e6d-4ce3-90f8-901ab3c6291f.html=-9.39406663105
/Users/like/Desktop/cs572/hw4/nytimes/842210c7-0ef1-4e25-9aba-fafee1e91534.html=-11.3253692135
/Users/like/Desktop/cs572/hw4/nytimes/3826fd2b-774e-4c55-b3b0-af31f20266f3.html=-9.93743755067
/Users/like/Desktop/cs572/hw4/nytimes/50a377da-0c10-43fd-bf5c-16ad0f9d1a06.html=-11.5210512361
/Users/like/Desktop/cs572/hw4/nytimes/6c7c31db-d992-4821-baa2-dc5c16457204.html=-11.4194695757
/Users/like/Desktop/cs572/hw4/nytimes/f8ef8218-bbca-4400-8877-2f4f9e4e554c.html=-11.2880774805
/Users/like/Desktop/cs572/hw4/nytimes/80376f82-c4b2-458a-8a17-bebd3daa8713.html=-11.4581477839
/Users/like/Desktop/cs572/hw4/nytimes/56897ce0-5ebd-4d9b-87c3-a8adaaeebc5f.html=-11.5212293011
/Users/like/Desktop/cs572/hw4/nytimes/13695412-ea2e-497d-a967-08d8cd7af908.html=-11.361834778
/Users/like/Desktop/cs572/hw4/nytimes/934ff594-6951-49b8-a745-124851f3c413.html=-11.2531469142
/Users/like/Desktop/cs572/hw4/nytimes/cbadd731-7a0f-45ec-a090-1e5e72ce23f9.html=-11.5313448003
/Users/like/Desktop/cs572/hw4/nytimes/79d8169f-f0e1-45eb-a730-7bfe1f4b0ac4.html=-11.4860339648
/Users/like/Desktop/cs572/hw4/nytimes/969beb26-d7b5-4088-a3b0-d041203f24a1.html=-11.4863949778
/Users/like/Desktop/cs572/hw4/nytimes/c5acb3cd-6425-4084-a37b-045d0274396b.html=-11.3228059923
```

3. Setting up Solr for PageRank algorithm compatibility
 - a. Modified managed-schema for using pageRankFile
 - i. `<FieldType name="external" keyField="id" defVal="0" class="solr.ExternalFileField" valType="pfloat" />`
 - ii. `<field name="pageRankFile" type="external" stored="false" indexed="false" />`
 - b. Modified solarconfig.xml file to reload external file
 - i. `<listener event="newSearcher" class="org.apache.solr.schema.ExternalFileFieldReloader" />`
 - ii. `<listener event="firstSearcher" class="org.apache.solr.schema.ExternalFileFieldReloader" />`
 - c. Reloaded Solr Interface using the reload button in the Core Admin section.
 - d. For using PageRank algorithm instead of default algorithm use the **parameter: pageRankFile desc**
4. Run the queries as described in the homework and computed the overlap.

● Screenshots

- a. The screenshot of initial page where I enter the query in input box

- b. Page which shows the results for Lucene(Default).

Comparing Lucene and PageRank Algorithm

Search:

☒ Lucene ☐ PageRank

[Submit Query](#)

Results 1 - 10 of 105:

[How Did a Virus From the Atlantic Infect Mammals in the Pacific?](#)

URL: <https://www.nytimes.com/2019/11/09/science/seals-distemper.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/73e479de-d240-43e0-8630-e6b8351bce9f.html

Description: Thawing sea ice may have opened the door, allowing the infection to cross oceans, a new study suggests.

[Scientists Were Hunting for the Next Ebola. Now the U.S. Has Cut Off Their Funding.](#)

URL: <https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/e0f08feb-8216-4316-a138-bde83e38ecf5.html

Description: Predict, a government research program, sought to identify animal viruses that might infect humans and to head off new pandemics.

[Kay Hagan, Former North Carolina Senator, Dies at 66](#)

URL: <https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/e293ff76-9271-4362-a931-d126403384b2.html

Description: A Democrat, she served one term after defeating Elizabeth Dole. She was found to have a tick-borne virus in 2016.

[Trump Plan to Stop Spread of H.I.V. Will Target 'Hot Spot' Areas](#)

URL: <https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/e59cb7d1-fa94-4ae5-88db-600de167adde.html

Description: The plan to end transmission of the virus in the United States by 2030, outlined in the president's State of the Union address, will focus on 48 counties where about half of new infections occur.

[Her Daughter, Born With a Dangerous Infection, Needs Close Care](#)

URL: <https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/90cac8ee-3c87-4c33-95a3-8f8fdded256e0.html

Description: Separated from her husband, with few others to rely on, Luisa Estefany Jimenez struggles to raise three daughters, including one with special needs.

[Trump Plan to Stop Spread of H.I.V. Will Target 'Hot Spot' Areas](#)

URL: <https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/e59cb7d1-fa94-4ae5-88db-600de167adde.html

Description: The plan to end transmission of the virus in the United States by 2030, outlined in the president's State of the Union address, will focus on 48 counties where about half of new infections occur.

[Her Daughter, Born With a Dangerous Infection, Needs Close Care](#)

URL: <https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/90cac8ee-3c87-4c33-95a3-8f8fdded256e0.html

Description: Separated from her husband, with few others to rely on, Luisa Estefany Jimenez struggles to raise three daughters, including one with special needs.

[When Is the Best Time to Get Your Flu Shot?](#)

URL: <https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/82a5705a-fcb2-4173-a99b-450a4eeab8f7.html

Description: Immunity to the flu virus may wane over the course of the flu season, so you don't want to get a flu shot too early, or too late.

[Crispr Takes Its First Steps in Editing Genes to Fight Cancer](#)

URL: <https://www.nytimes.com/2019/11/06/health/crispr-cancer-leukemia.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/4ecf471c-e93b-4a71-922b-d04d843b9fcb.html

Description: So far, a procedure that turbocharges the immune system to attack tumors seems safe, but it's too soon to tell whether it helps patients.

[Why Sexually Transmitted Infections Can't Shake Their Stigma](#)

URL: <https://www.nytimes.com/2019/08/13/style/sti-stigma-sexual-transmitted-infections.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/e964bd90-a6ae-4fc9-85a1-81002bf5d703.html

Description: We live in an era of sex positivity — until we get positive test results. And that's unfortunate, because S.T.I.s are on the rise.

[You're Swabbing a Dead Gorilla for Ebola. Then It Gets Worse.](#)

URL: <https://www.nytimes.com/2019/10/14/health/ebola-gorillas.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/d27db420-1993-4e74-94a1-2c715c722ccf.html

Description: Carrion flies inside your hood. Sweat turns your gloves into water balloons. This is tough work, but it could predict disease outbreaks.

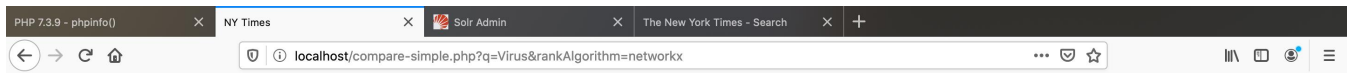
[Trump Expected to Announce Plan to Stop Spread of H.I.V. in the U.S. by 2030](#)

URL: <https://www.nytimes.com/2019/02/04/us/politics/trump-hiv-aids-plan.html>

Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/...nytimes/9437a8e1-d355-43f7-8225-38cb2b6352a4.html

Description: In his State of the Union address on Tuesday, the president is expected to announce a plan that calls for an expanded effort to prevent the virus that causes AIDS.

c. Page which shows the results for PageRank



Comparing Lucene and PageRank Algorithm

Search:

☐ Lucene ☒ PageRank

[Submit Query](#)

Results 1 - 10 of 105:

[New TB Vaccine Could Save Millions of Lives, Study Suggests](#)

URL: <https://www.nytimes.com/2019/10/29/health/tuberculosis-vaccine.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/fc8ea51c-37c0-4c46-b6cb-b6fb1ee17328.html
Description: There are 10 million new cases each year of tuberculosis, now the leading infectious cause of death worldwide. Even a partly effective vaccine could help turn the tide.

[I Hacked an Election. So Can the Russians.](#)

URL: <https://www.nytimes.com/video/opinion/100000005790489/i-hacked-an-election-so-can-the-russians.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/a0c1970b-e03e-4b13-b2a6-4790627753bf.html
Description: It's time America's leaders got serious about voting security.

[Searching for a Rectangular Sun Above the Arctic Circle](#)

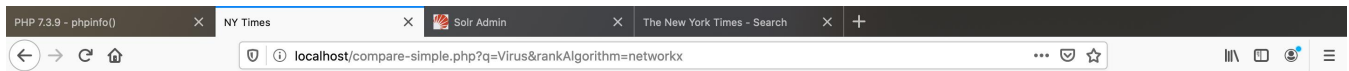
URL: <https://www.nytimes.com/2019/10/28/science/arctic-sunset.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/9163d5f1-b8f7-4681-a589-ad361f781439.html
Description: As polar winter sets in during an expedition on a research vessel, a reporter hopes she'll catch a glimpse of a brilliant mirage.

[Making Connections: 53 Teenagers Suggest Creative Ways to Link School Curriculum to the World of 2019](#)

URL: <https://www.nytimes.com/2019/03/07/learning/making-connections-53-teenagers-suggest-creative-ways-to-link-school-curriculum-to-the-world-of-2019.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/a13baff8-1681-4f03-b582-5a49ffe189c2.html
Description: The best of over 2,000 submissions that linked literature, history, science, math and art to recent news and opinion pieces in The Times.

[Trump Administration Sues Gilead, Maker of H.I.V.-Prevention Drugs](#)

URL: <https://www.nytimes.com/2019/11/07/business/gilead-truvada-hiv-lawsuit.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/cbef5a7a-3b98-4c34-a5de-4a2390d1899d.html



[Making Connections: 53 Teenagers Suggest Creative Ways to Link School Curriculum to the World of 2019](#)

URL: <https://www.nytimes.com/2019/03/07/learning/making-connections-53-teenagers-suggest-creative-ways-to-link-school-curriculum-to-the-world-of-2019.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/a13baff8-1681-4f03-b582-5a49ffe189c2.html
Description: The best of over 2,000 submissions that linked literature, history, science, math and art to recent news and opinion pieces in The Times.

[Trump Administration Sues Gilead, Maker of H.I.V.-Prevention Drugs](#)

URL: <https://www.nytimes.com/2019/11/07/business/gilead-truvada-hiv-lawsuit.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/cbef5a7a-3b98-4c34-a5de-4a2390d1899d.html
Description: The lawsuit accused the pharmaceutical company of profiting billions of dollars off taxpayer research without paying royalties.

[Opinion | Xi Jinping's Chinese Dream](#)

URL: <https://www.nytimes.com/2013/06/05/opinion/global/xi-jinpings-chinese-dream.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/a4738bce-2fb9-462f-8b14-88f3d476f638.html
Description: American policy makers must understand that the new Chinese leader needs to be a nationalist to be a reformer.

[As Congo's Ebola Outbreak Drags On, Untracked Cases Sow Confusion](#)

URL: <https://www.nytimes.com/2019/09/16/health/ebola-congo-azar-fauci.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/0ea37bbb-42e8-405f-9749-699e307cfa6d.html
Description: On tour in Africa, American officials said the U.S. would keep providing aid. But Congo's response has been uneven, and the former health minister has been jailed.

[The 25 Works of Art That Define the Contemporary Age](#)

URL: <https://www.nytimes.com/2019/07/15/t-magazine/most-important-contemporary-art.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/3a5436bc-9e3c-43a4-ae70-f81e59500d3f.html
Description: Three artists and a pair of curators came together at The New York Times to attempt to make a list of the era's essential artworks. Here's their conversation.

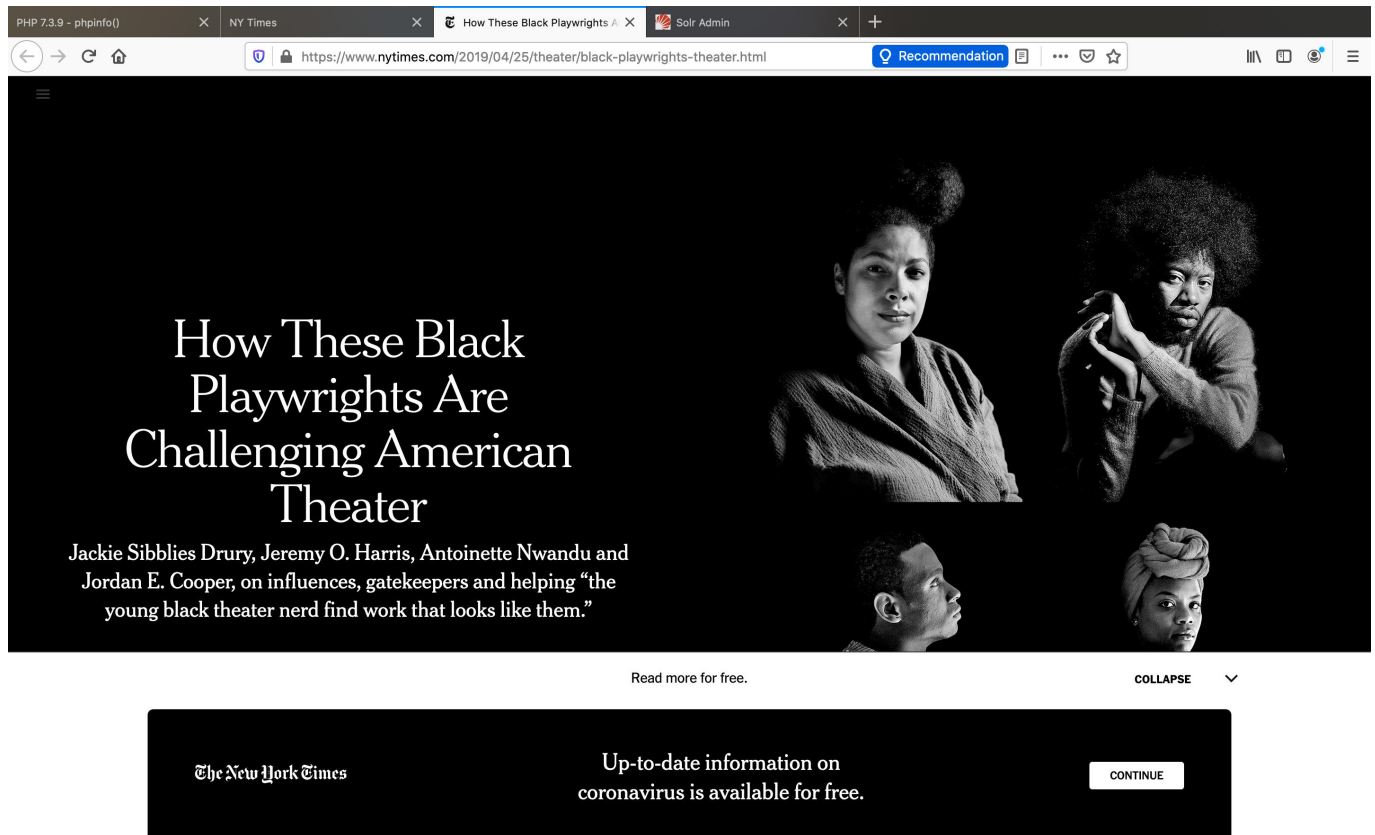
[The Strange and Beautiful Universe of Walter Van Beirendonck](#)

URL: <https://www.nytimes.com/2019/09/03/t-magazine/walter-van-beirendonck.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/b43e4021-7d5f-4893-baaf-8f20bf3f7499.html
Description: For more than three decades, the Belgian designer has been offering gleefully irreverent alternatives to the staid men's wear of old.

[Why Do Parrots Waste So Much Food?](#)

URL: <https://www.nytimes.com/2019/11/02/science/why-parrots-waste-food.html>
Id: /Users/like/Desktop/cs572/hw4/solr-7.7.2/.../nytimes/d63b9352-0a71-4419-b255-d28cd1101913.html
Description: The behavior is so consistent across species and in the wild that some scientists think it must be intentional.

d. Screenshot of the actual webpage that opens when clicking on one of the result's link.



- Explanation regarding why some pages have higher PageRank values than others. (One paragraph).

PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. It is computed based on the number of links to a particular webpage (in-links) and the number of links out from the webpage (out-links). The more in-links there are on the page, the higher the PageRank. In addition, the higher the PageRank of the in-links to the web page, the higher the PageRank. This is why some pages have a higher PageRank than others.

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

Cannes Comparation		
LUCENE		PAGERANK
1	https://www.nytimes.com/2018/04/12/movies/cannes-lineup-netflix-boycott.html	https://www.nytimes.com/2019/04/25/theaterblack-playwrights-theater.html
2	https://www.nytimes.com/2019/05/24/movies/cannes-almodovar-kechiche.html	https://www.nytimes.com/2019/10/10/movies/parasite-review.html
3	https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html	https://www.nytimes.com/2019/10/04/movies/new-york-film-festival.html
4	https://www.nytimes.com/2019/05/25/movies/cannes-film-festival-winners-parasite.html	https://www.nytimes.com/2018/10/03/movies/new-york-film-festival.html
5	https://www.nytimes.com/2019/05/16/movies/cannes-film-festival-jim-jarmusch.html	https://www.nytimes.com/2017/10/05/us/harvey-weinstein-harassment-allegations.html
6	https://www.nytimes.com/2019/05/21/movies/cannes-film-festival-antonio-banderas-pedro-almodovar.html	https://www.nytimes.com/2018/09/28/obituaries/inge-feltrinelli-dead.html
7	https://www.nytimes.com/2019/05/27/movies/cannes-oscars.html	https://www.nytimes.com/2019/01/09/movies/black-panther-oscars-ryan

		coogler.html
8	https://www.nytimes.com/2019/05/20/arts/cannes-film-festival-black-director.html	https://www.nytimes.com/search?query=Nicolas%20Rapold&module=inline
9	https://www.nytimes.com/2019/05/22/movies/quentin-tarantino-margot-robbie.html	https://www.nytimes.com/2019/09/06/movies/fall-movies.html
10	https://www.nytimes.com/2019/05/22/movies/dicaprio-pltt.html	https://www.nytimes.com/2019/10/03/movies/pedro-almodovar-pain-glory.html

Congress Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html	https://www.nytimes.com/1999/10/13/us/philip-morris-admits-evidence-shows-smoking-causes-cancer.html
2	https://www.nytimes.com/2019/11/05/magazine/congress-president-impeachment.html	https://www.nytimes.com/2019/09/25/us/politics/ukraine-trump-whistleblower.html
3	https://www.nytimes.com/2019/07/25/us/politics/mueller-impeachment.html	https://www.nytimes.com/2019/04/09/learning/is-horse-racing-ethical.html
4	https://www.nytimes.com/2019/07/26/us/politics/donald-trump-impeachment.html	https://www.nytimes.com/2019/10/17/us/politics/mick-mulvaney-trump-ukraine.html
5	https://www.nytimes.com/2018/09/17/us/politics/bloomberg-president-2020-democrat.html	https://www.nytimes.com/2019/02/10/us/politics/amy-klobuchar-president-2020.html
6	https://www.nytimes.com/2018/11/21/obituaries/james-billington-dead.html	https://www.nytimes.com/2019/05/28/us/politics/biden-education-plan-2020.html
7	https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html	https://www.nytimes.com/1998/09/04/us/senate-democrat-rebukes-clinton.html
8	https://www.nytimes.com/2019/08/07/us/politics/foreign-aid-freeze-congress.html	https://www.nytimes.com/2019/03/13/technology/facebook-data-deals-investigation.html
9	https://www.nytimes.com/1973/05/13/archives/constitutional-ervin-ah-thi-nk-the-senate-is-determined-to-recover.html	https://www.nytimes.com/2019/03/26/us/politics/boeing-faa.html
10	https://www.nytimes.com/2017/12/21/world/americas/peru-president-impeachment.html	https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html

Democrats Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2018/09/04/us/politics/ayanna-pressley-massachusetts.html	https://www.nytimes.com/video/us/100000004557512/the-dnc-music-festival.html
2	https://www.nytimes.com/2018/11/07/us/politics/election-history-firsts-blackburn-pressley.html	https://www.nytimes.com/video/us/elections/100000006582230/democratic-debate-highlights.html
3	https://www.nytimes.com/2019/10/16/us/politics/democrats-impeachment.html	https://www.nytimes.com/2018/06/04/upshot/school-funding-still-lags-after-recession-ended.html
4	https://www.nytimes.com/2019/09/29/us/politics/pelosi-trump-impeachment.html	https://www.nytimes.com/2019/09/25/us/politics/ukraine-trump-whistleblower.html
5	https://www.nytimes.com/2018/01/23/us/politics/impeach-trump-democrats-tom-steyer.html	https://www.nytimes.com/video/us/politics/100000004549242/tim-kaine-the-harmonica-player.html
6	https://www.nytimes.com/2019/09/24/us/politics/trump-democrats.html	https://www.nytimes.com/2019/10/17/us/politics/mick-mulvaney-trump-ukraine.html
7	https://www.nytimes.com/2019/10/28/us/politics/trump-impeachment-subpoena-house.html	https://www.nytimes.com/2019/02/10/us/politics/amy-klobuchar-president-2020.html
8	https://www.nytimes.com/2019/09/11/opinion/black-voters-democrats-2020-election.html	https://www.nytimes.com/video/us/politics/100000006761447/democratic-debate-cnn-nyt-10-15-19-clip-7.html
9	https://www.nytimes.com/2019/11/05/opinion/trump-democrats-2020.html	https://www.nytimes.com/2019/09/20/opinion/biden-warren-2020-elections.html
10	https://www.nytimes.com/2019/11/01/us/politics/house-republican-box-prank.html	https://www.nytimes.com/2019/10/21/us/elizabeth-warren-education-plan.html

Patriot Movement Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2019/08/14/opinion/joe-walsh-trump-primary.html	https://www.nytimes.com/2019/11/04/climate/trump-paris-agreement-climate.html

2	https://www.nytimes.com/1976/08/27/archives/lewis-michaux-92-dies-ran-bookstore-in-harlem.html	https://www.nytimes.com/2018/06/04/upshot/school-funding-still-lags-after-recession-ended.html
3	https://www.nytimes.com/2019/05/15/us/politics/us-iraq-embassy-evacuation.html	https://www.nytimes.com/video/us/politics/10000004549242/tim-kaine-the-harmonica-player.html
4	https://www.nytimes.com/2019/01/07/obituaries/moshe-arens-dead.html	https://www.nytimes.com/2019/04/09/learning/is-horse-racing-ethical.html
5	https://www.nytimes.com/2019/10/27/obituaries/john-conyers-jr-dead.html	https://www.nytimes.com/2019/02/10/us/politics/amy-klobuchar-president-2020.html
6	https://www.nytimes.com/2018/12/29/obituaries/shehu-shagari-dead.html	https://www.nytimes.com/1970/10/17/archives/jewelry-that-has-sense-of-movement.html
7	https://www.nytimes.com/2019/10/22/sports/football/jets-patriots.html	https://www.nytimes.com/2019/02/23/business/cell-phone-addiction.html
8	https://www.nytimes.com/2019/09/23/opinion/democrats-republicans.html	https://www.nytimes.com/2019/10/21/us/elizabeth-warren-education-plan.html
9	https://www.nytimes.com/2019/10/30/opinion/alexander-vindman-trump-ukraine.html	https://www.nytimes.com/video/movies/100000006156533/first-man-ryan-gosling.html
10	https://www.nytimes.com/2019/09/20/us/data-privacy-fbi.html	https://www.nytimes.com/2019/03/14/world/europe/climate-action-strikes-youth.html

Republicans Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html	https://www.nytimes.com/video/us/10000004557512/the-dnc-music-festival.html
2	https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html	https://www.nytimes.com/video/us/elections/10000006582230/democratic-debate-highlights.html
3	https://www.nytimes.com/2019/10/31/us/politics/house-impeachment-partisan-vote.html	https://www.nytimes.com/2019/10/29/opinion/trump-zuckerberg.html
4	https://www.nytimes.com/2019/11/04/us/politics/trump-republicans-impeachment.html	https://www.nytimes.com/2018/06/04/upshot/school-funding-still-lags-after-recession-ended.html
5	https://www.nytimes.com/2019/10/04/us/politics/republicans-trump-defense.html	https://www.nytimes.com/2019/10/17/us/politics/mick-mulvaney-trump-ukraine.html
6	https://www.nytimes.com/2019/09/30/opinion/republicans-trump-impeachment.html	https://www.nytimes.com/2019/02/10/us/politics/amy-klobuchar-president-2020.html
7	https://www.nytimes.com/2019/11/09/us/politics/republican-strategy-impeachment-trump.html	https://www.nytimes.com/1998/09/04/us/senate-democrat-rebukes-clinton.html
8	https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html	https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html
9	https://www.nytimes.com/2019/11/06/us/trump-senate-republicans-courts.html	https://www.nytimes.com/2019/09/25/opinion/impeachment-democrats.html
10	https://www.nytimes.com/2019/10/08/us/politics/republicans-trump-syria.html	https://www.nytimes.com/2019/09/20/us/politics/trump-whistle-blower-ukraine.html

Senate Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2019/10/01/us/politics/senate-trump-impeachment.html	https://www.nytimes.com/1999/10/13/us/philip-morris-admits-evidence-shows-smoking-causes-cancer.html
2	https://www.nytimes.com/2019/10/26/us/trump-senate-presidency.html	https://www.nytimes.com/video/us/politics/10000004549242/tim-kaine-the-harmonica-player.html
3	https://www.nytimes.com/2019/08/03/us/politics/senate-votes-mcconnell.html	https://www.nytimes.com/2019/10/17/us/politics/mick-mulvaney-trump-ukraine.html
4	https://www.nytimes.com/2019/08/22/us/politics/john-hickenlooper-senate-2020.html	https://www.nytimes.com/2019/02/10/us/politics/amy-klobuchar-president-2020.html
5	https://www.nytimes.com/2019/11/08/us/impeachment-senate-democrats.html	https://www.nytimes.com/2019/05/28/us/politics/biden-education-plan-2020.html
6	https://www.nytimes.com/2019/08/31/us/politics/senate-race-2020.html	https://www.nytimes.com/1998/09/04/us/senate-democrat-rebukes-clinton.html
7	https://www.nytimes.com/2019/10/24/us/politics/senate-republicans-trump.html	https://www.nytimes.com/2019/03/26/us/politics/boeing-faa.html
8	https://www.nytimes.com/2019/10/18/us/politics/mcconnell-impeachment-senate.html	https://www.nytimes.com/2019/09/07/us/politics/congress-assault-weapons-ban.html
9	https://www.nytimes.com/2019/09/05/opinion/the-argument-senate-democrats-2020.html	https://www.nytimes.com/2019/09/25/opinion/impeachment-democrats.html
10	https://www.nytimes.com/2019/09/05/opinion/the-argument-senate-democrats-2020.html	https://www.nytimes.com/2019/07/11/science/moon-apollo-11-archaeology-preservation.html

Olympics 2020 Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html	https://www.nytimes.com/2015/06/25/t-magazine/ed-ruscha-cactus-omelet-recipe.html
2	https://www.nytimes.com/2018/02/15/learning/should-technology-in-sports-be-limited.html	https://www.nytimes.com/2019/11/04/climate/trump-paris-agreement-climate.html
3	https://www.nytimes.com/2019/10/01/travel/what-you-need-to-know-the-tokyo-2020-olympics.html	https://www.nytimes.com/2019/08/01/arts/dorothy-toy-dead.html
4	https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html	https://www.nytimes.com/2019/10/09/arts/television/whats-on-tv-wednesday-nancy-drew-goliath.html
5	https://www.nytimes.com/video/sports/olympics/100000005750317/johnny-weir-tara-lipinski-figure-skating.html	https://www.nytimes.com/2019/11/06/learning/lesson-of-the-day-new-library-is-a-41-5-million-masterpiece-but-about-those-stairs.html
6	https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html	https://www.nytimes.com/2018/06/19/obituaries/dr-adel-mahmoud-76-dies-credited-with-major-vaccines.html
7	https://www.nytimes.com/video/sports/olympics/100000005752444/south-korea-speedskaters.html	https://www.nytimes.com/2019/11/07/sports/milwaukee-bucks-clippers-Giannis-Antetokounmpo.html
8	https://www.nytimes.com/2017/12/05/sports/olympics/ioc-russia-winter-olympics.html	https://www.nytimes.com/video/us/100000004557512/the-dnc-music-festival.html
9	https://www.nytimes.com/video/sports/football/100000005827420/nfl-cheerleaders-and-the-gender-gap.html	https://www.nytimes.com/video/opinion/100000003473921/hotel-22.html
10	https://www.nytimes.com/video/sports/football/100000005827420/nfl-cheerleaders-and-the-gender-gap.html	https://www.nytimes.com/2019/09/24/world/europe/holocaust-diary-rena-spiegel.html

Stock Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2019/05/02/business/tesla-stock-fundraising.html	https://www.nytimes.com/2019/11/07/sports/milwaukee-bucks-clippers-Giannis-Antetokounmpo.html
2	https://www.nytimes.com/2018/12/16/technology/tech-workers-company-stock-shareholder-activism.html	https://www.nytimes.com/2019/10/29/opinion/trump-zuckerberg.html
3	https://www.nytimes.com/2019/05/10/technology/uber-stock-price-ipo.html	https://www.nytimes.com/2019/05/08/business/gm-lordstown-trump.html
4	https://www.nytimes.com/2019/11/04/business/under-armour-stock-investigation.html	https://www.nytimes.com/2018/11/29/t-magazine/food-as-art.html
5	https://www.nytimes.com/2019/09/20/technology/airbnb-employees-ipo-payouts.html	https://www.nytimes.com/2019/06/21/books/review/crime-fiction-marilyn-stasio-kate-atkinson-laura-purcell.html
6	https://www.nytimes.com/2019/10/24/style/rebag-clair-handbag-stock-market.html	https://www.nytimes.com/2018/12/20/obituaries/donald-moffat-dead.html
7	https://www.nytimes.com/2019/05/30/technology/uber-stock-earnings.html	https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html
8	https://www.nytimes.com/2018/11/12/obituaries/david-pearson-nascar-dead.html	https://www.nytimes.com/2019/08/20/business/dealbook/business-roundtable-corporate-responsibility.html
9	https://www.nytimes.com/2019/04/03/business/tesla-sales-elon-musk.html	https://www.nytimes.com/2001/12/31/nyregion/the-giuliani-years-the-overview-a-man-who-became-more-than-a-mayor.html
10	https://www.nytimes.com/2019/10/28/business/stock-market-record.html	https://www.nytimes.com/2019/06/04/us/politics/republicans-mexico-tariffs.html

Virus Comparison		
LUCENE		PAGERANK
1	https://www.nytimes.com/2019/11/09/science/seals-distemper.html	https://www.nytimes.com/2019/10/29/health/tuberculosis-vaccine.html
2	https://www.nytimes.com/2019/10/25/health/predict-usaid-viruses.html	https://www.nytimes.com/video/opinion/100000005790489/i-hacked-an-election-so-can-the-russians.html
3	https://www.nytimes.com/2019/10/28/us/politics/kay-hagan-dead.html	https://www.nytimes.com/2019/10/28/science/arctic-sunset.html
4	https://www.nytimes.com/2019/02/05/us/politics/trump-hiv-aids-plan.html	https://www.nytimes.com/2019/03/07/learning/making-connections-53-teenagers-suggest-creative-ways-to-link-school-curriculum-to-the-world-of-2019.html
5	https://www.nytimes.com/2019/01/23/neediest-cases/daughter-cmv.html	https://www.nytimes.com/2019/11/07/business/gilead-truvada-hiv-lawsuit.html
6	https://www.nytimes.com/2019/09/27/well/live/when-is-the-best-time-to-get-your-flu-shot.html	https://www.nytimes.com/2013/06/05/opinion/global/xi-jinpings-chinese-dream.html

7	https://www.nytimes.com/2019/11/06/health/crispr-cancer-leukemia.html	https://www.nytimes.com/2019/09/16/health/ebola-congo-azar-fauci.html
8	https://www.nytimes.com/2019/08/13/style/sti-stigma-sexual-transmitted-infections.html	https://www.nytimes.com/2019/07/15/t-magazine/most-important-contemporary-art.html
9	https://www.nytimes.com/2019/10/14/health/ebola-gorillas.html	https://www.nytimes.com/2019/09/03/t-magazine/walter-van-beirendonck.html
10	https://www.nytimes.com/2019/02/04/us/politics/trump-hiv-aids-plan.html	https://www.nytimes.com/2019/11/02/science/why-parrots-waste-food.html

● OVERLAPS

QUERY	NO. OF OVERLAPS
Cannes	0
Democrats	0
Congress	0
Patriot Movement	0
Republicans	0
Senate	0
Olympics 2020	0
Stock	0
Virus	0

There were no overlaps in any query. So, no bar graph has been drawn as per the instructions given.