

Information Retrieval and Web Search Engines

Homework 4: Indexing the Web Using Solr

REPORT

Name - Mukund Murali

USC ID - 1264824477

Steps followed to complete the assignment

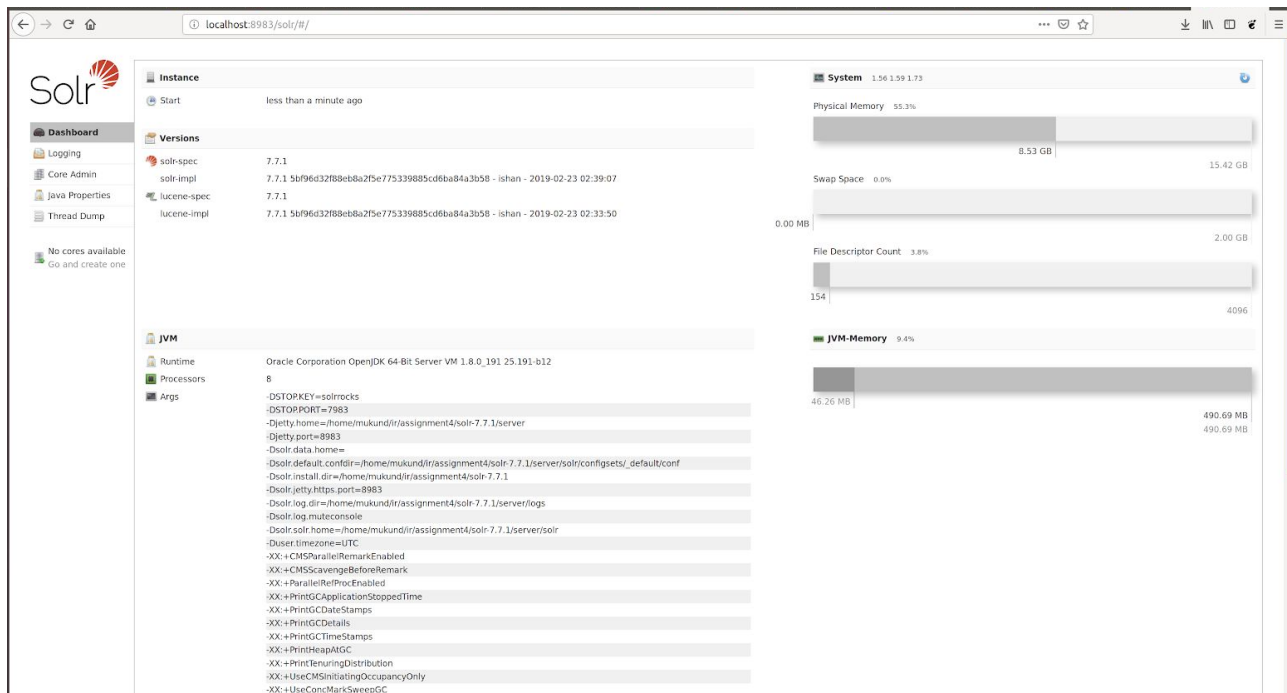
The steps followed to complete the assignment are detailed in the following sections.

Download crawled web pages

Downloaded the crawled web pages for the news site **Reuters** from,
https://drive.google.com/drive/u/0/folders/1EFkdy9cLjq0hNuC_bRQ0UO75djtvrwj7?usp=sharing

Solr installation, configuration and indexing

1. Downloaded Solr from [solr-7.7.1.zip](#)
2. Unzipped the content using `unzip solr-7.7.1`
3. Checked if Solr is starting,
 - a. `cd solr-7.7.1/bin`
 - b. `./solr start`
 - c. Screenshot of solr homepage



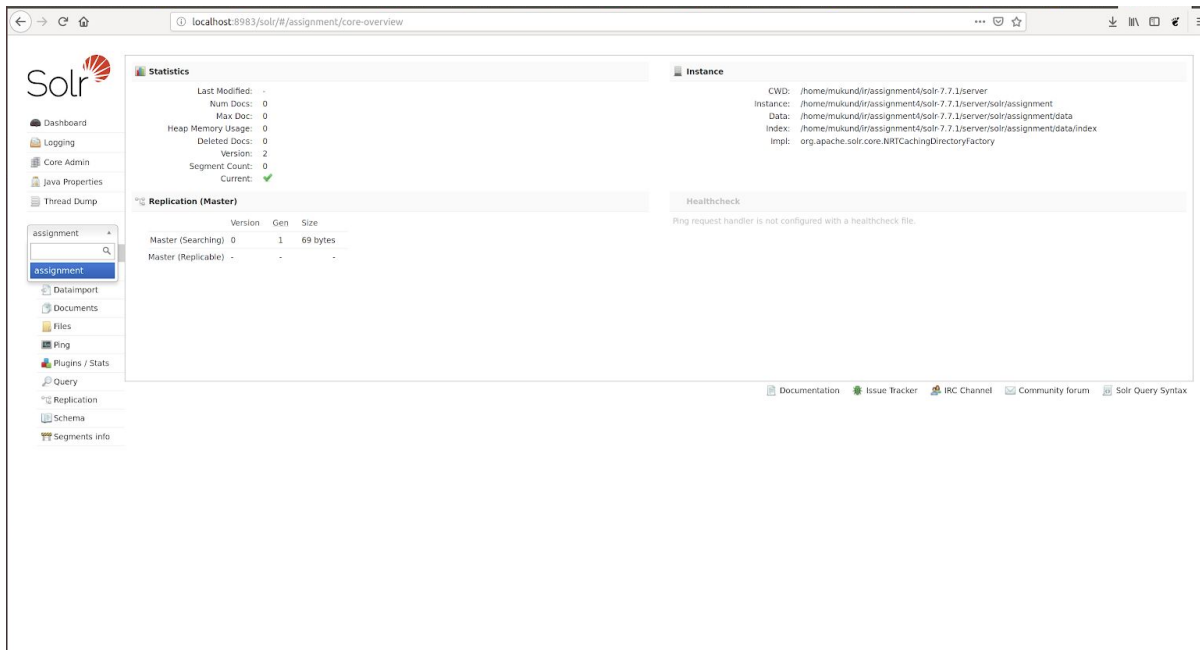
4. Created a new core named **assignment**

- `./solr create -c assignment`
- Screenshot of core creation

```
(base) mukund@mukund:~/ir/assignment4/solr-7.7.1/bin$ ./solr create -c assignment
WARNING: Using default configset with data driven schema functionality. NOT RECOMMENDED for production use.
To turn off: bin/solr config -c assignment -p 8983 -action set-user-property -property update.autoCreateFields -value false
INFO - 2019-04-09 22:12:02.633; org.apache.solr.util.configuration.SSLCredentialProviderFactory; Processing SSL Credential Provider chain: env;sysprop

Created new core 'assignment'
(base) mukund@mukund:~/ir/assignment4/solr-7.7.1/bin$
```

c. Screenshot of Solr dashboard with the created core



5. Made the following changes in `solr-7.7.1/server/solr/assignment/conf/managed-schema` file,

- Uncommented `<copyField source="*" dest="_text_"/>`

6. Configured Solr to default querying from the field `_text_` by defining the default field in the requestHandler in the `solrconfig.xml` file. Made the following changes to `solr-7.7.1/server/solr/assignment/conf/solrconfig.xml` file,

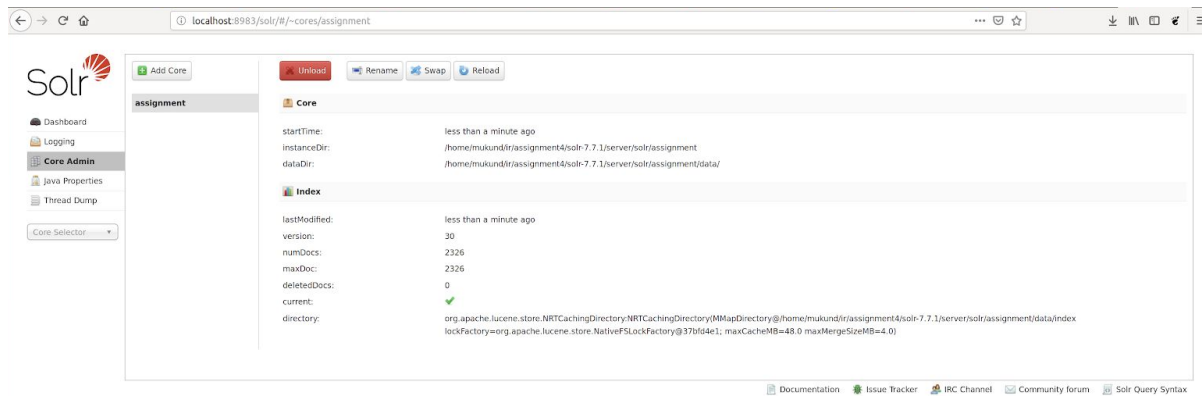
- Uncommented and modified the line as, `<str name="df">_text_</str>`

7. Indexing is performed by using the command,

- `solr-7.7.1/bin/post -filetypes html -c assignment`
`/home/mukund/ir/assignment4/Reuters/reutersnews/`
- Pages will get indexed by Solr

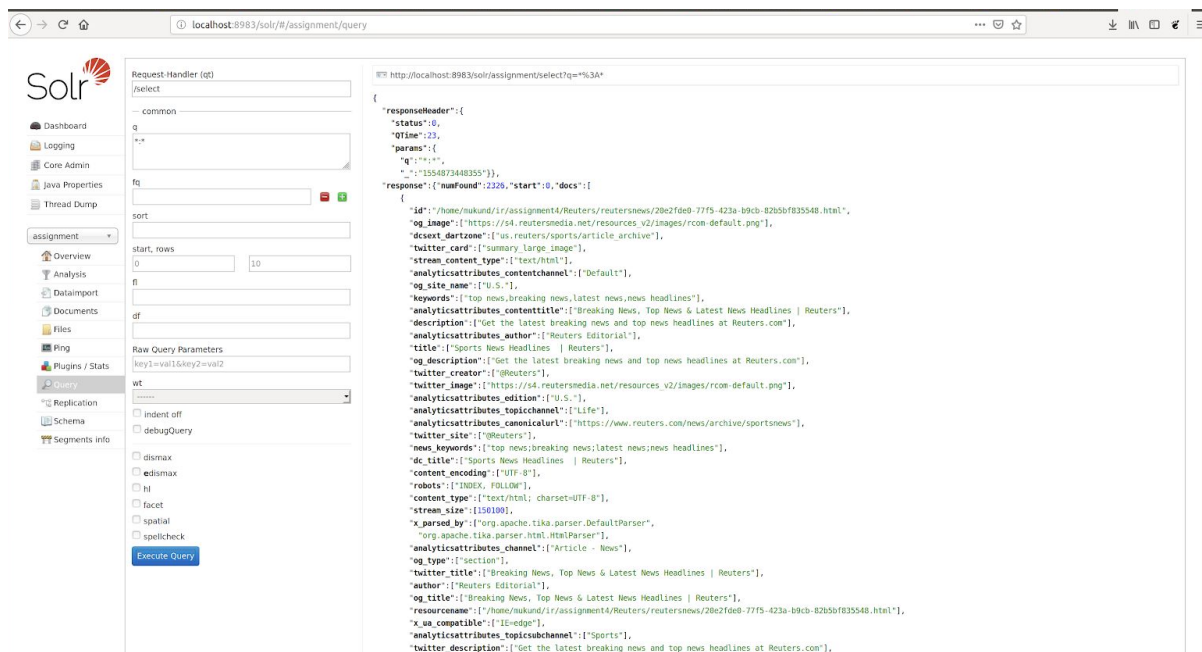
```
.POSTing file 2e1a70ee-4d97-4bb2-9cf7-857449b55eb5.html (text/html) to (base)/extract
.POSTing file 7d951303-535f-4da5-8864-37f59d1fa0f1.html (text/html) to (base)/extract
.POSTing file 19b3f827-c489-4d56-ae1-7e8db82a8e41.html (text/html) to (base)/extract
.POSTing file df814e05-f966-45bc-afba-49428d581e9a.html (text/html) to (base)/extract
.POSTing file 8e55eed1-342c-415b-b8ad-63f8b5524a44.html (text/html) to (base)/extract
.POSTing file 1cd3d1fe-b260-4fa2-b147-ae79cf8c8457.html (text/html) to (base)/extract
.POSTing file 9189ae16-3c21-4348-b2b-a19961152e08.html (text/html) to (base)/extract
.POSTing file 7ead3ba7-491b-4c9a-8ea7-b11168bd1aed.html (text/html) to (base)/extract
.POSTing file ac3e74a8-cd54-4138-be7e-f277d08263fd.html (text/html) to (base)/extract
.POSTing file 8c99f81b-6836-4379-88cc-b34c448b52af.html (text/html) to (base)/extract
.POSTing file df7d6a65-6593-40f8-a87d-991b47ad5f66.html (text/html) to (base)/extract
.POSTing file 3ba57593-4d5a-44f6-82ab-fbed01b40669.html (text/html) to (base)/extract
.POSTing file 7b8c94e1-38c2-4981-ad51-0c4ec65af8f8.html (text/html) to (base)/extract
.POSTing file 4d253a8b-83d4-40c5-8246-e1be0f5c9593.html (text/html) to (base)/extract
.POSTing file fa34d894-c7c7-4d20-98c7-d770863b387.html (text/html) to (base)/extract
.POSTing file bbc5ae45-0582-48e7-8245-43a5dc93c43c.html (text/html) to (base)/extract
.POSTing file ee38573e-3b31-47dd-9df4-c89de29aae86.html (text/html) to (base)/extract
.POSTing file b6326086-3fda-4d65-b997-85247b03d5bf.html (text/html) to (base)/extract
.POSTing file 6fd3bb8d-de6a-463c-83f8-3057efe14738.html (text/html) to (base)/extract
2326 files indexed.
COMMITTING Solr index changes to http://localhost:8983/solr/assignment/update...
Time spent: 0:01:09.178
(base) mukund@mukund:~/ir/assignment4$
```

8. Reloaded the core by going inside Core Admin



The screenshot shows the Solr Core Admin interface for the 'assignment' core. The left sidebar contains navigation links: Dashboard, Logging, Core Admin (selected), Java Properties, Thread Dump, and Core Selector. The main content area displays the core's configuration and status. At the top, there are buttons for 'Add Core', 'Reload', 'Rename', 'Swap', and 'Reload'. Below this, the 'Core' section shows the start time as 'less than a minute ago', the instance directory as '/home/mukund/ir/assignment4/solr-7.7.1/server/solr/assignment', and the data directory as '/home/mukund/ir/assignment4/solr-7.7.1/server/solr/assignment/data'. The 'Index' section shows the last modified time as 'less than a minute ago', the version as 30, the number of documents as 2326, the maximum document size as 2326, and the current status as 'OK'. The directory path is also listed.

9. Screenshot of results for default query



The screenshot shows the Solr Query interface. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query (selected), Schema, and Segments info. The main content area displays the query results. The 'Request Handler (q)' is set to '/select'. The 'q' field contains the query 'http://localhost:8983/solr/assignment/select?q=*%3A*'. The 'wt' field is set to 'json'. The 'indent' checkbox is checked. The 'debugQuery' checkbox is checked. The 'Execute Query' button is visible. The results are displayed in a JSON format, showing a single document with various fields including 'id', 'og_image', 'descrip', 'twitter_card', 'stream_content_type', 'analyticsattributes_contentchannel', 'og_site_name', 'keywords', 'analyticsattributes_contenttitle', 'description', 'analyticsattributes_author', 'title', 'og_description', 'twitter_creator', 'twitter_image', 'analyticsattributes_edition', 'analyticsattributes_topichannel', 'analyticsattributes_canonicalurl', 'twitter_site', 'news_keywords', 'dc_title', 'content_encoding', 'robots', 'content_type', 'stream_size', 'x_parsed_by', 'org.apache.tika.parser.html.HtmlParser', 'analyticsattributes_channel', 'og_type', 'twitter_site', 'author', 'og_title', 'resourceName', 'x_ua_compatible', 'analyticsattributes_topichannel', and 'twitter_description'.

10. Screenshot of results for sample query

Computing Page Rank using Python code

1. Used the Python NetworkX library to compute page rank using the edgeList file generated above
2. To import NetworkX, `import networkx as nx`
3. Loaded the webgraph to NetworkX using,
`G = nx.read_edgelist("edgeList.txt", create_using=nx.DiGraph())`
4. Used PageRank() function to computes a ranking of the nodes in the graph G based on the structure of the incoming links,
`myPR = nx.pagerank(G, alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None)`
5. The output file generated is `external_pageRankFile.txt`. It is of the format,

```
/home/nukund/r/assignment4/Reuters/reutersnews/a8c907ad-01fb-4cd8-9eeb-22b422b81cef.html=6.575515427463519e-05
/home/nukund/r/assignment4/Reuters/reutersnews/ac5c8a84-d2e2-4411-ae49-4f65c3b0d2dc.html=0.0002956849347264689
/home/nukund/r/assignment4/Reuters/reutersnews/20f9094b-23c3-40f2-809c-4b99a88c791e.html=6.843056165806882e-05
/home/nukund/r/assignment4/Reuters/reutersnews/c475223-d0b3-462b-884c-6392c5e80202.html=0.00030851330374372
/home/nukund/r/assignment4/Reuters/reutersnews/40b69e2f-2cce-4ac8-9984-3f77468d59dc.html=6.575515427463519e-05
/home/nukund/r/assignment4/Reuters/reutersnews/9e3ce3db-43a2-4b36-93ea-38a11d4e8c99.html=6.799511904018767e-05
/home/nukund/r/assignment4/Reuters/reutersnews/b77c3c6d-83b4-4308-a73e-e5ae6dabe092.html=6.575515427463519e-05
/home/nukund/r/assignment4/Reuters/reutersnews/0818c9f0-b195-4d2e-b080-e4762e465176.html=0.0001206571806972616
/home/nukund/r/assignment4/Reuters/reutersnews/26fcb2b8-9a0f-4d81-a7cc-ac64a721e97f.html=0.00022414925631303624
/home/nukund/r/assignment4/Reuters/reutersnews/06980714-cd5d-49ed-8599-a2ca3a60d4ba.html=0.0002420808637140478
/home/nukund/r/assignment4/Reuters/reutersnews/0a6e2f38-f097-45ce-b815-4a7cb6eb3aea.html=0.000122912201560777
/home/nukund/r/assignment4/Reuters/reutersnews/0b7d1a09-649f-4d9c-81c9-5711f1a9147f.html=6.575515427463519e-05
/home/nukund/r/assignment4/Reuters/reutersnews/e77b27ad-cfb9-45a6-ace5-007b04f73438.html=6.828251351863688e-05
/home/nukund/r/assignment4/Reuters/reutersnews/e9f882a3-5607-4079-9e40-b7e91d61928f.html=0.00024008512148531702
/home/nukund/r/assignment4/Reuters/reutersnews/0e22d4c5-5683-4dcb-8531-4f089a42b09e.html=0.00013432222398795024
/home/nukund/r/assignment4/Reuters/reutersnews/239e42cc-4010-459f-8dca-b39f76a02b06.html=7.126196444625243e-05
/home/nukund/r/assignment4/Reuters/reutersnews/c48e8841-8a6d-40a2-b2f5-00bbd993fe33.html=6.575515427463519e-05
/home/nukund/r/assignment4/Reuters/reutersnews/60e9c332-12e3-4bf3-a4de-39f5bd73101c.html=9.300954380363075e-05
/home/nukund/r/assignment4/Reuters/reutersnews/7009d0ff-1086-4bdb-8ac3-4ba2fbc0c779.html=0.0002444437905371668
external_pageRankFile.txt' 22861, 256339c
```

2286,1

Bot

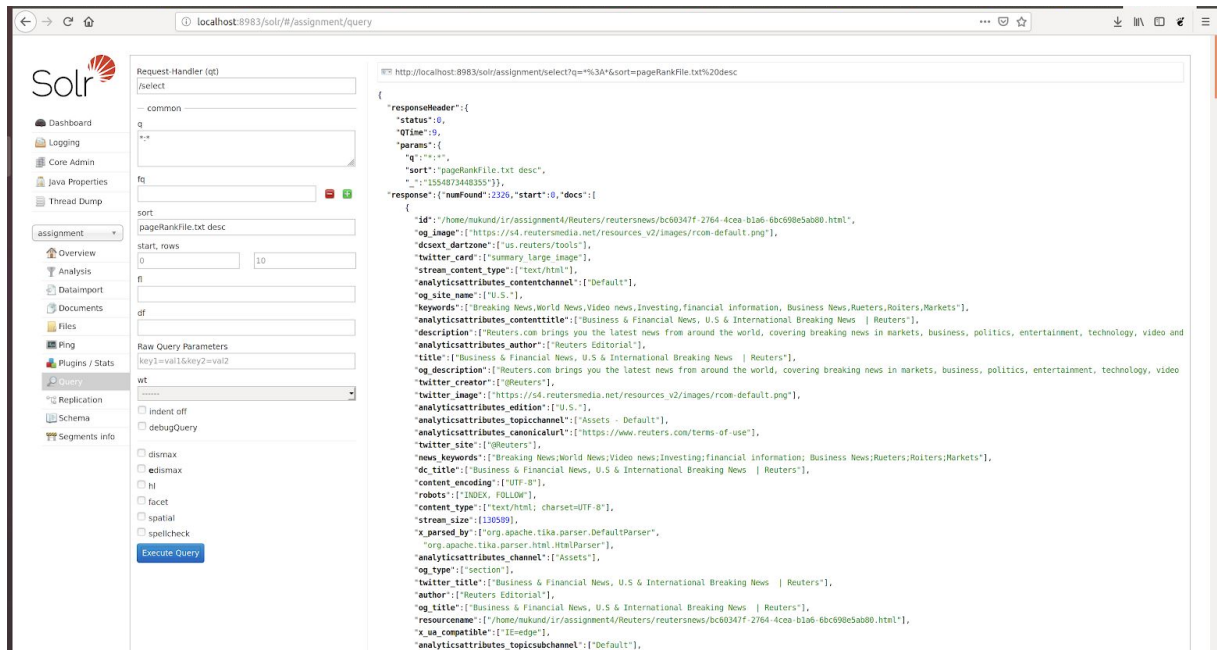
Changes to Solr in order to include the external page rank file

1. Placed the `external_pageRankFile.txt` into the `solr-7.7.1/server/solr/assignment/data/` directory
2. Added the following lines to `solr-7.7.1/server/solr/assignment/conf/managed-schema` file,

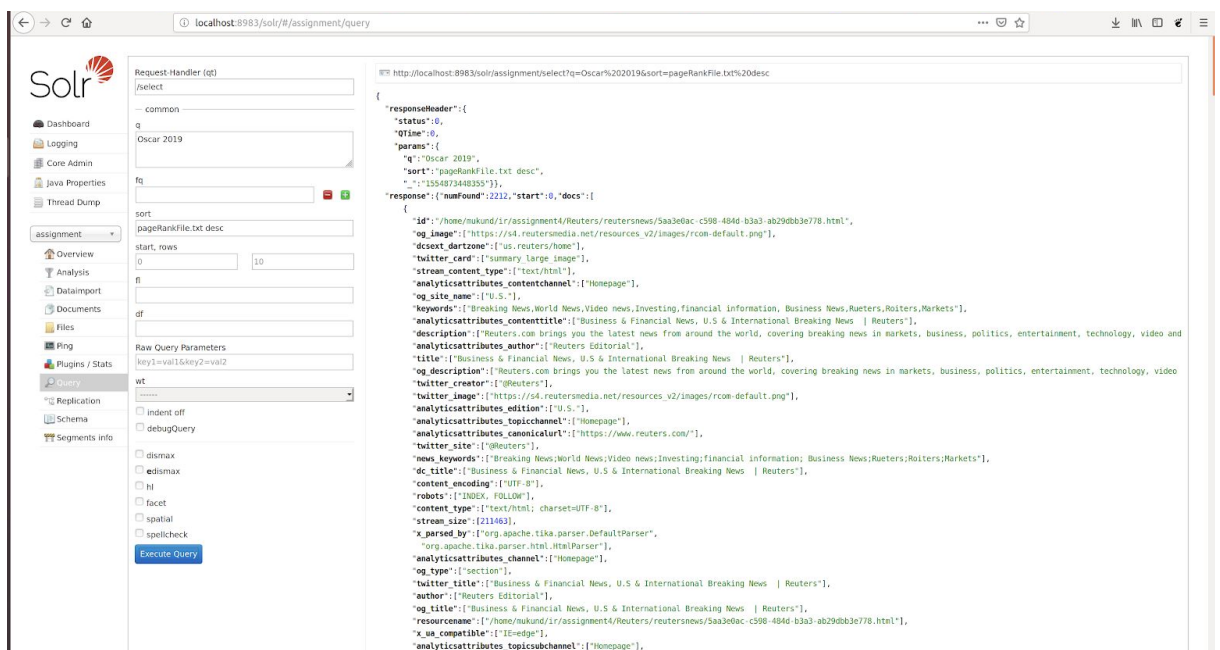
```
<fieldType name="external" keyField="id" defVal="">
class="solr.ExternalTextField" />
<field name="pageRankFile.txt" type="external" stored="false"
indexed="false" />
```

Added the following lines to `solr-7.7.1/server/solr/assignment/conf/solrconfig.xml` file,

```
<listener event="newSearcher" class=
"org.apache.solr.schema.ExternalTextFieldReloader" />
<listener event="firstSearcher" class=
"org.apache.solr.schema.ExternalTextFieldReloader" />
```
3. Reloaded the core by going inside Core Admin
4. In the query tab, in the `sort` field, added the external page rank file by using the command. Also specified the sort order as descending,
`pageRankFile.txt desc`
5. Screenshot of default query,



6. Screenshot of sample query,



Apache Server and PHP installation

1. Installed XAMPP version for Linux in order to run apache server and php code
2. Cloned the solr-php client in the htdocs folder
 - a. `git clone https://github.com/PTCInc/solr-php-client.git`
3. Used the template code given in the homework description

4. Added code for radio button (to choose between Lucene and page rank)
 - a. For Lucene, the logic would be,
`$results = $solr->search($query, 0, $limit);`
 - b. For page rank, the logic would be,
`$additionalParameters = array('sort' => 'pageRankFile.txt desc');`
`$results = $solr->search($query, 0, $limit, $additionalParameters);`
5. Added code to display the response of the search properly in tabular format
6. Added code to handle null description, title and url

Why some pages have higher PageRank values than others?

Some pages have higher PageRank than the others mainly because there are more number of pages linking to that page. PageRank can be considered as a vote, by all the other pages denoting how important a page is. If there is a link to a page, it is counted as a vote of support. If not, there is no support. In general, a PageRank algorithm is based on the # of incoming links and the weight of the linking page i.e., if a web page has an incoming link from a page with higher page rank, then its page rank will also be higher. In this assignment, edge list for a directed graph was computed and the external page rank file was computed using PageRank() function provided by NetworkX python library. This method computes a ranking of the nodes in the graph G based on the structure of the incoming links. This method also works on the same logic as briefed above.

Flow for one single query - Venezuela

Screenshot of initial page



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



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

localhost/ir/displayResultsForSearch.php?q=Venezuela&search_algo=lucene

Search Engine Comparison Using Solr

Search Term:

☒ Lucene ☐ Page Rank

Results 1 - 10 of 1709:

1	<p>Title : How Russia sank billions of dollars into Venezuelan quicksand URL : https://www.reuters.com/investigates/special-report/venezuela-russia-rosneft/ ID : /home/mukund/ir/assignment4/Reuters/reutersnews/3ac186d9-6b06-4550-a2c1-3a83ffd9cb84.html Description : A Russian oil company spent a fortune in Venezuela despite suspecting it was being ripped off. Sources say the Kremlin wanted to help its South American ally.</p>
2	<p>Title : Special Report: How Russia sank billions of dollars into Venezuela quicksand Reuters URL : https://www.reuters.com/article/us-venezuela-russia-rosneft-special-repo-idUSKCN1QV1HN ID : /home/mukund/ir/assignment4/Reuters/reutersnews/2d57e926-437b-4249-98be-1f0d359d814.html Description : At the end of 2015, managers at Rosneft, the Russian state-controlled oil firm, ...</p>
3	<p>Title : U.S. considers sanctions to restrict Visa, Mastercard in Venezuela: official Reuters URL : https://www.reuters.com/article/us-usa-venezuela-sanctions-idUSKCN1QV2VS ID : /home/mukund/ir/assignment4/Reuters/reutersnews/e49e96d0-9a07-4890-9c5d-302862160ed0.html Description : The United States is considering imposing financial sanctions that could prohibi...</p>
4	<p>Title : U.S. envoy says Venezuela oil production dropping steadily Reuters URL : https://www.reuters.com/article/us-venezuela-politics-usa-abrams-idUSKCN1QW2HM ID : /home/mukund/ir/assignment4/Reuters/reutersnews/c725f3e2-eda2-47a6-97b2-8840c3ad3e89.html Description : The U.S. special representative for Venezuela said on Friday that Venezuela's ...</p>
5	<p>Title : Shipbroker McQuilling ends tanker contracts with Venezuela's PDVSA Reuters URL : https://www.reuters.com/article/us-venezuela-politics-tankers-idUSKCN1QW2L6 ID : /home/mukund/ir/assignment4/Reuters/reutersnews/5434ebb-7882-4ae9-9962-83ed792b2306.html Description : A second maritime firm has told Venezuela's state-run oil firm Petroleos de...</p>
6	<p>Title : Billions spent, little to show Reuters URL : https://www.reuters.com/article/us-venezuela-russia-rosneft-money-idUSKCN1QV1HS ID : /home/mukund/ir/assignment4/Reuters/reutersnews/8930da40-6b0b-4276-9c99-83217309951f.html Description : Reuters estimates that after sinking more than \$9 billion into Venezuela in loan...</p>
7	<p>Title : Columns Reuters.com URL : https://www.reuters.com/journalists/nelson-bocanegra ID : /home/mukund/ir/assignment4/Reuters/reutersnews/c6803977-3c52-4814-9936-8870c797c68b.html Description : Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.</p>
8	<p>Title : Columns Reuters.com URL : https://www.reuters.com/journalists/lesley-wroughton ID : /home/mukund/ir/assignment4/Reuters/reutersnews/57d346b9-8ffc-40f8-a9a8-194082e1d149.html Description : Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.</p>
9	<p>Title : Columns Reuters.com URL : https://www.reuters.com/journalists/roberta-rampton ID : /home/mukund/ir/assignment4/Reuters/reutersnews/d12d761f-755c-41d2-a68d-565f8009dcf2.html Description : Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.</p>

Screenshot of page which shows the results for PageRank

localhost/ir/displayResultsForSearch.php?q=Venezuela&search_algo=page_rank

Search Engine Comparison Using Solr

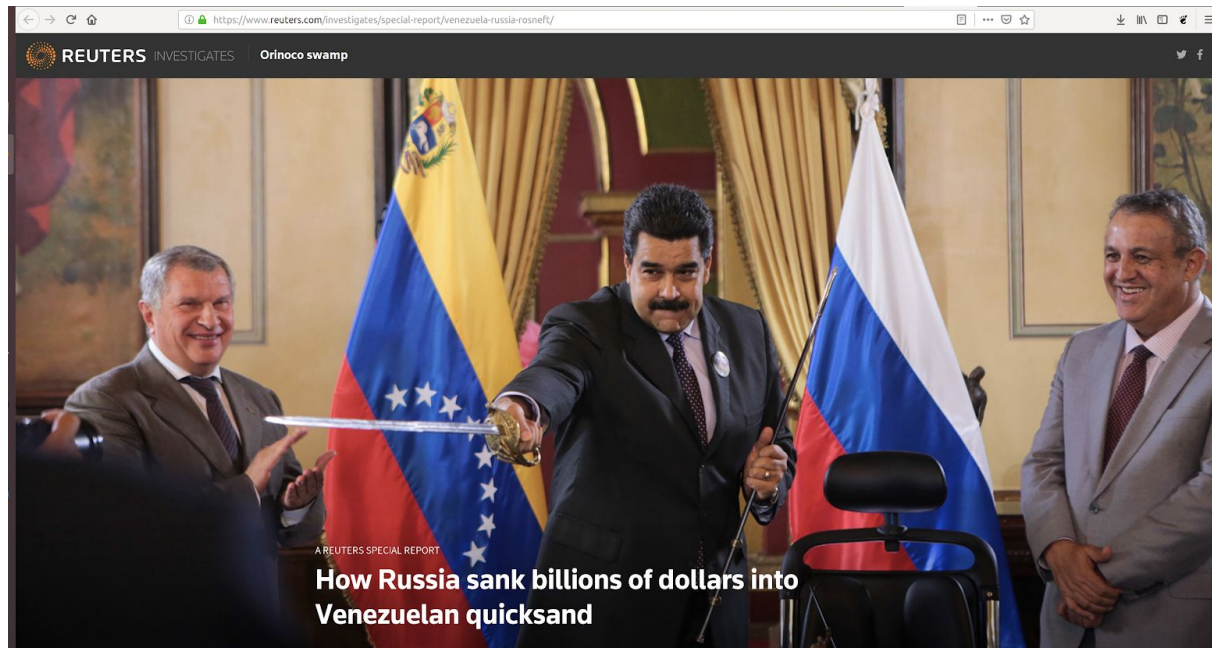
Search Term:

☐ Lucene ☒ Page Rank

Results 1 - 10 of 1709:

1	<p>Title : Business & Financial News, U.S. & International Breaking News Reuters URL : https://www.reuters.com/ ID : /home/mukund/ir/assignment4/Reuters/reutersnews/5aa3e0ac-c598-484d-b3a3-ab29d8b3e778.html Description : Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.</p>
2	<p>Title : Politics & Political News Reuters URL : https://www.reuters.com/politics ID : /home/mukund/ir/assignment4/Reuters/reutersnews/b3557cce-8ba0-43d4-a477-8334f685745a.html Description : Reuters.com is your politics news source for up to the minute, impartial political news coverage on the leaders, policies and agendas that shape our world.</p>
3	<p>Title : Markets & Finance News Reuters.com URL : https://www.reuters.com/finance/markets ID : /home/mukund/ir/assignment4/Reuters/reutersnews/dcca6198-104a-469f-9b07-e324bcae4275.html Description : Connect to the most current information on stocks and bonds on Reuters.com. Finance and investing news on NYSE, Dow Jones, S&P, and Nasdaq at Reuters.com.</p>
4	<p>Title : World News Reuters.com URL : https://www.reuters.com/news/world ID : /home/mukund/ir/assignment4/Reuters/reutersnews/ee5caaa-d84a-4cb7-b670-f042b665123a.html Description : Find late breaking world news from every corner of the globe at Reuters.com, your online source for international news coverage.</p>
5	<p>Title : Breakingviews Video - Reuters.com URL : https://www.reuters.com/video/breakingviews ID : /home/mukund/ir/assignment4/Reuters/reutersnews/19bb0d07-8fba-496c-ad7c-8608353d5f8b.html Description : After scandal enveloped a member of South Korean boy band Big Bang, shares in the industry's talent agencies plummeted this week. Robyn Mak and Jeffrey Goldfarb discuss how YouTube and Spotify should be able to help the popular music export pick up...</p>
6	<p>Title : Top News & Business Videos - Reuters.com URL : https://www.reuters.com/video ID : /home/mukund/ir/assignment4/Reuters/reutersnews/ca69ffcd-c30f-45bb-bd16-100a7ee7ee14.html Description : Reuters has learned that when he was growing up in El Paso, Texas, the kid who would grow up to be Democratic presidential hopeful Beto O'Rourke was a member of one of the earliest - and most influential - hacker collectives.</p>
7	<p>Title : Reuters Editorial Leadership Reuters URL : https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE25D ID : /home/mukund/ir/assignment4/Reuters/reutersnews/b7e1d031-d9dc-4acb-a788-5b5adda4f566.html Description : Stephen J. Adler Editor-in-Chief</p>
8	<p>Title : Special Reports from Reuters journalists around the world URL : https://www.reuters.com/investigates/section/homepage/ ID : /home/mukund/ir/assignment4/Reuters/reutersnews/59797e64-1360-4c9e-9cce-b68ab6e3b0b3.html Description : The latest Reuters investigations and other in-depth stories, plus a full archive of the past five years</p>
9	<p>Title : Rates & Bonds URL : https://www.reuters.com/markets/bonds</p>

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



URLs of top ten results produced by Lucene and Page Rank

Venezuela

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/investigates/special-report/venezuela-russia-rosneft/	https://www.reuters.com/
2	https://www.reuters.com/article/us-venezuela-russia-rosneft-special-report-idUSKCN1QV1HN	https://www.reuters.com/politics
3	https://www.reuters.com/article/us-usa-venezuela-sanctions-idUSKCN1QV2VS	https://www.reuters.com/finance/markets
4	https://www.reuters.com/article/us-venezuela-politics-usa-abrams-idUSKCN1QW2HM	https://www.reuters.com/news/world

5	https://www.reuters.com/article/us-venezuela-politics-tankers-idUSKCN1QW2L6	https://www.reuters.com/video/breakingviews
6	https://www.reuters.com/article/us-venezuela-russia-rosneft-money-idUSKCN1QV1HS	https://www.reuters.com/video
7	https://www.reuters.com/journalists/nelson-bocanegra	https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE2SD
8	https://www.reuters.com/journalists/lesley-wroughton	https://www.reuters.com/investigates/section/homepage/
9	https://www.reuters.com/journalists/roberta-rampton	https://www.reuters.com/markets/bonds
10	https://www.reuters.com/article/us-venezuela-politics-usa-banks-idUSKCN1QN2QE	https://www.reuters.com/markets/stocks

Senate

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/politics
2	https://www.reuters.com/video/2019/03/04/senate-has-votes-to-block-border-emergen?videoid=521958334&videoChannel=-10465	https://www.reuters.com/video/breakingviews
3	https://www.reuters.com/video/2019/03/05/mcconnell-senate-has-votes-to-block-bord?videoid=521939215&videoChannel=-10465	https://www.reuters.com/video

4	https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N	https://www.reuters.com/news/media
5	https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N	https://www.reuters.com/places/mexico
6	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	https://www.reuters.com/news/pictures
7	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	https://www.reuters.com/subjects/sustainable-business
8	https://www.reuters.com/politics	https://www.reuters.com/video/2019/03/14/orourke-hits-the-ground-running-in-iowa?videoId=525923095&videoChannel=1003
9	https://www.reuters.com/article/us-congo-politics-idUSKCN1QW28R	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK
10	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV00E	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H

Democrats

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/article/us-usa-election-medicare-idUSKBN1QU189	https://www.reuters.com/
2	https://www.reuters.com/article/us-usa-election-medicare-idUSKBN1QU189	https://www.reuters.com/politics
3	https://www.reuters.com/article/us-usa-congress-whitaker-idUSKCN1PX1LD	https://www.reuters.com/breakingviews

4	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoid=520293110&videoChannel=13976	https://www.reuters.com/subjects/supreme-court
5	https://www.reuters.com/article/us-usa-election-biden-idUSKCN1QH1EX	https://www.reuters.com/news/archive/politicsNews
6	https://www.reuters.com/article/us-usa-trump-budget-idUSKBN1QS12D	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK
7	https://www.reuters.com/video/2019/03/05/the-witch-hunt-continues-trump-on-democr?videoid=522293719&videoChannel=13976	https://www.reuters.com/article/us-usa-election-sanders-idUSKCN1QW2ZJ
8	https://www.reuters.com/article/us-usa-trump-russia-congress-idUSKCN1QV23E	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H
9	https://www.reuters.com/politics	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY
10	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-election-booker-idUSKCN1QW2P4

Republicans

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/politics
2	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoid=520293110&videoChannel=13976	https://www.reuters.com/news/us

3	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/news/archive/politicsNews
4	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK
5	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY
6	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C
7	https://www.reuters.com/article/us-california-fires-trump-idUSKCN1P31ND	https://www.reuters.com/article/us-usa-immigration-wall-insight-idUSKCN1QW1GR
8	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	https://www.reuters.com/article/us-usa-trump-russia-congress-idUSKCN1QV23E
9	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8
10	https://www.reuters.com/article/us-usa-politics-trump-idUSKCN1QH29L	https://www.reuters.com/article/us-usa-court-census-idUSKCN1QW2MC

Patriot Movement

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/article/us-usa-trump-turkey-idUSKCN1QV328	https://www.reuters.com/markets/bonds
2	https://www.reuters.com/article/us-france-protests-idUSKCN1Q325A	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C

3	https://www.reuters.com/article/us-britain-climatechange-education-idUSKCN1QW2QU	https://www.reuters.com/article/canada-forex-idUSL1N2121OB
4	https://www.reuters.com/article/uk-people-rkelly-metoo-idUKKCN1QC0V7	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S
5	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/article/us-slovakia-election-president-idUSKCN1QW2Z3
6	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/article/us-climate-change-youth-usa-idUSKCN1QW1BE
7	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/article/us-climate-change-youth-usa-idUSKCN1QW1BE
8	https://www.reuters.com/article/us-italy-china-analysis-idUSKCN1QW1E2	https://www.reuters.com/investigates/special-report/myanmar-suukyi-history/
9	https://www.reuters.com/article/us-italy-china-analysis-idUSKCN1QW1E2	https://www.reuters.com/article/us-auction-christies-georgemichael-art-idUSKCN1QP1EY
10	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C	https://www.reuters.com/article/us-italy-china-analysis-idUSKCN1QW1E2

Oscar 2019

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/article/us-award-s-oscar-women-idUSKCN1Q7130	https://www.reuters.com/
2	https://www.reuters.com/article/us-award-s-oscar-diversity-idUSKCN1QE0N4	https://www.reuters.com/tools/mobile
3	https://www.reuters.com/article/us-avian-caholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/politics
4	https://www.reuters.com/article/us-avian-caholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/finance

5	https://www.reuters.com/article/us-avian-caholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/finance/markets
6	https://www.reuters.com/journalists/pei-li	https://www.reuters.com/news/world
7	https://www.reuters.com/article/us-britain-lgbt-celebrities-idUSKCN1QW2KY	https://www.reuters.com/video/breakingviews
8	https://www.reuters.com/theWire	https://www.reuters.com/video
9	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI	https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE2SD
10	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI	https://www.reuters.com/investigates/section/homepage/

Channel

Rank	Solr-Lucene	External Page Rank using NetworkX
1	https://www.reuters.com/video/2019/01/28/brooklyn-grocer-goes-packaging-free?videoid=508863563&videoChannel=118172	https://www.reuters.com/politics
2	https://www.reuters.com/video/2019/02/27/cohen-dishes-on-donald-jr-in-testimony?videoid=519943059&videoChannel=13976	https://www.reuters.com/video/breakingviews
3	https://www.reuters.com/video/2019/03/11/exclusive-trump-to-ask-for-86-billion-fo?videoid=524332535&videoChannel=101	https://www.reuters.com/video
4	https://www.reuters.com/video/2019/03/13/breakingviews-tv-peltz-jackpot?videoid=525533725&videoChannel=117766	https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE2SD

5	https://www.reuters.com/video/2019/03/11/exclusive-trump-to-ask-for-86-billion-fo?videoid=524332535&videoChannel=-10465	https://www.reuters.com/markets/bonds
6	https://www.reuters.com/video/2019/03/05/the-witch-hunt-continues-trump-on-democr?videoid=522293719&videoChannel=13976	https://www.reuters.com/news/entertainment/arts
7	https://www.reuters.com/video/2019/03/15/breakingviews-tv-k-pop-idles?videoid=526312336&videoChannel=117766	https://www.reuters.com/news/media
8	https://www.reuters.com/video/2019/03/06/japans-demon-festival-revised-after-12-y?videoid=522528399&videoChannel=-10464	https://www.reuters.com/subjects/supreme-court
9	https://www.reuters.com/video/2019/02/21/macron-announces-anti-semitism-measures?videoid=517540456&videoChannel=13421	https://www.reuters.com/news/lifestyle
10	https://www.reuters.com/video/2019/01/15/greenwashing-warning-amid-renewable-push?videoid=503830819&videoChannel=118285	https://www.reuters.com/places/china

Wall

Rank	Solr-Lucene	External Page Rank using NetworkX
1	http://www.reuters.com/investigates/section/wall-streets-way/	https://www.reuters.com/
2	https://www.reuters.com/article/us-usa-trump-budget-wall-exclusive-idUSKBN1QR0CW	https://www.reuters.com/politics

3	https://www.reuters.com/article/us-usa-immigration-wall-insight-idUSKCN1QW1GR	https://www.reuters.com/finance
4	https://www.reuters.com/politics	https://www.reuters.com/finance/markets
5	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/video/breaking-views
6	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/video
7	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/investigates/section/homepage/
8	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/finance/wealth
9	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/finance/funds
10	https://www.reuters.com/journalists/noel-randewich	https://www.reuters.com/finance/EarningsUS

Overlap graph

Query	# of overlaps
Venezuela	0
Senate	2
Democrats	2
Republicans	3
Patriot Movement	3

Oscar 2019	0
Channel	0
Wall	1

of overlaps between Solr-Lucene and External Page Rank

