Neekita Salvankar - USC ID: 8591-3366-93

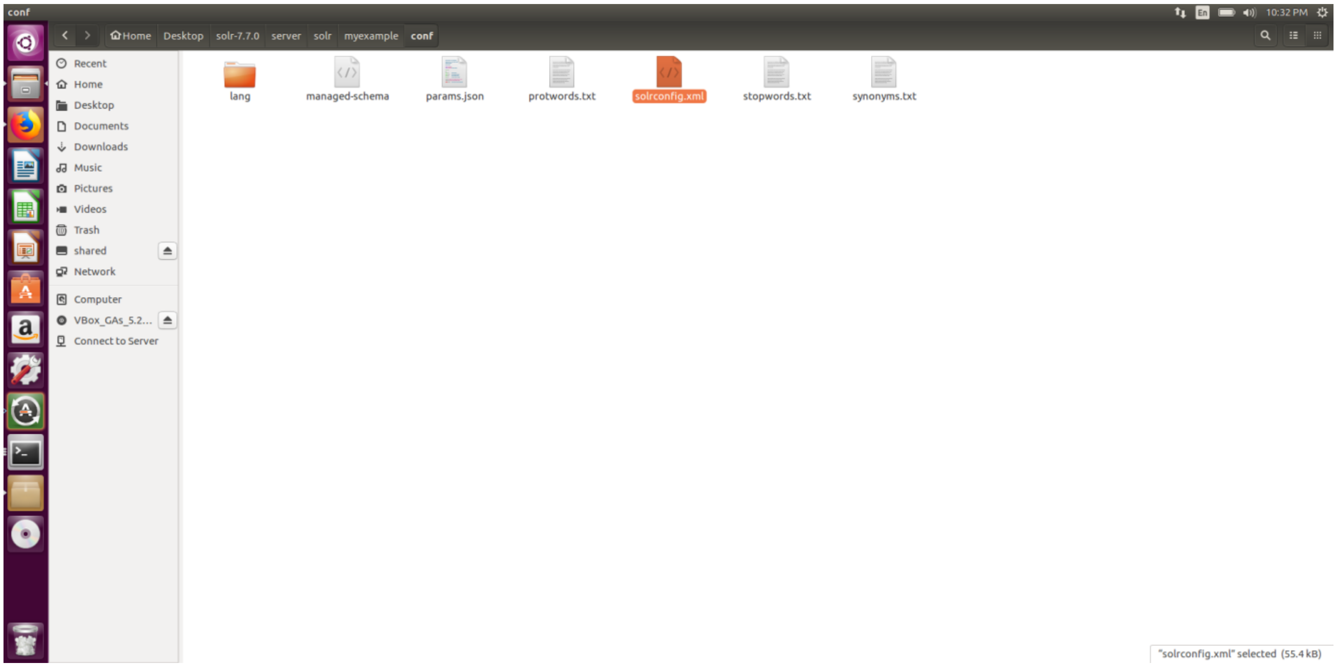
Homework 4 Report (All steps and screenshots are explained in detailed manner)

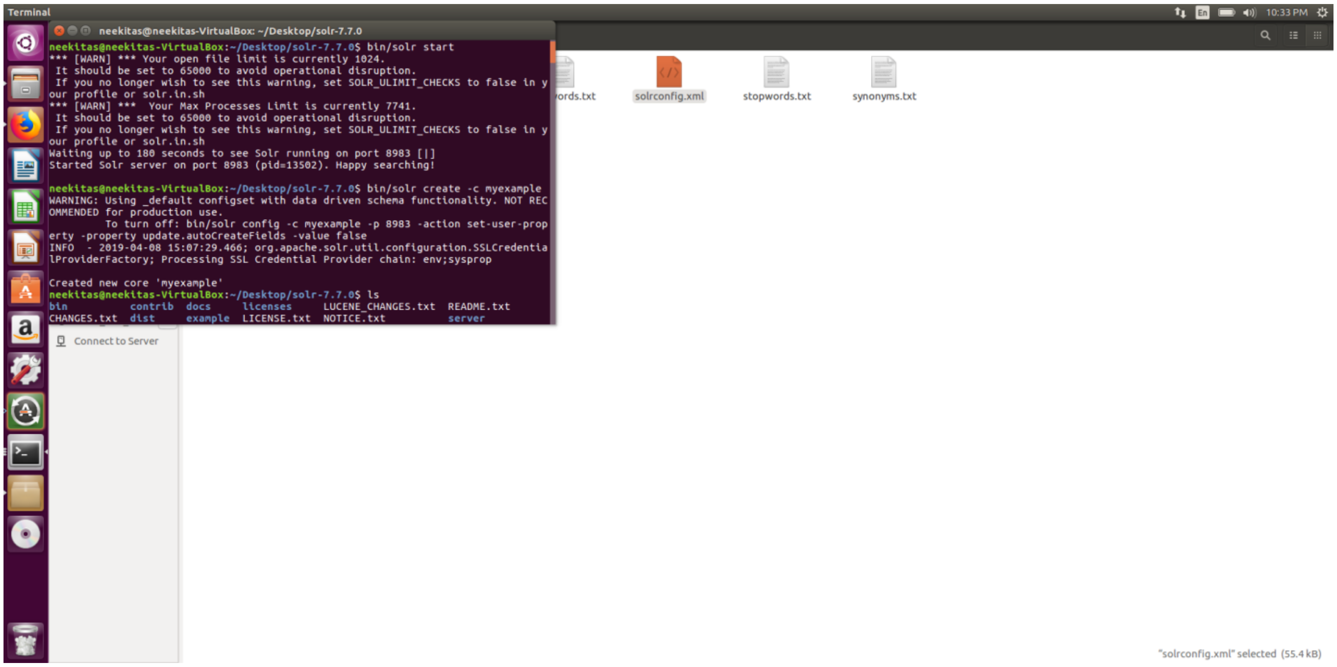
1. Installation of Virtual Machine and Ubuntu

a. Installed and completed set up of Virtual Machine and Ubuntu using the tutorial provided on class website

2. Downloaded and installed Solr

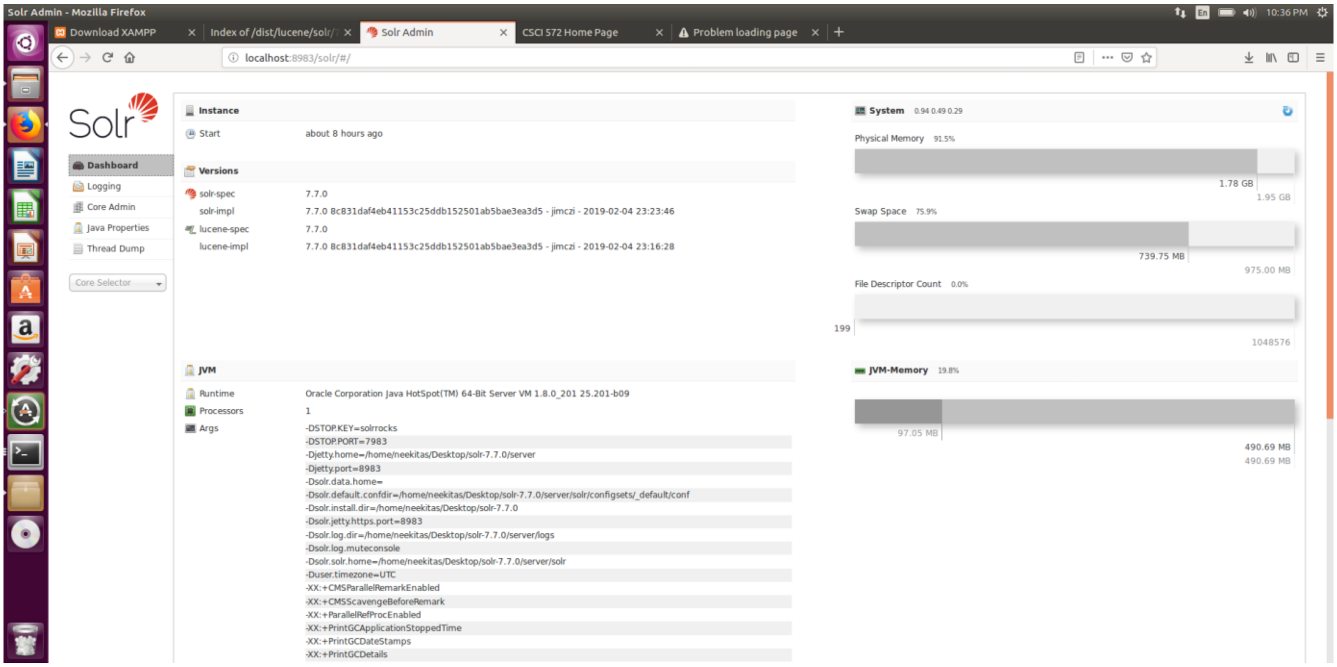
a. Completed the download and installation of Solr using the tutorial provided on class website. Then I started running Solr using “bin/solr start” command . I then created my own core named “myexample” .




Terminal
10:33 PM
neekitas@neekitas-VirtualBox: ~/Desktop/solr-7.7.0
ords.txt
solrconfig.xml
stopwords.txt
synonyms.txt
Connect to Server
"solrconfig.xml" selected (55.4 kB)

```
neekitas@neekitas-VirtualBox:~/Desktop/solr-7.7.0$ bin/solr start
*** [WARN] *** Your open file limit is currently 1024.
 It should be set to 65000 to avoid operational disruption.
 If you no longer wish to see this warning, set SOLR_ULIMIT_CHECKS to false in y
our profile or solr.in.sh
*** [WARN] ***  Your Max Processes Limit is currently 7741.
 It should be set to 65000 to avoid operational disruption.
 If you no longer wish to see this warning, set SOLR_ULIMIT_CHECKS to false in y
our profile or solr.in.sh
Waiting up to 180 seconds to see Solr running on port 8983 [|]
Started Solr server on port 8983 (pid=13502). Happy searching!

neekitas@neekitas-VirtualBox:~/Desktop/solr-7.7.0$ bin/solr create -c myexample
WARNING: Using _default configset with data driven schema functionality. NOT REC
OMMENDED for production use.
         To turn off: bin/solr config -c myexample -p 8983 -action set-user-prop
erty -property update.autoCreateFields -value false
INFO  - 2019-04-08 15:07:29.466; org.apache.solr.util.configuration.SSLCredentia
lProviderFactory; Processing SSL Credential Provider chain: env;sysprop

Created new core 'myexample'
neekitas@neekitas-VirtualBox:~/Desktop/solr-7.7.0$ ls
bin          contrib  docs     licenses     LUCENE_CHANGES.txt  README.txt
CHANGES.txt  dist     example  LICENSE.txt  NOTICE.txt          server
```

b. I typed “localhost:8983” on my Mozilla browser to run the Solr admin and loaded my core to check everything worked fine

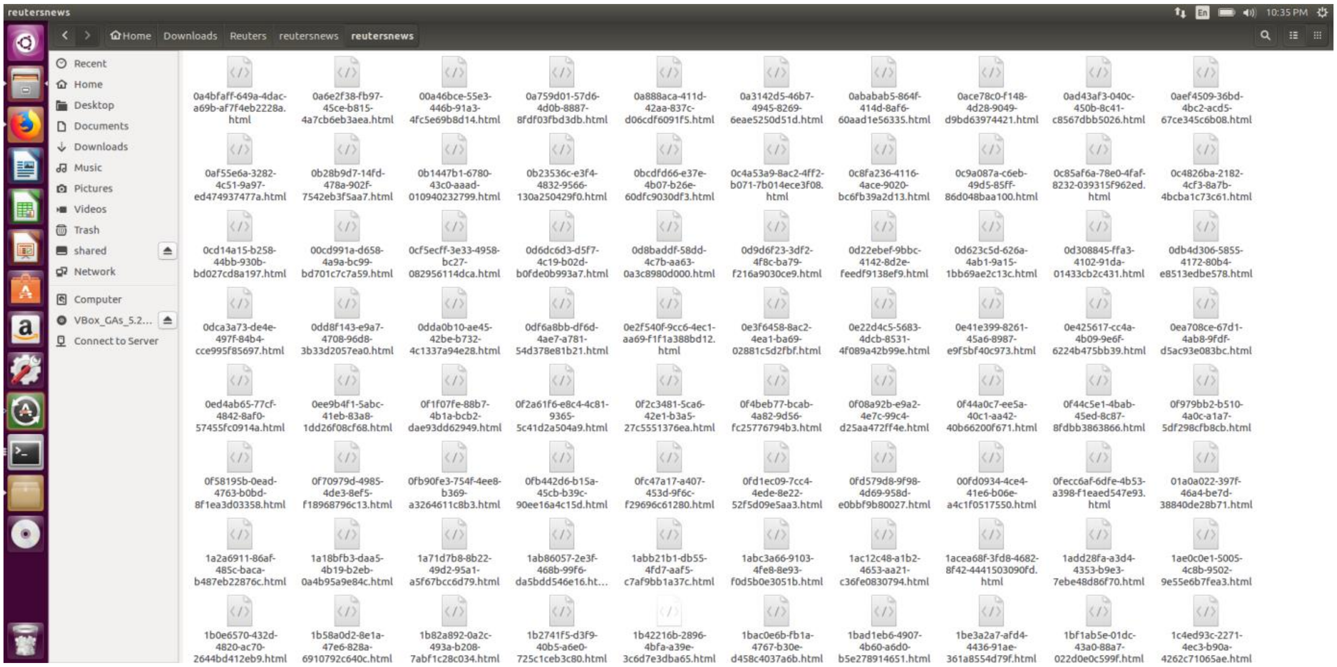


```
Instance
Start          about 8 hours ago

Versions
solr-spec      7.7.0
solr-impl      7.7.0 8c831daf4eb41153c25ddb152501ab5bae3ea3d5 - jimczi - 2019-02-04 23:23:46
lucene-spec    7.7.0
lucene-impl    7.7.0 8c831daf4eb41153c25ddb152501ab5bae3ea3d5 - jimczi - 2019-02-04 23:16:28

JVM
Runtime        Oracle Corporation Java HotSpot(TM) 64-Bit Server VM 1.8.0_201 25.201-b09
Processors     1
Args           -DSTOP.KEY=solrrocks
               -DSTOP.PORT=7983
               -Djetty.home=/home/neekitas/Desktop/solr-7.7.0/server
               -Djetty.port=8983
               -Dsolr.data.home=
               -Dsolr.default.confdir=/home/neekitas/Desktop/solr-7.7.0/server/solr/configsets/_default/conf
               -Dsolr.install.dir=/home/neekitas/Desktop/solr-7.7.0
               -Dsolr.jetty.https.port=8983
               -Dsolr.log.dir=/home/neekitas/Desktop/solr-7.7.0/server/logs
               -Dsolr.log.muteconsole
               -Dsolr.solr.home=/home/neekitas/Desktop/solr-7.7.0/server/solr
               -Duser.timezone=UTC
               -XX:+CMSParallelRemarkEnabled
               -XX:+CMSScavengeBeforeRemark
               -XX:+ParallelRefProcEnabled
               -XX:+PrintGCApplicationStoppedTime
               -XX:+PrintGCDateStamps
               -XX:+PrintGCDetails

System   0.94 0.49 0.29
Physical Memory   91.5%
  1.78 GB / 1.95 GB
Swap Space   75.9%
  739.75 MB / 975.00 MB
File Descriptor Count   0.0%
  199 / 1048576
JVM-Memory   19.8%
  97.05 MB / 490.69 MB / 490.69 MB
```

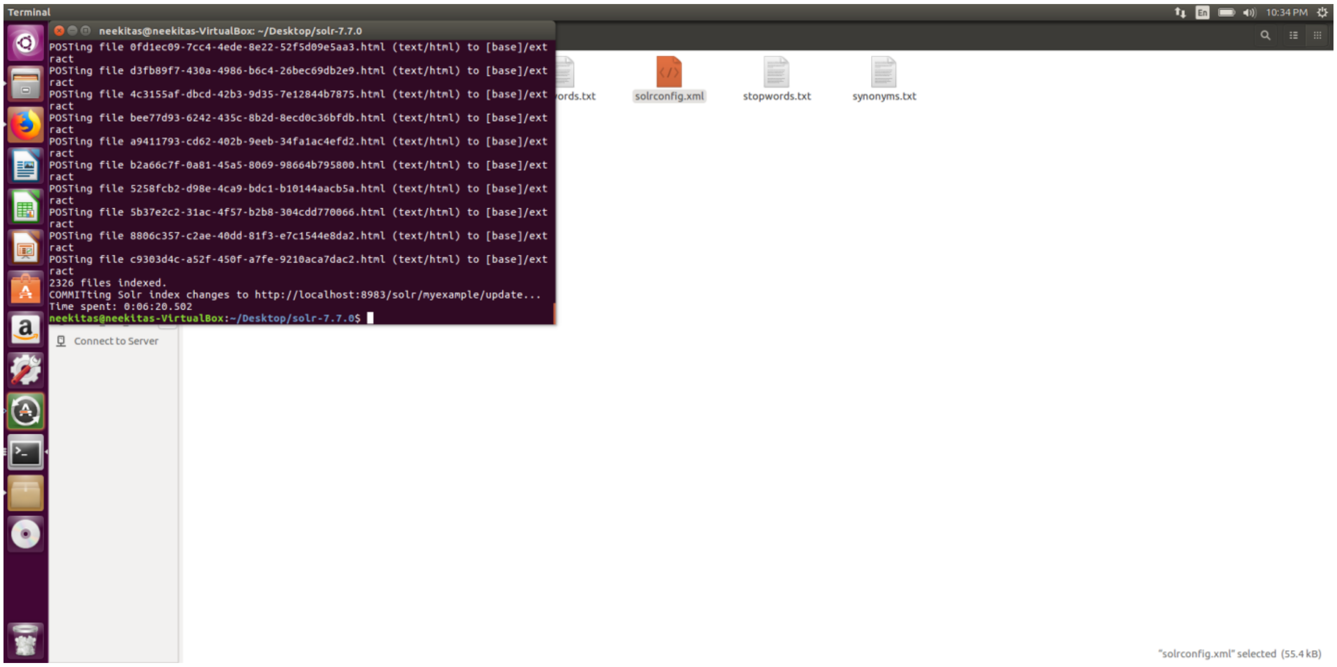
c. I made changes to 2 conf files :

1. managed-schema file : uncommented copyfield source tag
2. solrconfig.xml file: to default querying from the field text by defining default field in requestHandler and uncommented the str element with name “df”, to specify the default query field to be “_text_”

d. Downloaded the folder in Google Drive containing files for REUTER news as per my USC ID

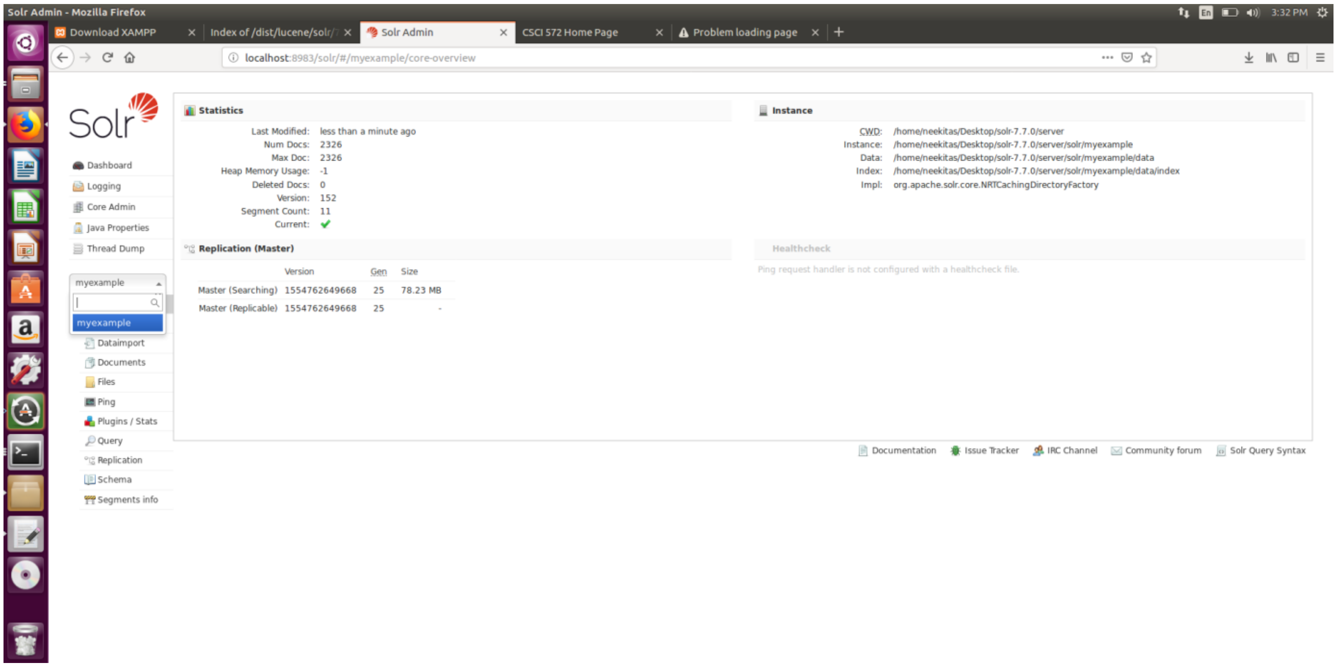

reutersnews
10:35 PM
Home
Downloads
Reuters
reutersnews
reutersnews
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
shared
Network
Computer
VBox_GAs_5.2...
Connect to Server
0a4bfaff-649a-4dac-a69b-af7f4eb2228a.html
0a6e2f38-fb97-45ce-b815-4a7cb6eb3aea.html
00a46bce-55e3-446b-91a3-4fc5e69b8d14.html
0a759d01-57d6-4d0b-8887-8fdf03fbd3db.html
0a888aca-411d-42aa-837c-d06cdf6091f5.html
0a3142d5-46b7-4945-8269-6eae5250d51d.html
0abababs5-864f-414d-8af6-60aad1e56335.html
0ace78c0-f148-4d28-9049-d9bd63974421.html
0ad43af3-040c-450b-8c41-c8567dbb5026.html
0aef4509-36bd-4bc2-acd5-67ce345c6b08.html
0af55e6a-3282-4c51-9a97-ed474937477a.html
0b28b9d7-14fd-478a-902f-7542eb3f5aa7.html
0b1447b1-6780-43c0-aaad-010940232799.html
0b23536c-e3f4-4832-9566-130a250429f0.html
0bcdfd66-e37e-4b07-b26e-60dfc9030df3.html
0c4a53a9-8ac2-4ff2-b071-7b014ece3f08.html
0c8fa236-4116-4ace-9020-bc6fb39a2d13.html
0c9a087a-c6eb-49d5-85ff-86d048baa100.html
0c85af6a-78e0-4faf-8232-039315f962ed.html
0c4826ba-2182-4cf3-8a7b-4bcba1c73c61.html
0cd14a15-b258-44bb-930b-bd027cd8a197.html
00cd991a-d658-4a9a-bc99-bd701c7c7a59.html
0cf5ecff-3e33-4958-bc27-082956114dca.html
0d6dc6d3-d5f7-4c19-b02d-b0fde0b993a7.html
0d8baddf-58dd-4c7b-aa63-0a3c8980d000.html
0d9d6f23-3df2-4f8c-ba79-f216a9030ce9.html
0d22ebef-9bbc-4142-8d2e-feedf9138ef9.html
0d623c5d-626a-4ab1-9a15-1bb69ae2c13c.html
0d308845-ffa3-4102-91da-01433cb2c431.html
0db4d306-5855-4172-80b4-e8513edbe578.html
0dca3a73-de4e-497f-84b4-cce995f85697.html
0dd8f143-e9a7-4708-96d8-3b33d2057ea0.html
0dda0b10-ae45-42be-b732-4c1337a94e28.html
0df6a8bb-df6d-4ae7-a781-54d378e81b21.html
0e2f540f-9cc6-4ec1-aa69-f1f1a388bd12.html
0e3f6458-8ac2-4ea1-ba69-02881c5d2fbf.html
0e22d4c5-5683-4dcb-8531-4f089a42b99e.html
0e41e399-8261-45a6-8987-e9f5bf40c973.html
0e425617-cc4a-4b09-9e6f-6224b475bb39.html
0ea708ce-67d1-4ab8-9fdf-d5ac93e083bc.html
0ed4ab65-77cf-4842-8af0-57455fc0914a.html
0ee9b4f1-5abc-41eb-83a8-1dd26f08cf68.html
0f1f07fe-88b7-4b1a-bcb2-dae93dd62949.html
0f2a61f6-e8c4-4c81-9365-5c41d2a504a9.html
0f2c3481-5ca6-42e1-b3a5-27c5551376ea.html
0f4beb77-bcab-4a82-9d56-fc25776794b3.html
0f08a92b-e9a2-4e7c-99c4-d25aa472ff4e.html
0f44a0c7-ee5a-40c1-aa42-40b66200f671.html
0f44c5e1-4bab-45ed-8c87-8fdbb3863866.html
0f979bb2-b510-4a0c-a1a7-5df298cfb8cb.html
0f58195b-0ead-4763-b0bd-8f1ea3d03358.html
0f70979d-4985-4de3-8ef5-f18968796c13.html
0fb90fe3-754f-4ee8-b369-a3264611c8b3.html
0fb442d6-b15a-45cb-b39c-90ee16a4c15d.html
0fc47a17-a407-453d-9f6c-f29696c61280.html
0fd1ec09-7cc4-4ede-8e22-52f5d09e5aa3.html
0fd579d8-9f98-4d69-958d-e0bbf9b80027.html
00fd0934-4ce4-41e6-b06e-a4c1f0517550.html
0fecc6af-6dfe-4b53-a398-f1eaed547e93.html
01a0a022-397f-46a4-be7d-38840de28b71.html
1a2a6911-86af-485c-baca-b487eb22876c.html
1a18bfb3-daa5-4b19-b2eb-0a4b95a9e84c.html
1a71d7b8-8b22-49d2-95a1-a5f67bcc6d79.html
1ab86057-2e3f-468b-99f6-da5bdd546e16.ht...
1abb21b1-db55-4fd7-aaf5-c7af9bb1a37c.html
1abc3a66-9103-4fe8-8e93-f0d5b0e3051b.html
1ac12c48-a1b2-4653-aa21-c36fe0830794.html
1acea68f-3fd8-4682-8f42-4441503090fd.html
1add28fa-a3d4-4353-b9e3-7ebe48d86f70.html
1ae0c0e1-5005-4c8b-9502-9e55e6b7fea3.html
1b0e6570-432d-4820-ac70-2644bd412eb9.html
1b58a0d2-8e1a-47e6-828a-6910792c640c.html
1b82a892-0a2c-493a-b208-7abf1c28c034.html
1b2741f5-d3f9-40b5-a6e0-725c1ceb3c80.html
1b42216b-2896-4bfa-a39e-3c6d7e3dba65.html
1bac0e6b-fb1a-4767-b30e-d458c4037a6b.html
1bad1eb6-4907-4b60-a6d0-b5e278914651.html
1be3a2a7-afd4-4436-91ae-361a8554d79f.html
1bf1ab5e-01dc-43a0-88a7-022d0e0c599f.html
1c4ed93c-2271-4ec3-b90a-4262c71065ae.html

e. Indexed html files specified in the Reuter folder.


Terminal
10:34 PM
neekitas@neekitas-VirtualBox: ~/Desktop/solr-7.7.0
ords.txt
solrconfig.xml
stopwords.txt
synonyms.txt
Connect to Server
"solrconfig.xml" selected (55.4 kB)

```
POSTing file 0fd1ec09-7cc4-4ede-8e22-52f5d09e5aa3.html (text/html) to [base]/ext
ract
POSTing file d3fb89f7-430a-4986-b6c4-26bec69db2e9.html (text/html) to [base]/ext
ract
POSTing file 4c3155af-dbcd-42b3-9d35-7e12844b7875.html (text/html) to [base]/ext
ract
POSTing file bee77d93-6242-435c-8b2d-8ecd0c36bfdb.html (text/html) to [base]/ext
ract
POSTing file a9411793-cd62-402b-9eeb-34fa1ac4efd2.html (text/html) to [base]/ext
ract
POSTing file b2a66c7f-0a81-45a5-8069-98664b795800.html (text/html) to [base]/ext
ract
POSTing file 5258fcb2-d98e-4ca9-bdc1-b10144aacb5a.html (text/html) to [base]/ext
ract
POSTing file 5b37e2c2-31ac-4f57-b2b8-304cdd770066.html (text/html) to [base]/ext
ract
POSTing file 8806c357-c2ae-40dd-81f3-e7c1544e8da2.html (text/html) to [base]/ext
ract
POSTing file c9303d4c-a52f-450f-a7fe-9210aca7dac2.html (text/html) to [base]/ext
ract
2326 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/myexample/update...
Time spent: 0:06:20.502
neekitas@neekitas-VirtualBox:~/Desktop/solr-7.7.0$
```

f. Reloaded the core



Statistics

Last Modified: less than a minute ago
Num Docs: 2326
Max Doc: 2326
Heap Memory Usage: -1
Deleted Docs: 0
Version: 152
Segment Count: 11
Current: ✓

Replication (Master)

	Version	Gen	Size
Master (Searching)	1554762649668	25	78.23 MB
Master (Replicable)	1554762649668	25	-

Instance

CWD: /home/neekitas/Desktop/solr-7.7.0/server
Instance: /home/neekitas/Desktop/solr-7.7.0/server/solr/myexample
Data: /home/neekitas/Desktop/solr-7.7.0/server/solr/myexample/data
Index: /home/neekitas/Desktop/solr-7.7.0/server/solr/myexample/data/index
Impl: org.apache.solr.core.NRTCachingDirectoryFactory

Healthcheck

Ping request handler is not configured with a healthcheck file.

3. Creating EdgeList file

I wrote a Java program and used the Jsoup library to create a file with edgelists. I used Eclipse to develop the same.

4. Creating External PageRank file

a. Used network library of python to create an external pageRank file. The input was the edgelist file created in the previous step.
b. I loaded the web graph using read_edgeList function with edgelist and digraph a parameters.
c. Used the pagerank function with following parameters:
 alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight = ‘weight’ , dangling= None

5. I made changes to the conf files to add the external page rank files and reloaded the solr admin (as mentioned in the tutorial)

a. Managed-schema file :
Added the following:

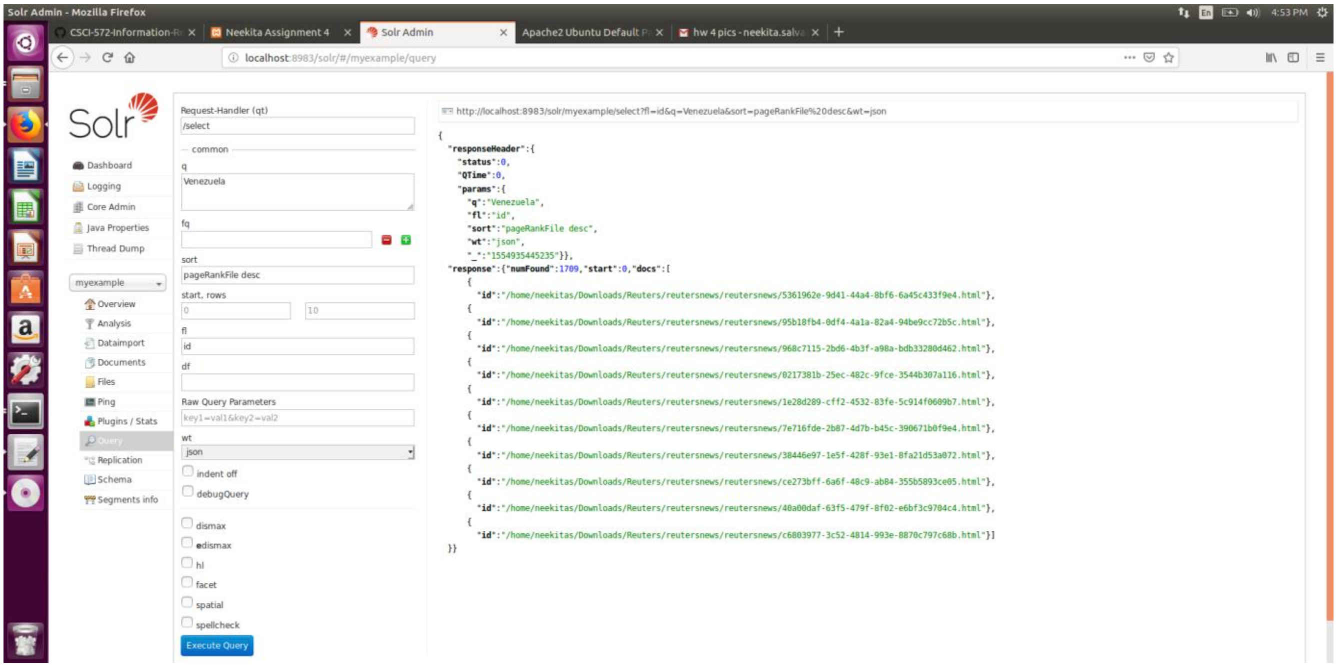
```
<fieldType name=”external” keyField=”id” defVal=”0”
class=”solr.ExternalFileField” />
<field name=”pageRankFile” type=”external” stored=”false”
indexed=”false” />
```

b. Solrconfig.xml file:
Added the following:

```
<listener event = ″newSearcher″ class =
″org.apache.solr.schema.ExternalFileFieldReloader″/>
<listener event = ″firstSearcher″ class =
″org.apache.solr.schema.ExternalFileFieldReloader″/>
```

6. I reloaded the core and put in a sample query and got results both by default method and PageRank method

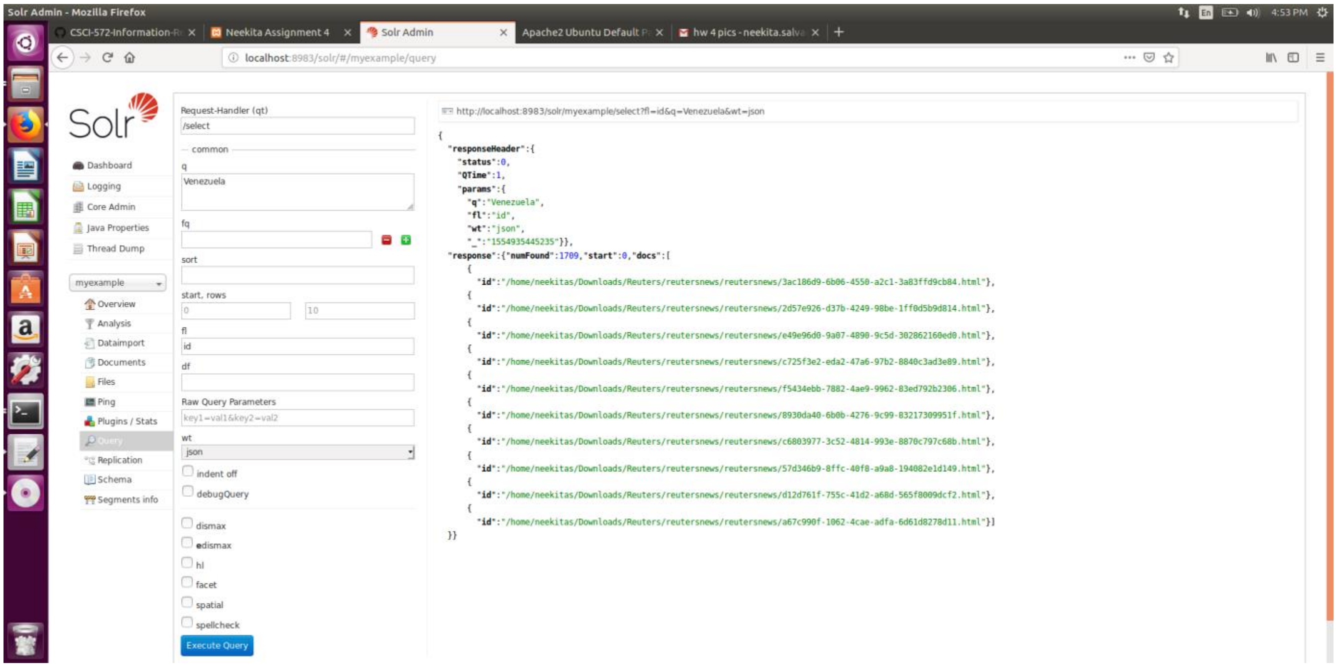
Below are the screenshots of both working:



Request-Handler (qt): /select
common
q: Venezuela
fq:
sort: pageRankFile desc
start, rows: 0 10
fl: id
df:
Raw Query Parameters: key1=val1&key2=val2
wt: json
□ indent off
□ debugQuery
□ dismax
□ edismax
□ hl
□ facet
□ spatial
□ spellcheck

http://localhost:8983/solr/myexample/select?fl=id&q=Venezuela&sort=pageRankFile%20desc&wt=json

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"Venezuela",
      "fl":"id",
      "sort":"pageRankFile desc",
      "wt":"json",
      "_":"1554935445235"}},
  "response":{"numFound":1709,"start":0,"docs":[
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/5361962e-9d41-44a4-8bf6-6a45c433f9e4.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/95b18fb4-0df4-4a1a-82a4-94be9cc72b5c.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/968c7115-2bd6-4b3f-a98a-bdb33280d462.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/0217381b-25ec-482c-9fce-3544b307a116.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/1e28d289-cff2-4532-83fe-5c914f0609b7.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/7e716fde-2b87-4d7b-b45c-390671b0f9e4.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/38446e97-1e5f-428f-93e1-8fa21d53a072.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/ce273bff-6a6f-48c9-ab84-355b5893ce05.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/40a00daf-63f5-479f-8f02-e6bf3c9704c4.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/c6803977-3c52-4814-993e-8870c797c68b.html"}]
  }}
```



Request-Handler (qt): /select

q: Venezuela

start, rows: 0, 10

fl: id

wt: json

http://localhost:8983/solr/myexample/select?fl=id&q=Venezuela&wt=json

```
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"Venezuela",
      "fl":"id",
      "wt":"json",
      "_":"1554935445235"}},
  "response":{"numFound":1709,"start":0,"docs":[
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/3ac186d9-6b06-4550-a2c1-3a83ffd9cb84.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/2d57e926-d37b-4249-98be-1ff0d5b9d814.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/e49e96d0-9a07-4890-9c5d-302862160ed0.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/c725f3e2-eda2-47a6-97b2-8840c3ad3e89.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/f5434ebb-7882-4ae9-9962-83ed792b2306.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/8930da40-6b0b-4276-9c99-83217309951f.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/c6803977-3c52-4814-993e-8870c797c68b.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/57d346b9-8ffc-40f8-a9a8-194082e1d149.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/d12d761f-755c-41d2-a68d-565f8009dcf2.html"},
      {
        "id":"/home/neekitas/Downloads/Reuters/reutersnews/reutersnews/a67c990f-1062-4cae-adfa-6d61d8278d11.html"}]
  }}
```

Below are the results of 8 queries and also graph of the overlap

1. **Venezuela**

No	Default	PageRank
1	https://www.reuters.com/investigates/special-report/venezuela-russia-rosneft/	https://www.reuters.com/article/us-ethiopia-airplane-airlines-idUSKCN1QW2P8
2	https://www.reuters.com/article/us-venezuela-russia-rosneft-special-repo-idUSKCN1QV1HN	https://www.reuters.com/article/us-barbie-anniversary-idUSKBN1QP2HG
3	https://www.reuters.com/article/us-usa-venezuela-sanctions-idUSKCN1QV2VS	https://www.reuters.com/news/archive/russia

4	https://www.reuters.com/article/us-venezuela-politics-usa-abrams-idUSKCN1QW2HM	https://www.reuters.com/article/aviva-ceo-idUSFWN20Q09G
5	https://www.reuters.com/article/us-venezuela-politics-tankers-idUSKCN1QW2L6	https://www.reuters.com/journalists/julia-symmes-cobb
6	https://www.reuters.com/article/us-venezuela-russia-rosneft-money-idUSKCN1QV1HS	https://www.reuters.com/article/us-soccer-spain-usa-idUSKCN1N652I
7	https://www.reuters.com/journalists/nelson-bocanegra	https://www.reuters.com/article/us-northkorea-dissidents-idUSKCN1QW2Z
8	https://www.reuters.com/journalists/lesley-wroughton	https://www.reuters.com/finance/currencies/quote?destAmt=&srcAmt=1&srcCurr=USD&destCurr=CNY
9	https://www.reuters.com/journalists/roberta-rampton	https://www.reuters.com/article/us-usa-trump-russia-deripaska-idUSKCN1QW2CA
10	https://www.reuters.com/article/us-venezuela-politics-usa-banks-idUSKCN1QN2QE	https://www.reuters.com/journalists/nelson-bocanegra

2. Senate

No	Default	PageRank
1	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK
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/2019/03/06/breakingviews-tv-carry-trades?videoId=522657667&videoChannel=117766

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/2019/02/01/images-of-january?videoId=510237001&videoChannel=118069
4	https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N	https://www.reuters.com/news/picture/photos-of-the-week-idUSRTX6R5YJ
5	https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N	https://www.reuters.com/article/us-usa-trump-russia-timing-explainer-idUSKBN1QU1E8
6	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	https://www.reuters.com/article/us-myanmar-journalists-reaction-factbox-idUSKBN1FU030
7	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	https://www.reuters.com/article/us-column-miller-socialsecurity-idUSKCN1PP1AN
8	https://www.reuters.com/politics	https://www.reuters.com/video/2019/03/15/tesla-launches-the-model-y-electric-cros?videoId=526236973&videoChannel=118169
9	https://www.reuters.com/article/us-congo-politics-idUSKCN1QW28R	https://www.reuters.com/politics
10	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV00E	https://www.reuters.com/article/us-usa-trump-russia-cohen-idUSKCN1QH0K6

3. Democrats

No	Default	PageRank
1	https://www.reuters.com/article/us-usa-election-medicare-idUSKBN1QU189	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK
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/article/us-usa-trump-russia-cohen-idUSKCN1QH0K6
4	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoId=520293110&videoChannel=13976	https://www.reuters.com/news/archive/politicsNews
5	https://www.reuters.com/article/us-usa-election-biden-idUSKCN1QH1EX	https://www.reuters.com/article/us-usa-politics-trump-idUSKCN1QH29L
6	https://www.reuters.com/article/us-usa-trump-budget-idUSKBN1QS12D	https://www.reuters.com/article/usa-election-orourke-idINKCN1QV130
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-census-ross-idUSKCN1QV27O
8	https://www.reuters.com/article/us-usa-trump-russia-congress-idUSKCN1QV23E	https://www.reuters.com/journalists/nathan-layne
9	https://www.reuters.com/politics	https://www.reuters.com/article/us-usa-election-booker-idUSKCN1QW2P4
10	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/germany-saudi-arms-airbus-idUSL5N20M7OT

4. Republicans

No	Default	PageRank
1	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK

2	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoId=520293110&videoChannel=13976	https://www.reuters.com/article/us-column-miller-socialsecurity-idUSKCN1PP1AN
3	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/politics
4	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-trump-russia-cohen-idUSKCN1QH0K6
5	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/news/archive/politicsNews
6	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-politics-trump-idUSKCN1QH29L
7	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	https://www.reuters.com/article/us-usa-census-ross-idUSKCN1QV27O
8	https://www.reuters.com/article/us-california-fires-trump-idUSKCN1P31ND	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoId=520293110&videoChannel=13976
9	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	https://www.reuters.com/article/us-usa-immigration-wall-insight-idUSKCN1QW1GR
10	https://www.reuters.com/article/us-usa-politics-trump-idUSKCN1QH29L	https://www.reuters.com/article/us-usa-court-census-idUSKCN1QW2MC

5. Patriot Movement

No	Default	PageRank

1	https://www.reuters.com/article/us-usa-trump-turkey-idUSKCN1QV328	"https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01
2	https://www.reuters.com/article/us-france-protests-idUSKCN1Q325A	https://www.reuters.com/journalists/fergal-smith
3	https://www.reuters.com/article/uk-people-rkelly-metoo-idUKKCN1QC0V7	https://www.reuters.com/article/uk-britain-sterling-idUSKBN1QS0YM
4	https://www.reuters.com/article/us-britain-climatechange-education-idUSKCN1QW2QU	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI
5	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	"https://www.reuters.com/article/us-taiwan-china-hongkong-idUSKCN1QW0NT
6	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/article/canada-forex-idUSL1N2121OB
7	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01	https://www.reuters.com/journalists/crispian-balmer
8	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C	https://www.reuters.com/article/us-usa-directors-diversity-idUSKCN1Q20E8
9	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C	https://www.reuters.com/journalists/jillian-kitchener
10	https://www.reuters.com/article/us-italy-china-analysis-idUSKCN1QW1E2	https://www.reuters.com/article/us-myanmar-rakhine-investment-idUSKCN1QB0HO

6. Oscar 2019

No	Default	PageRank
1	https://www.reuters.com/article/us-awards-oscars-women-idUSKCN1Q7130	https://www.reuters.com/article/us-ethiopia-airplane-airlines-idUSKCN1QW2P8
2	https://www.reuters.com/article/us-awards-oscars-diversity-idUSKCN1QE0N4	https://www.reuters.com/article/us-barbie-anniversary-idUSKBN1QP2HG
3	https://www.reuters.com/article/us-aviancaholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/news/archive/russia
4	https://www.reuters.com/article/us-aviancaholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/article/aviva-ceo-idUSFWN20Q09G
5	https://www.reuters.com/article/us-aviancaholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/journalists/julia-symmes-cobb
6	https://www.reuters.com/journalists/pei-li	https://www.reuters.com/article/us-soccer-spain-usa-idUSKCN1N652I
7	https://www.reuters.com/article/us-britain-lgbt-celebrities-idUSKCN1QW2KY	https://www.reuters.com/article/us-northkorea-dissidents-idUSKCN1QW2ZL
8	https://www.reuters.com/theWire	https://www.reuters.com/finance/currencies/quote?destAmt=&srcAmt=1&srcCurr=USD&destCurr=CNY
9	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI	https://www.reuters.com/news/archive/esgnews
10	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI	https://www.reuters.com/news/archive/esgnews

7. Channel

No	Default	PageRank
1	https://www.reuters.com/video/2019/03/06/breakingviews-tv-carry-trades?videoId=522657667&videoChannel=117766	https://www.reuters.com/article/us-ethiopia-airplane-airlines-idUSKCN1QW2P8
2	"https://www.reuters.com/video/2019/03/13/liverpool-result-at-bayern-wont-affect-p?videoId=525185681&videoChannel=79	https://www.reuters.com/article/us-barbie-anniversary-idUSKBN1QP2HG
3	https://www.reuters.com/video/2019/03/14/boeing-fix-could-take-weeks-us-lawmakers?videoId=526002472&videoChannel=5	https://www.reuters.com/article/aviva-ceo-idUSFWN20Q09G
4	https://www.reuters.com/video/2019/01/09/trump-threatens-california-wildfire-aid?videoId=501404962&videoChannel=118262	https://www.reuters.com/article/us-soccer-spain-usa-idUSKCN1N652I
5	https://www.reuters.com/video/2019/03/10/trump-asks-congress-for-750b-to-fund-mil?videoId=524322587&videoChannel=-10465	https://www.reuters.com/article/us-northkorea-dissidents-idUSKCN1QW2ZL
6	https://www.reuters.com/video/2019/03/10/brexit-in-peril-if-pm-mays-deal-is-rejec?videoId=524245642&videoChannel=118261	https://www.reuters.com/article/us-usa-trump-russia-deripaska-idUSKCN1QW2CA
7	https://www.reuters.com/video/2019/03/15/malawi-floods-wreak-havoc-cyclone-idai-e?videoId=525916362&videoChannel=73	https://www.reuters.com/article/us-soccer-germany-turkey-ozil-idUSKBN1KG1OX
8	https://www.reuters.com/video/2019/02/27/cohen-dishes-on-donald-jr-in-testimony?videoId=519943059&videoChannel=13976	https://www.reuters.com/article/ecuador-imf-idUSL1N20Y1O6
9	https://www.reuters.com/video/2019/03/13/woods-neck-not-painful-ahead-of-players?videoId=525458826&videoChannel=79	https://www.reuters.com/article/china-consumerday-idUSL8N2102SG"

10	https://www.reuters.com/video/2019/02/15/sustainable-is-the-new-sexy-at-new-york?videoId=515332676&videoChannel=118285	https://www.reuters.com/article/us-italy-china-mou-factbox-idUSKCN1QW1EB

8. Wall

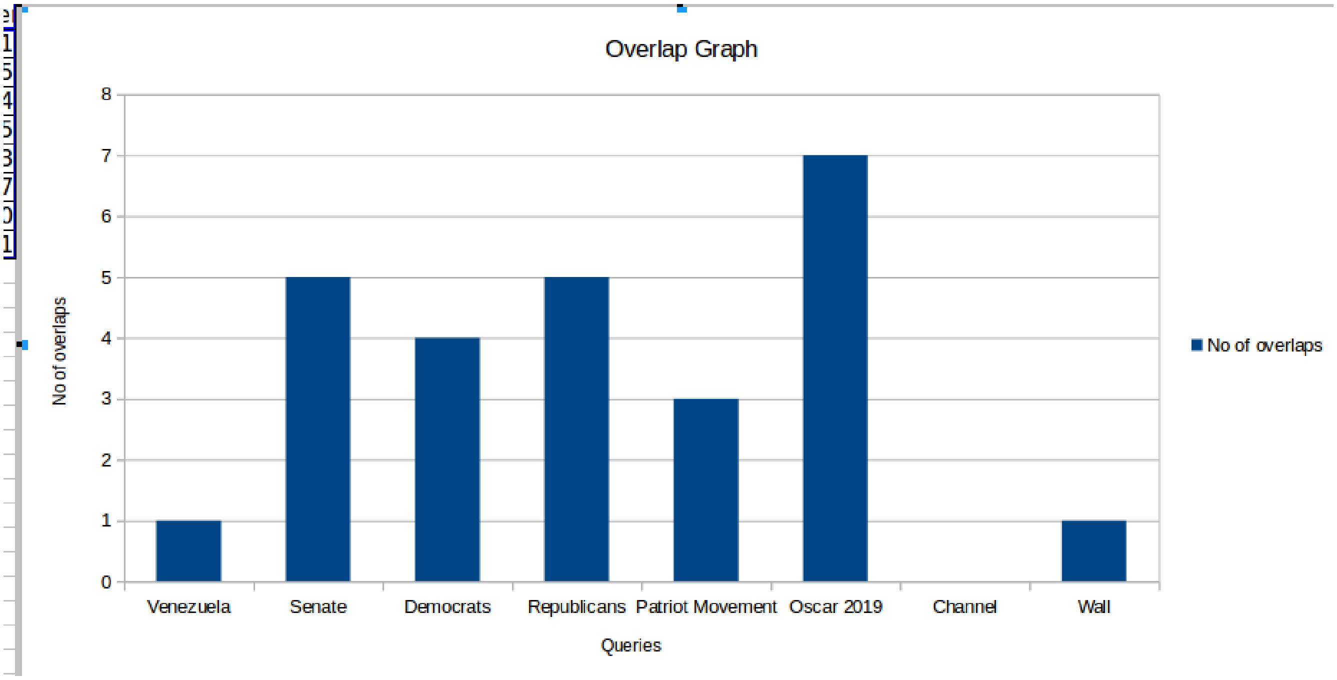
No	Default	PageRank
1.	http://www.reuters.com/investigates/section/wall-streets-way/"	https://www.reuters.com/news/archive/russia
2.	https://www.reuters.com/article/us-usa-trump-budget-wall-exclusive-idUSKBN1QR0CW	https://www.reuters.com/news/archive/esgnews
3.	https://www.reuters.com/article/us-usa-immigration-wall-insight-idUSKCN1QW1GR	https://www.reuters.com/finance/wealth
4.	https://www.reuters.com/politics"	https://www.reuters.com/finance/funds
5.	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/journalists/susan-mathew
6.	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK
7.	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK"	https://www.reuters.com/video/2019/03/06/breakingviews-tv-carry-trades?videoId=522657667&videoChannel=117766

8.	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK"	https://www.reuters.com/video/2019/02/01/images-of-january?videoId=510237001&videoChannel=118069
9.	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	"https://www.reuters.com/article/us-usa-trump-russia-timing-explainer-idUSKBN1QU1E8"
10.	https://www.reuters.com/journalists/noel-randewich	https://www.reuters.com/news/archive/japan"

Number of overlaps (URL and ID were same)

No	Query	No of overlaps
1	Venezuela	1
2	Senate	5
3	Democrats	4
4	Republicans	5
5	Patriot Movement	3
6	Oscar 2019	7
7	Channel	0
8	Wall	1

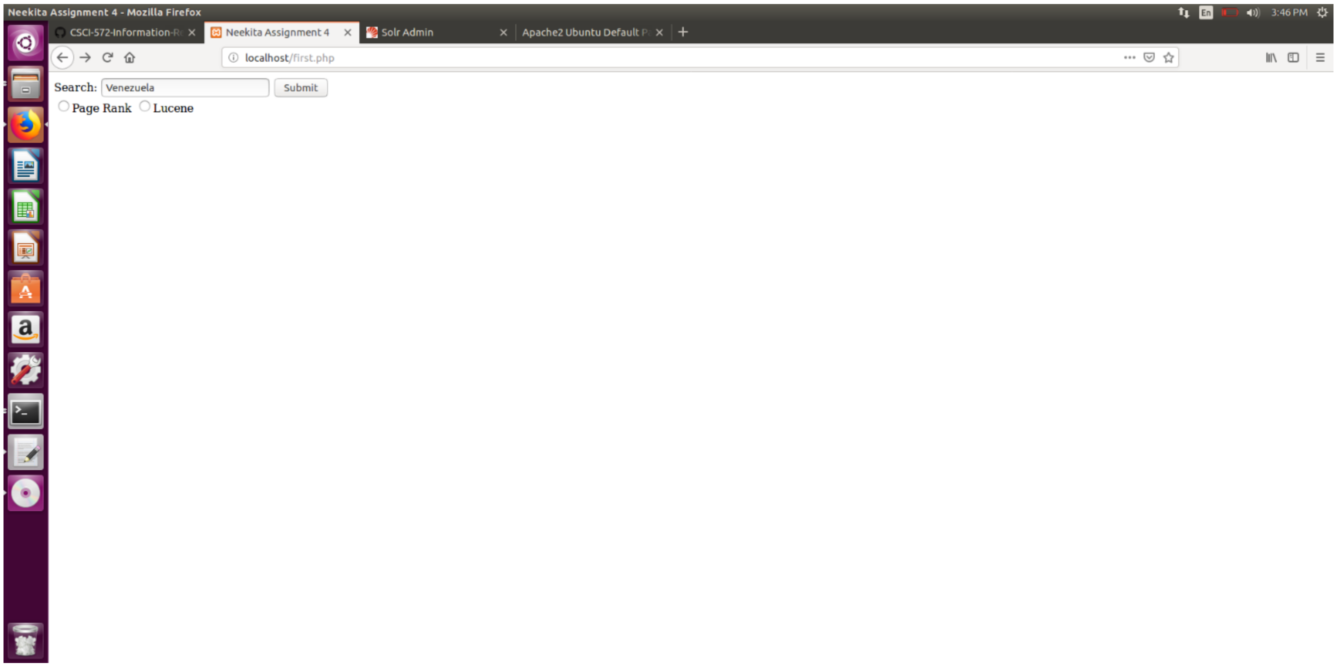
Graph of Overlaps



7. Installed Apache server and PHP on Ubuntu

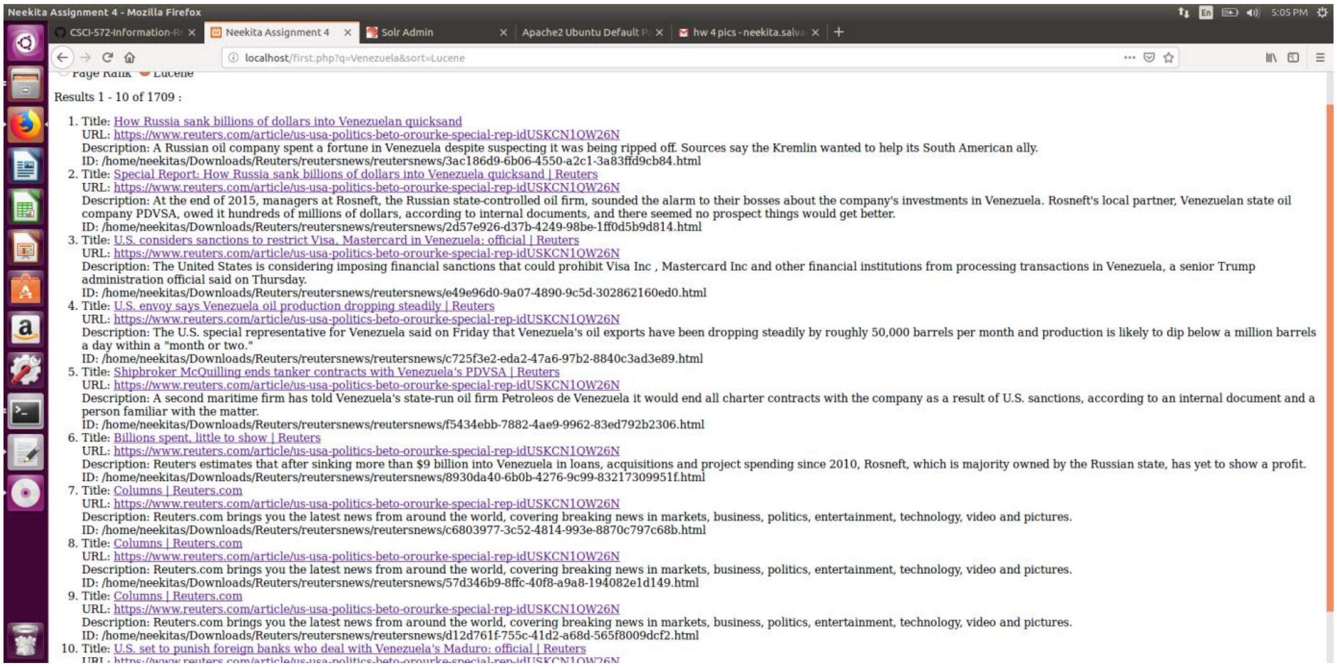
8. I created the UI using PHP and below are the screenshots. Also, I cloned the git ‘solr-php-repository’ and placed it in the same folder as UI.

Added proper paths to Apache Sever and csv files with links. Added radio buttons for user to choose between Lucene and PageRank. Displayed appropriate values and set limit to 10.



Search: Venezuela Submit

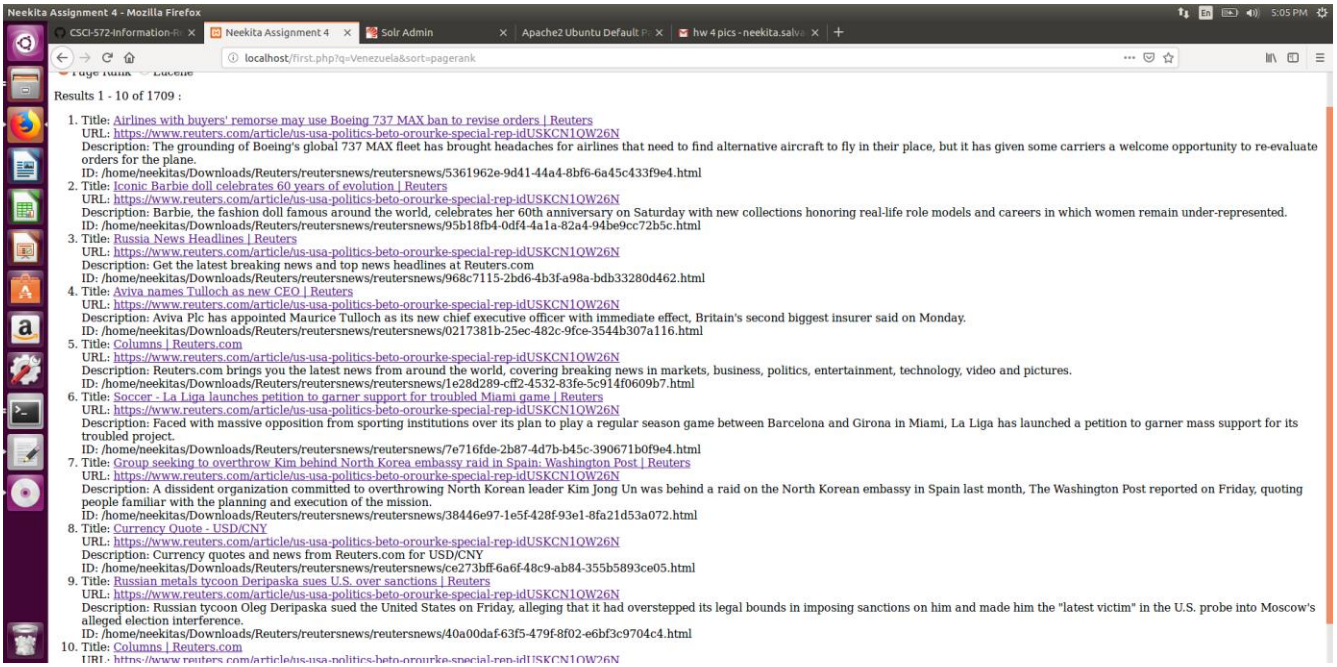
○ Page Rank ○ Lucene


Neekita Assignment 4 - Mozilla Firefox
CSCI-572-Information-R
Neekita Assignment 4
Solr Admin
Apache2 Ubuntu Default P
hw 4 pics - neekita.salva
5:05 PM
localhost/first.php?q=Venezuela&sort=Lucene

○ Page Rank ● Lucene

Results 1 - 10 of 1709 :

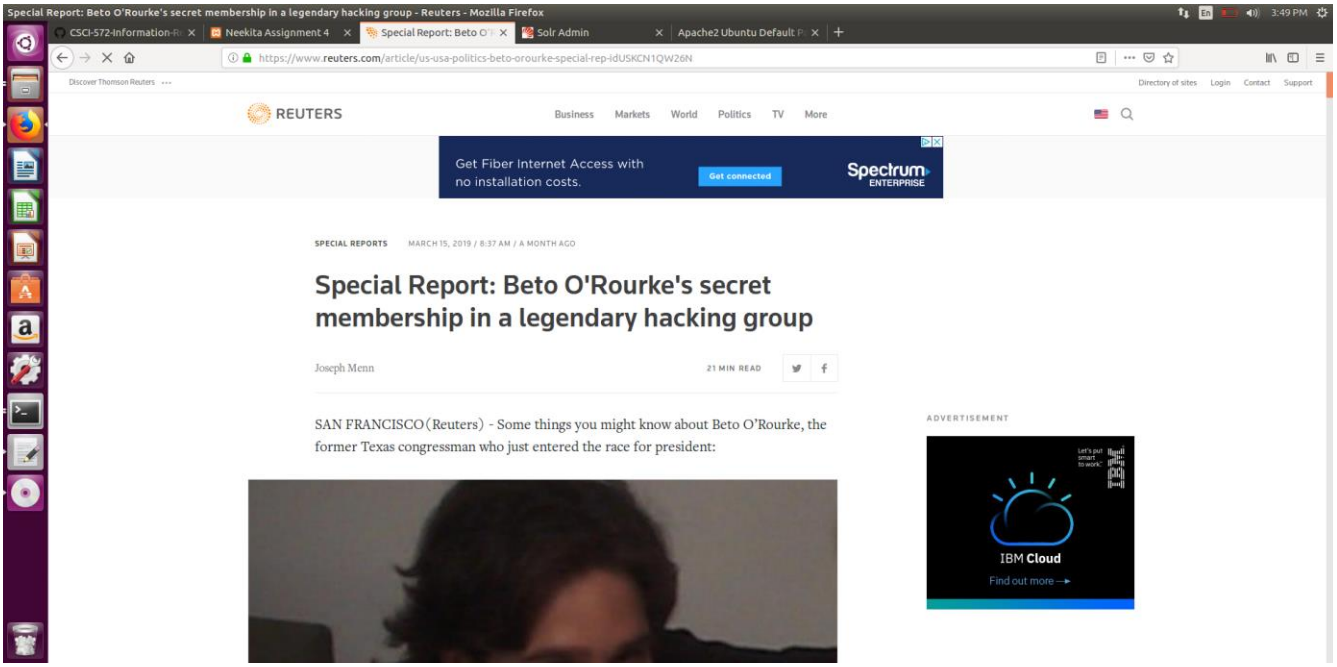
1. Title: How Russia sank billions of dollars into Venezuelan quicksand
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: A Russian oil company spent a fortune in Venezuela despite suspecting it was being ripped off. Sources say the Kremlin wanted to help its South American ally.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/3ac186d9-6b06-4550-a2c1-3a83ffd9cb84.html
2. Title: Special Report: How Russia sank billions of dollars into Venezuela quicksand | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: At the end of 2015, managers at Rosneft, the Russian state-controlled oil firm, sounded the alarm to their bosses about the company's investments in Venezuela. Rosneft's local partner, Venezuelan state oil company PDVSA, owed it hundreds of millions of dollars, according to internal documents, and there seemed no prospect things would get better.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/2d57e926-d37b-4249-98be-1ff0d5b9d814.html
3. Title: U.S. considers sanctions to restrict Visa, Mastercard in Venezuela: official | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: The United States is considering imposing financial sanctions that could prohibit Visa Inc , Mastercard Inc and other financial institutions from processing transactions in Venezuela, a senior Trump administration official said on Thursday.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/e49e96d0-9a07-4890-9c5d-302862160ed0.html
4. Title: U.S. envoy says Venezuela oil production dropping steadily | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: The U.S. special representative for Venezuela said on Friday that Venezuela's oil exports have been dropping steadily by roughly 50,000 barrels per month and production is likely to dip below a million barrels a day within a "month or two."
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/c725f3e2-eda2-47a6-97b2-8840c3ad3e89.html
5. Title: Shipbroker McQuilling ends tanker contracts with Venezuela's PDVSA | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: A second maritime firm has told Venezuela's state-run oil firm Petroleos de Venezuela it would end all charter contracts with the company as a result of U.S. sanctions, according to an internal document and a person familiar with the matter.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/f5434ebb-7882-4ae9-9962-83ed792b2306.html
6. Title: Billions spent, little to show | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Reuters estimates that after sinking more than $9 billion into Venezuela in loans, acquisitions and project spending since 2010, Rosneft, which is majority owned by the Russian state, has yet to show a profit.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/8930da40-6b0b-4276-9c99-83217309951f.html
7. Title: Columns | Reuters.com
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/c6803977-3c52-4814-993e-8870c797c68b.html
8. Title: Columns | Reuters.com
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/57d346b9-8ffc-40f8-a9a8-194082e1d149.html
9. Title: Columns | Reuters.com
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/d12d761f-755c-41d2-a68d-565f8009dcf2.html
10. Title: U.S. set to punish foreign banks who deal with Venezuela's Maduro: official | Reuters
    URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N


Neekita Assignment 4 - Mozilla Firefox
CSCI-572-Information-R
Neekita Assignment 4
Solr Admin
Apache2 Ubuntu Default P
hw 4 pics - neekita.salva
5:05 PM
localhost/first.php?q=Venezuela&sort=pagerank
Page Rank
Lucene

Results 1 - 10 of 1709 :

1. Title: Airlines with buyers' remorse may use Boeing 737 MAX ban to revise orders | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: The grounding of Boeing's global 737 MAX fleet has brought headaches for airlines that need to find alternative aircraft to fly in their place, but it has given some carriers a welcome opportunity to re-evaluate orders for the plane.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/5361962e-9d41-44a4-8bf6-6a45c433f9e4.html
2. Title: Iconic Barbie doll celebrates 60 years of evolution | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Barbie, the fashion doll famous around the world, celebrates her 60th anniversary on Saturday with new collections honoring real-life role models and careers in which women remain under-represented.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/95b18fb4-0df4-4a1a-82a4-94be9cc72b5c.html
3. Title: Russia News Headlines | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Get the latest breaking news and top news headlines at Reuters.com
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/968c7115-2bd6-4b3f-a98a-bdb33280d462.html
4. Title: Aviva names Tulloch as new CEO | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Aviva Plc has appointed Maurice Tulloch as its new chief executive officer with immediate effect, Britain's second biggest insurer said on Monday.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/0217381b-25ec-482c-9fce-3544b307a116.html
5. Title: Columns | Reuters.com
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Reuters.com brings you the latest news from around the world, covering breaking news in markets, business, politics, entertainment, technology, video and pictures.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/1e28d289-cff2-4532-83fe-5c914f0609b7.html
6. Title: Soccer - La Liga launches petition to garner support for troubled Miami game | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Faced with massive opposition from sporting institutions over its plan to play a regular season game between Barcelona and Girona in Miami, La Liga has launched a petition to garner mass support for its troubled project.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/7e716fde-2b87-4d7b-b45c-390671b0f9e4.html
7. Title: Group seeking to overthrow Kim behind North Korea embassy raid in Spain: Washington Post | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: A dissident organization committed to overthrowing North Korean leader Kim Jong Un was behind a raid on the North Korean embassy in Spain last month, The Washington Post reported on Friday, quoting people familiar with the planning and execution of the mission.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/38446e97-1e5f-428f-93e1-8fa21d53a072.html
8. Title: Currency Quote - USD/CNY
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Currency quotes and news from Reuters.com for USD/CNY
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/ce273bff-6a6f-48c9-ab84-355b5893ce05.html
9. Title: Russian metals tycoon Deripaska sues U.S. over sanctions | Reuters
   URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N
   Description: Russian tycoon Oleg Deripaska sued the United States on Friday, alleging that it had overstepped its legal bounds in imposing sanctions on him and made him the "latest victim" in the U.S. probe into Moscow's alleged election interference.
   ID: /home/neekitas/Downloads/Reuters/reutersnews/reutersnews/40a00daf-63f5-479f-8f02-e6bf3c9704c4.html
10. Title: Columns | Reuters.com
    URL: https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N

On clicking the URL:



SPECIAL REPORTS MARCH 15, 2019 / 8:37 AM / A MONTH AGO

Special Report: Beto O'Rourke's secret membership in a legendary hacking group

Joseph Menn

SAN FRANCISCO (Reuters) - Some things you might know about Beto O'Rourke, the former Texas congressman who just entered the race for president:

Question : Why PageRank of some pages is more than PageRank of others?

Answer: PageRank algorithm is based on how the web page is linked, that is , the number of in links and out links. If the page has many in links, it means the page is of relevance and is assigned a higher pagerank. Moreover, if these incoming links are also relevant, that is, they are pointed by many pages, then the PageRank value is higher. In our implementation, we created the EdgeList and assigned it as an input parameter to python code to generate a graph. We computed the rank using this graph.