

CSCI572 Assignment 4 Report:

Name: Prerana Teligi Harapanahalli Math

USCID: 7795750397

1. Steps:

1. Operating System used: **MacOS (High Sierra)**. There was no need to install Linux VM

2. **Solr installation:** Followed the instructions on http://lucene.apache.org/solr/guide/7_3/solr-tutorial.html and installed solr-7.3.0 by downloading solr-7.3.0.zip file from <http://www.apache.org/dyn/closer.lua/lucene/solr/7.3.0>

3. Unzipped the file on Desktop. This is the entire path to the directory
/Users/prerana/Desktop/solr-7.3.0

4. **Solr setup:** cd into solr-7.3.0 directory. Followed instruction mentioned on <http://www-scf.usc.edu/~csci572/2018Spring/hw4/IndexingwithTIKA V3.pdf>

- a. **bin/solr** – start Solr server
- b. **bin/solr create -c assignment4** - creates new core named assignment4 inside
/Users/prerana/Desktop/solr-7.3.0/server/solr/

5. Downloaded the NBC_News data from google drive link provided
<https://drive.google.com/drive/folders/1ZPWjQZMB7xeegjaeZfzqeJKq9oKuukio>

and unzip the HTML files and store those in the following path

/Users/prerana/Desktop/solr-7.3.0/NBC_News/

Store the provided CSV file named UrlToHtml_NBCNews.csv in this path

/Users/prerana/Desktop/solr-7.3.0/

6. Use this command to index the HTML files in NBC_News folder using Solr
bin/post -c assignment4 -filetypes html NBC_News/

7. You can view the indexing by going to this link <http://localhost:8983/solr/> and selecting your core i.e **assignment4** and to see how TIKa parser parsed the html pages and what fields were created by using TIKa. Select the query option, and just submit the default query (*:*). Click on overview option to view your core statistics. Before that to configure default querying using Solr, uncomment the str element with name “df”, and replace the value with **_text_**, where we specify that the default query field has to be “_text_”. Now when you click on Execute Query button on Solr Query page it returns the index results with “**numFound**”:15725 (i.e. number of documents).

8. **assignment4.php:** This file is created to send the form query to Solr through Apache server and retrieve the results. We also need to download solr-php-client in order to run solr service from php client (i.e. from our php file) by doing cloning this git repo using **git clone** <https://github.com/PTCInc/solr-php-client.git> command. We also need to download XAMPP for

running Apache server or you can also use the default server on Mac. I used XAMPP. After starting Apache on XAMPP, place the assignment4.php file and solar-php-client folder in htdocs folder of XAMPP i.e. /Applications/XAMPP/xamppfiles/htdocs/ .

9. PageRank: Wrote Java code to create edgelist.txt by providing NBC_News folder containing HTML files and UrlToHtml_NBCNews.csv file as input. Using JSoup, urlFileMap and fileUrlMap functions. Imported **JSoup jar** into eclipse and ran the java code.

10. Wrote python code to generate **external_pageRankFile.txt** file from **edgelist.txt**.

11. Type **sudo pip install networkx** command in command line to install networkx python library. This is used to run the python file. I have named my python file as **networkX.py**. and used these parameters: alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None. Run this to generate external_pageRankFile.txt.

12. After generating **external_pageRankFile.txt** file place it in /Users/prerana/Desktop/solr-7.3.0/server/solr/assignment4/data/ folder.

13. Modify **managed-schema.xml** by adding the following lines:

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

14. Define listeners for query in **solrconfig.xml** by modifying these lines **<listener event="newSearcher" class="org.apache.solr.schema.ExternalFileFieldReloader" />**
<listener event="firstSearcher"
class="org.apache.solr.schema.ExternalFileFieldReloader" />

15. Add the following inside **solrconfig.xml** for ANDing the query terms for better retrieval results. **<requestHandler name="/select" class="solr.SearchHandler"> <lst> tag <str name="q.op">AND</str>**

16. Now reload the index, by going to the **Solr Dashboard UI ->Core Admin** and clicking on the **“Reload”** button.

17. Go to this URL <http://localhost/assignment4.php> to view you search application. Now you can give any query term and retrieve results from Solr.

18. Run the queries given in assignment description.

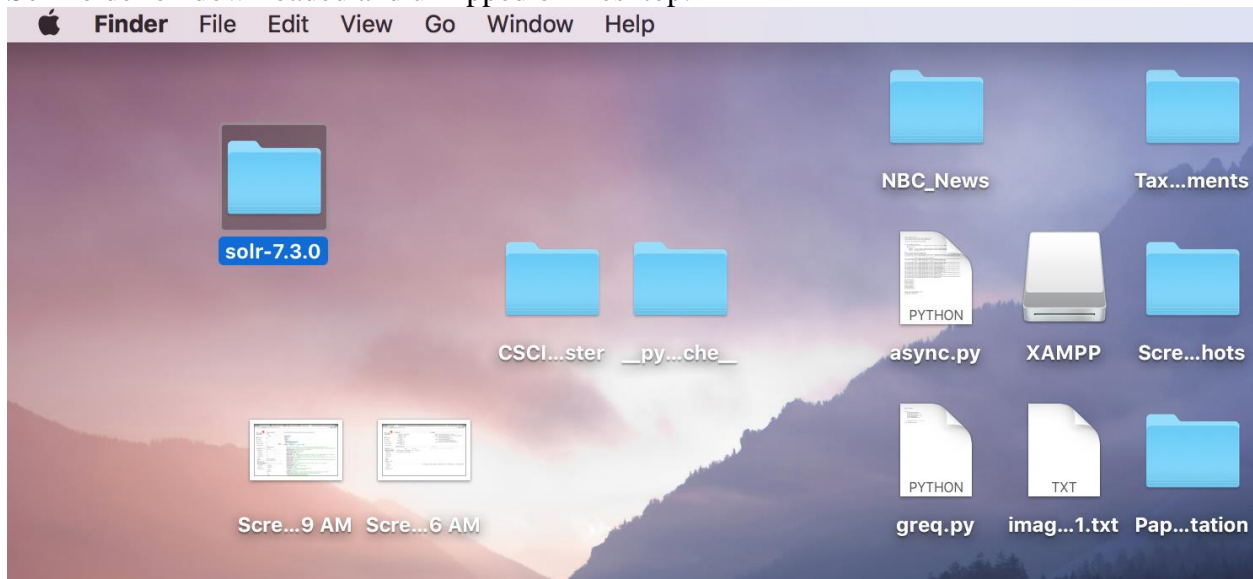
2. why some pages have higher PageRank values than others?

- PageRank algorithm assigns a numeric weight to each link element of DOM in a webpage for measuring its relative importance. This is achieved by measuring the number of outgoing links for a particular website. We create a graph with the list of edges pointing to the page (i.e. edgeList.txt). We use this to find importance of any given page. PageRank determines the weight

of the page and then outputs it in given order (Highest first(descending)). Some pages have more weight than others and hence they have higher PageRank values than others.

3. Screenshots:

Solr folder on downloaded and unzipped on Desktop:

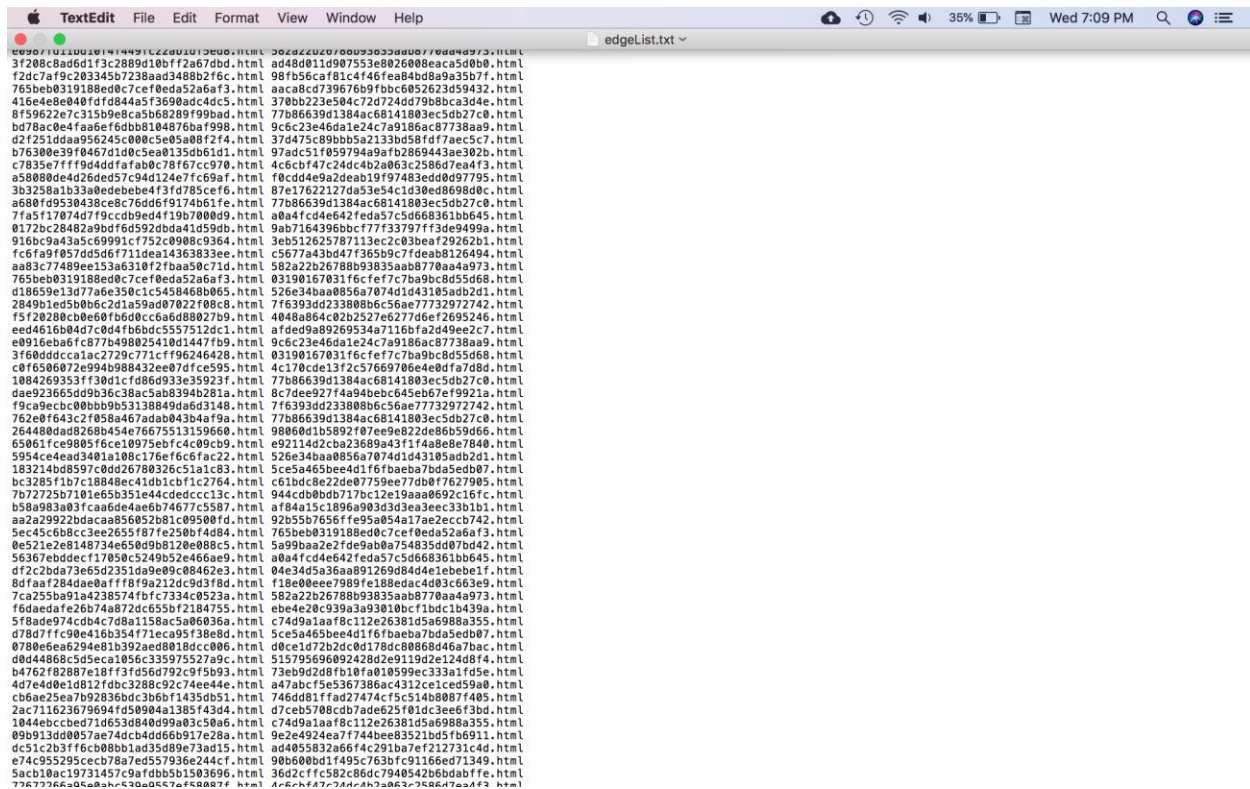


Solr indexing:

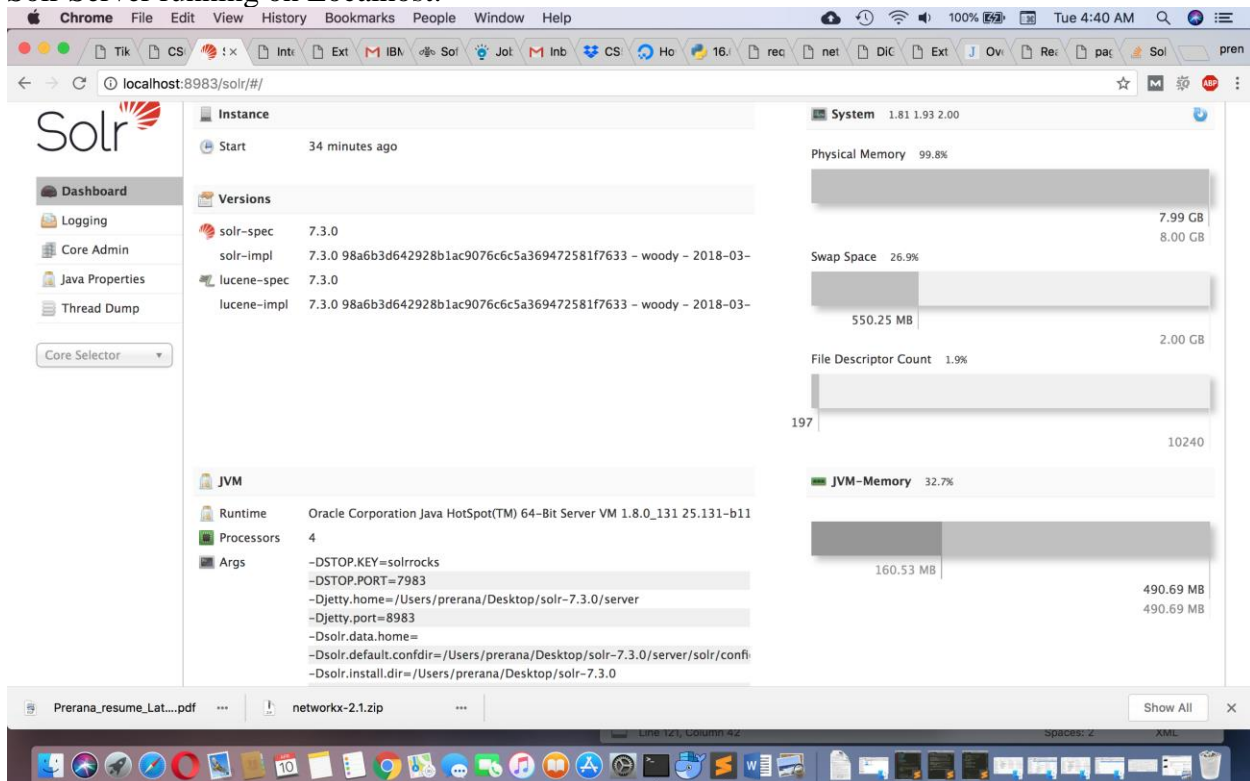
```
Created new core 'assignment4'
PreranasMacBook:solr-7.3.0 prerana$ bin/post -c assignment4 -filetypes html NBC_News/
java -classpath /Users/prerana/Desktop/solr-7.3.0/dist/solr-core-7.3.0.jar -Dautoyes -Dfiletypes=html -Dc=assignment4 -Ddatafiles -Drecursive=yes org.apache.solr.util.SimplePostTool NBC_News/
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/assignment4/update...
Entering auto mode. File endings considered are html
Entering recursive mode, max depth=999, delay=0s
Indexing directory NBC_News (15727 files, depth=0)
POSTing file e8f7e425f5db5ea8616caeca3d6d1c1.html (text/html) to [base]/extract
POSTing file 6497255c6ad7a31c9a048a5b5f6574d7.html (text/html) to [base]/extract
POSTing file dc59d99a54fcb0def24cf1817335fa88.html (text/html) to [base]/extract
POSTing file 4363a4985dd632c05431401109ecd98a.html (text/html) to [base]/extract
POSTing file 3261976e46ce2e0835261ad93211d2e9.html (text/html) to [base]/extract
POSTing file 9d548ee69e5a4f3dde86423abaa80a96.html (text/html) to [base]/extract
POSTing file e4d95edcd90e3394490f9b0e9cd0eb7.html (text/html) to [base]/extract
POSTing file 32715ee253494486c93c1b59af255cc4.html (text/html) to [base]/extract
POSTing file 235978a2194b79b99e9eac1c3579ae.html (text/html) to [base]/extract
POSTing file 20a49b7cba124aab28380e2b7d2d6421.html (text/html) to [base]/extract
POSTing file a8ccffbcac68f1acbf4f39129a09c5b6f.html (text/html) to [base]/extract
POSTing file 6878f9cb498f92bd91e2c631d1b50bd4.html (text/html) to [base]/extract
POSTing file cf82d2bd23f5e201b2870196a311e3c8.html (text/html) to [base]/extract
POSTing file cb1368316dcce0f0a2c8fa0b0c8dfcb7.html (text/html) to [base]/extract
POSTing file 40a480554cc2674619307985f8b0b7340.html (text/html) to [base]/extract
POSTing file ab3c1c184a5d43ee652d2ad56ac2e.html (text/html) to [base]/extract
POSTing file 5cc882829f58c8869cdc14ca0f5145d.html (text/html) to [base]/extract
POSTing file b29c42b6ab14254eccbbaa14706ee89.html (text/html) to [base]/extract
POSTing file d517e35a140e246df79d0bed8a20df15.html (text/html) to [base]/extract
POSTing file e3ecd43077b600efc9ca1c0ef31983e.html (text/html) to [base]/extract
POSTing file d1e2f8ac38d6b4610607231d239f229.html (text/html) to [base]/extract
POSTing file 44abd1b3d32e5b5b6e48e88059fdb159.html (text/html) to [base]/extract
POSTing file 386c0c48ba490a8ef9f18c0baef98af.html (text/html) to [base]/extract
POSTing file 6addadab87de37a4eeb1976896f408af.html (text/html) to [base]/extract
POSTing file 2b8e88da39cb1bc29834baba2cd3a81c.html (text/html) to [base]/extract
POSTing file fd82b2eadb79fbc7d47d5d73f966df7.html (text/html) to [base]/extract
POSTing file 25a67c6e1ac15d94737015a93598e500.html (text/html) to [base]/extract
POSTing file aa768f28371c46489548099336205cd.html (text/html) to [base]/extract
POSTing file 48b09d6fec34e10b083893fec08fac2e9.html (text/html) to [base]/extract
POSTing file 15240813f6cd3a087d52af8d1b058e.html (text/html) to [base]/extract
POSTing file cbb4bf3316a4312f5d5b89d8225312a.html (text/html) to [base]/extract
POSTing file b31737a560e3656791c9b85c7d606464.html (text/html) to [base]/extract
POSTing file e941028c923aaf96616c2ad19e1e0168.html (text/html) to [base]/extract
POSTing file df56dcee8760dc00a9172ef9ed9c3bd.html (text/html) to [base]/extract
POSTing file d5ba4887764a590a274ad799545d38b.html (text/html) to [base]/extract
POSTing file 29099f938362e66c03d36a1145e5b9a57.html (text/html) to [base]/extract
POSTing file 47aa86ca8dbf1e3fedaddb725a231c1.html (text/html) to [base]/extract
POSTing file e6583adbcb22ab0b08555c435831725a.html (text/html) to [base]/extract
POSTing file 1997a2020f14e9aa20b9a05b63cbbdc.html (text/html) to [base]/extract
POSTing file c88622482356ba03961122d0e943681f.html (text/html) to [base]/extract
POSTing file 2b8f2b4cfb46348991f571edc65acc85.html (text/html) to [base]/extract
POSTing file a0ac27b9953c48e00caf3283d1ade46c.html (text/html) to [base]/extract
POSTing file a97e637167f5cc54468a1727d907ec8.html (text/html) to [base]/extract
POSTing file 1f8ec99a0f15f938ff31d44f57299cd.html (text/html) to [base]/extract
POSTing file 419d1d9c0c6b219b0c857640d0baa618.html (text/html) to [base]/extract
POSTing file 5c3a39c00406ef43c354a19d4f9fb385.html (text/html) to [base]/extract
POSTing file ffc18b2c8bdd0e8f8ee5faeeb8e6193b.html (text/html) to [base]/extract

POSTing file 11ebe86a0b2706029040048293abf988.html (text/html) to [base]/extract
POSTing file d6d1be5202087e0569a47471cad21d2.html (text/html) to [base]/extract
POSTing file 4af54f0c807319c8f4d71fde20a80.html (text/html) to [base]/extract
POSTing file 32969f4607528ca3cd69d6126dc5a78a.html (text/html) to [base]/extract
POSTing file 81e77f155d65c8cf9fee46a52b02609.html (text/html) to [base]/extract
POSTing file 901de4945f9d00aa1ed077aafc636f254.html (text/html) to [base]/extract
POSTing file ed7c1e048286989ef0b0c1010744d004.html (text/html) to [base]/extract
POSTing file a256f3e198e348a639787406e4221086.html (text/html) to [base]/extract
POSTing file be0fa8d51505b834cd63996ba181175.html (text/html) to [base]/extract
POSTing file c3d09ebc762c0e4a6836af958e28a75.html (text/html) to [base]/extract
POSTing file 0608a9fcc5847d329a08bd2a824bacd64.html (text/html) to [base]/extract
POSTing file 408a72192ebfd7e671012c4b380497a.html (text/html) to [base]/extract
POSTing file d3df37bbc1e83c83853db0cf0fe45a33.html (text/html) to [base]/extract
POSTing file 774104390a95ce632a351cf79d38bbc.html (text/html) to [base]/extract
POSTing file c3d3f2b32fc1bb2799083ef4f8f5d100d.html (text/html) to [base]/extract
POSTing file 14c26fc833702511341a59300b7f08a.html (text/html) to [base]/extract
POSTing file 47b0a5b09b01a6bf0efde98225.html (text/html) to [base]/extract
POSTing file 526d31cf7c338d8e2d098be122496e.html (text/html) to [base]/extract
POSTing file 7220bf3a4858fbc9d5eb549d56e366e8.html (text/html) to [base]/extract
POSTing file cb95a3212b1ec96a3075329b588a0cc6.html (text/html) to [base]/extract
POSTing file 00a5d6a45d082094243f64abe0c0d52c.html (text/html) to [base]/extract
POSTing file 4a8eb0e8258ae835a4ae4b16a8ae7c3.html (text/html) to [base]/extract
POSTing file fe2e5242100d0e024479ca03ab92cf.html (text/html) to [base]/extract
POSTing file 4ff47a0ee383f06ecd57ad5da74463d.html (text/html) to [base]/extract
POSTing file c453bd53e29aa021aa095d3e00712619.html (text/html) to [base]/extract
POSTing file 0342fd5ffc2e6e0fe09e2639fb8b4d00.html (text/html) to [base]/extract
POSTing file 2ee9fcdcc75e6430d51d5cc195d23ff9.html (text/html) to [base]/extract
POSTing file a0f27f59df0eb4d41842e782a5b2679a.html (text/html) to [base]/extract
POSTing file 4fa843580a93a7adc269448d533bdf.html (text/html) to [base]/extract
POSTing file 4094790167c9238269aeb0840efed58.html (text/html) to [base]/extract
POSTing file e6177c037ea485765ea2400cd8827d00.html (text/html) to [base]/extract
POSTing file 347d991672d3157212e6efc7fe2f6c6c.html (text/html) to [base]/extract
POSTing file 80bc06ef58301baff04336a739770f64.html (text/html) to [base]/extract
POSTing file 7af2fd7b8830413f2a727839b30fe644.html (text/html) to [base]/extract
POSTing file f4883629734168caa3fd5444fe23e6ed.html (text/html) to [base]/extract
POSTing file 96bcd5a0161d2f7a8f02d524da5fdeed.html (text/html) to [base]/extract
POSTing file 08a7457c5e0f562ba98aac2e22107e.html (text/html) to [base]/extract
POSTing file cccb556384f76e788ec34b0e5a97a0e7.html (text/html) to [base]/extract
POSTing file 1fb992150a276a593a33bf7762d4b52d.html (text/html) to [base]/extract
POSTing file c42295aab2187aef774cda97733a7cd9.html (text/html) to [base]/extract
POSTing file 6ff5c2eb96e77a24c895f3ff08e7c174.html (text/html) to [base]/extract
POSTing file a98e7d79065a1101f7edad046a11bc78.html (text/html) to [base]/extract
POSTing file 230fd04e71230854fb4dc6435b990d1.html (text/html) to [base]/extract
POSTing file fd712227574aff31cdd0931b977a16a.html (text/html) to [base]/extract
POSTing file 22c237037841ab539072f05f6f4e6f5.html (text/html) to [base]/extract
POSTing file 57ee11a7b4473570edd37b98468ddc2.html (text/html) to [base]/extract
POSTing file 22ca2e61e053c1d424442692c7f3f808.html (text/html) to [base]/extract
POSTing file 6040a3369e996031146458df2175363.html (text/html) to [base]/extract
POSTing file 077bc29244b60e8829fe43dd20c12cf.html (text/html) to [base]/extract
POSTing file 185f2135c1d9f7c73bc240a42d07a.html (text/html) to [base]/extract
POSTing file 9ac230a6e4a4d050c13d7f62000137.html (text/html) to [base]/extract
POSTing file 6f57b0fdded37a5e579665f583c0e724f.html (text/html) to [base]/extract
15727 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/assignment4/update...
Time spent: 0:00:02.585
PreranasMacBook:solr-7.3.0 prerana$ █
```

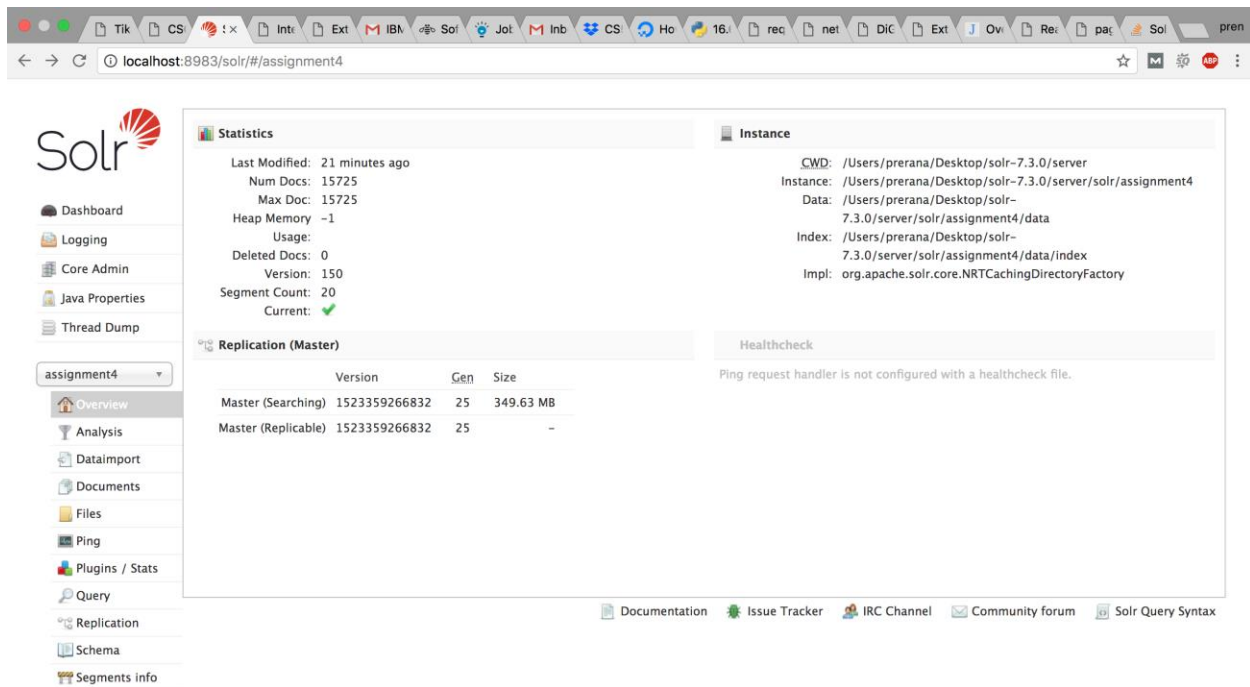
edgeList.txt:



Solr Server running on Localhost:



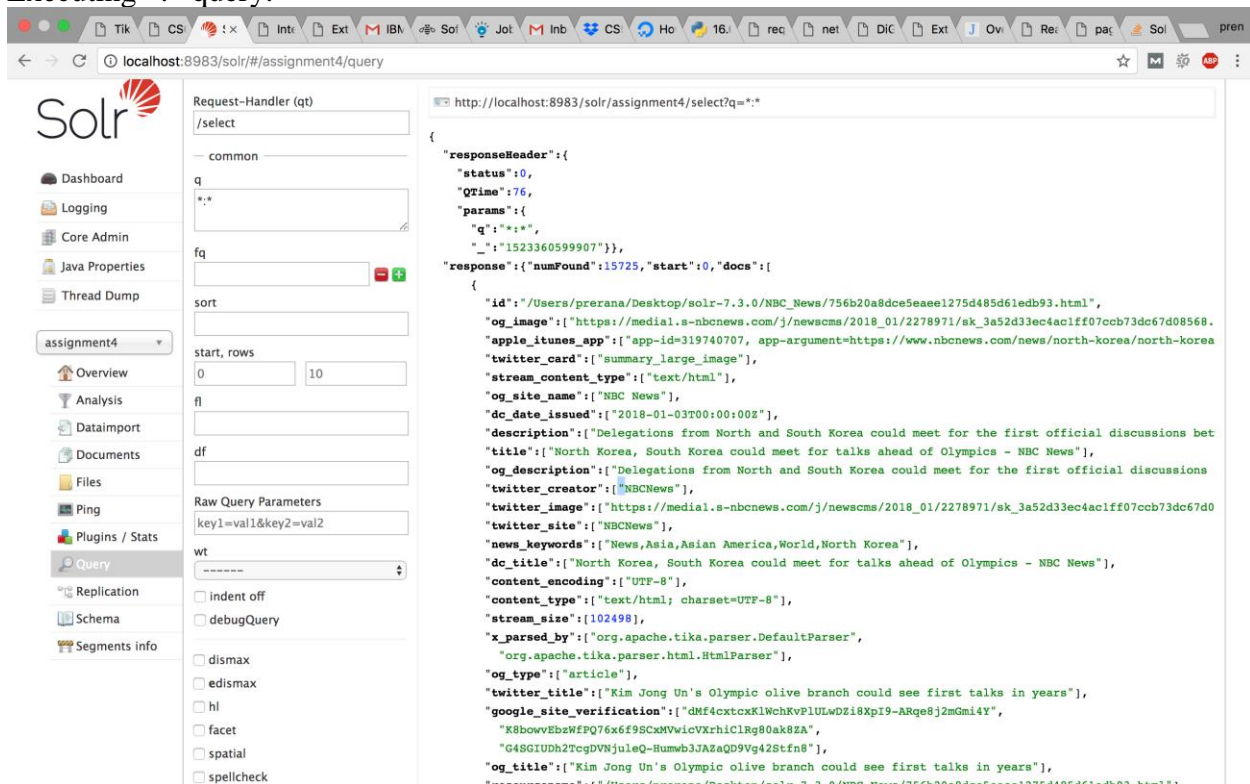
Page displayed after selecting Core. i.e. assignment4:



The image shows the Solr Admin UI for the 'assignment4' instance. The left sidebar contains navigation links: Dashboard, Logging, Core Admin, Java Properties, Thread Dump, assignment4 (selected), Overview, Analysis, Dataimport, Documents, Files, Ping, Plugins / Stats, Query, Replication, Schema, and Segments info. The main content area is divided into three sections: Statistics, Instance, and Replication (Master). The Statistics section shows metrics like Last Modified (21 minutes ago), Num Docs (15725), Max Doc (15725), Heap Memory (-1), Deleted Docs (0), Version (150), Segment Count (20), and Current status (green checkmark). The Instance section shows CWD, Instance path, Data path, Index path, and Impl. The Replication (Master) section shows a table with columns Version, Gen, and Size for Master (Searching) and Master (Replicable). A Healthcheck section at the bottom right indicates that the Ping request handler is not configured with a healthcheck file. At the bottom of the page, there are links for Documentation, Issue Tracker, IRC Channel, Community forum, and Solr Query Syntax.

	Version	Gen	Size
Master (Searching)	1523359266832	25	349.63 MB
Master (Replicable)	1523359266832	25	-

Executing *.* query:



The image shows the Solr Admin UI for the 'assignment4' instance, displaying the results of a query. The left sidebar is the same as the previous image. The main content area is divided into two sections: Request-Handler (qt) and Query. The Request-Handler (qt) section shows the selected handler as /select. The Query section shows the query parameters: q=*, fq=, sort=, start=0, rows=10, fl=, df=, raw Query Parameters (key1=val1&key2=val2), wt=, and various checkboxes for sorting and faceting. The Query Results section shows the response from the Solr server, including the status, QTime, params, and the response body. The response body is a JSON object containing the search results for the query *.*.

```

{
  "responseHeader": {
    "status": 0,
    "QTime": 76,
    "params": {
      "q": "*",
      "_:": "1523360599907"
    }
  },
  "response": {
    "numFound": 15725,
    "start": 0,
    "docs": [
      {
        "id": "/Users/prerana/Desktop/solr-7.3.0/NBC_News/756b20a8dce5aeae1275d485d61edb93.html",
        "og_image": ["https://medial.s-nbcnews.com/j/newscoms/2018_01/2278971/sk_3a52d33ec4ac1ff07ccb73dc67d08568"],
        "apple_itunes_app": {"app-id": 319740707, "app-argument": "https://www.nbcnews.com/news/north-korea/north-korea"},
        "twitter_card": {"summary_large_image": "https://www.nbcnews.com/news/north-korea/north-korea"},
        "stream_content_type": ["text/html"],
        "og_site_name": ["NBC News"],
        "dc_date_issued": ["2018-01-03T00:00:00Z"],
        "description": ["Delegations from North and South Korea could meet for the first official discussions bet"],
        "title": ["North Korea, South Korea could meet for talks ahead of Olympics - NBC News"],
        "og_description": ["Delegations from North and South Korea could meet for the first official discussions"],
        "twitter_creator": ["NBCNews"],
        "twitter_image": ["https://medial.s-nbcnews.com/j/newscoms/2018_01/2278971/sk_3a52d33ec4ac1ff07ccb73dc67d0"],
        "twitter_site": ["NBCNews"],
        "news_keywords": ["News,Asia,Asian America,World,North Korea"],
        "dc_title": ["North Korea, South Korea could meet for talks ahead of Olympics - NBC News"],
        "content_encoding": ["UTF-8"],
        "content_type": ["text/html; charset=UTF-8"],
        "stream_size": [102498],
        "x_parsed_by": ["org.apache.tika.parser.DefaultParser", "org.apache.tika.parser.html.HtmlParser"],
        "og_type": ["article"],
        "twitter_title": ["Kim Jong Un's Olympic olive branch could see first talks in years"],
        "google_site_verification": ["dMf4cxtcxKlWchKvP1ULwDZ18XpI9-ARqe8j2mGmi4Y", "K8bowvEbwWfPQ76x6f9SCxMvVlcVXrh1C1Rg80ak82A", "G4SGIUdh2TcgDvNjuleQ-Humwb3JAZaQD9Vg42Stfn8"],
        "og_title": ["Kim Jong Un's Olympic olive branch could see first talks in years"],
        "resourceName": "/Users/prerana/Desktop/solr-7.3.0/NBC_News/756b20a8dce5aeae1275d485d61edb93.html*1.

```

Executing query after sorting it in descending order using PageRank Algorithm:

Solr

Request-Handler (qt)
/select

common

q
.

fq

sort
pageRankFile desc

start, rows
0 10

fl

df

Raw Query Parameters
key1=val1&key2=val2

wt
json

assignment4

Overview
Analysis
Dataimport
Documents
Files
Ping
Plugins / Stats
Query
Replication
Schema
Segments info

http://localhost:8983/solr/assignment4/select?q=*&sort=pageRankFile desc

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 2,
    "params": {
      "q": " *.* ",
      "sort": "pageRankFile desc",
      "_: "1523439431133"}},
  "response": {
    "numFound": 15725, "start": 0, "docs": [
      {
        "id": "/Users/prerana/Desktop/solr-7.3.0/NBC_News/77b86639d1384ac68141803ec5db27c0.html",
        "og_image": ["https://nodeassets.nbcnews.com/cdnassets/projects/socialshareimages/og-nbcnews1200x630.png"],
        "apple_itunes_app": ["app-id=319740707, app-argument="],
        "twitter_card": ["summary_large_image"],
        "stream_content_type": ["text/html"],
        "og_site_name": ["NBC News"],
        "description": ["Go to NBCNews.com for breaking news, videos, and the latest top stories in world news, b"],
        "title": ["NBC News - Breaking News & Top Stories - Latest World, US & Local News"],
        "og_description": ["Go to NBCNews.com for breaking news, videos, and the latest top stories in world news"],
        "twitter_creator": ["NBCNews"],
        "twitter_image": ["https://nodeassets.nbcnews.com/cdnassets/projects/socialshareimages/og-nbcnews1024x512"],
        "twitter_site": ["NBCNews"],
        "news_keywords": ["Cover"],
        "dc_title": ["NBC News - Breaking News & Top Stories - Latest World, US & Local News"],
        "content_encoding": ["UTF-8"],
        "content_type": ["text/html; charset=UTF-8"],
        "stream_size": [514862],
        "x_parsed_by": ["org.apache.tika.parser.DefaultParser",
          "org.apache.tika.parser.html.HtmlParser"],
        "og_type": ["article"],
        "twitter_title": ["NBC News - Breaking News & Top Stories - Latest World, US & Local News"],
        "google_site_verification": ["dMf4cxtcxKlWchKvPlULwDzi8XpI9-ARqe8j2mGmi4Y",
          "K8bowvEbzWfPQ76x6f98CxmVvicVXrhICLRg80ak82A",
          "G4SGIUDh22cgDVNjuleQ-Humwb3JAZaQD9Vg42Stfn8"],

```

Running assignment4.php file on localhost:

CSCI 572 : Assignment 4

Search :

Enter Search Query

Ranking Algorithm

☒ Lucene ☐ PageRank

Running Lucene (Default) search algorithm and retrieving results from Solr:

localhost/assignment4.php?q=Bitcoin+Blockchain&pageRankType=lucene

CSCI 572 : Assignment 4

Search : Bitcoin Blockchain

Enter Search Query

Ranking Algorithm

☒ Lucene ☐ PageRank

Submit

TITLE:	How Bitcoin and other cryptocurrencies could kill off cash
FILE ID:	27640bb3f3bb0345f023330f021248e0.html
URL:	https://www.nbcnews.com/mach/amp/ncna857036
DESCRIPTION:	National governments are also looking into creating virtual currencies of their own.

TITLE:	Bitcoin pays for entire wedding as its price soars past \$16,000
FILE ID:	c346a4bb29558ea93d980d10cc15b5db.html
URL:	https://www.nbcnews.com/news/amp/ncna826776
DESCRIPTION:	With bitcoin hitting a record high of over \$16,000, one wedding in Indiana just got bigger.

TITLE:	What is Bitcoin? And should you invest in it?
FILE ID:	9ba27c9494397c038fe869bcb1e76a27.html
URL:	https://www.nbcnews.com/better/amp/ncna844106
DESCRIPTION:	Only 24 percent of Americans are familiar with Bitcoin. Here's what you need to know about this "cryptocurrency."

TITLE:	Middle America Is Crazy in Love With Bitcoin
FILE ID:	b4f26bcd9d8731708d5e1972ccf26fd7.html
URL:	https://www.nbcnews.com/news/amp/ncna789011
DESCRIPTION:	In May, there were more searches for Bitcoin than Beyonce. Cryptocoins are going mainstream as the middleclass wants in on the "Wild West" action.

Running PageRank search algorithm and retrieving results from Solr:

localhost/assignment4.php?q=Bitcoin+Blockchain&pageRankType=pageRankAlgo

CSCI 572 : Assignment 4

Search : Bitcoin Blockchain

Enter Search Query

Ranking Algorithm

☐ Lucene ☒ PageRank

Submit

TITLE:	Bitcoin heist: 600 powerful computers stolen in Iceland - NBC News
FILE ID:	8ae709c12a5f14d6ed1d255cc213db10.html
URL:	https://www.nbcnews.com/news/world/bitcoin-heist-600-powerful-computers-stolen-iceland-n852811
DESCRIPTION:	Some 11 people were arrested, including a security guard, in what Icelandic media have dubbed the "Big Bitcoin Heist."

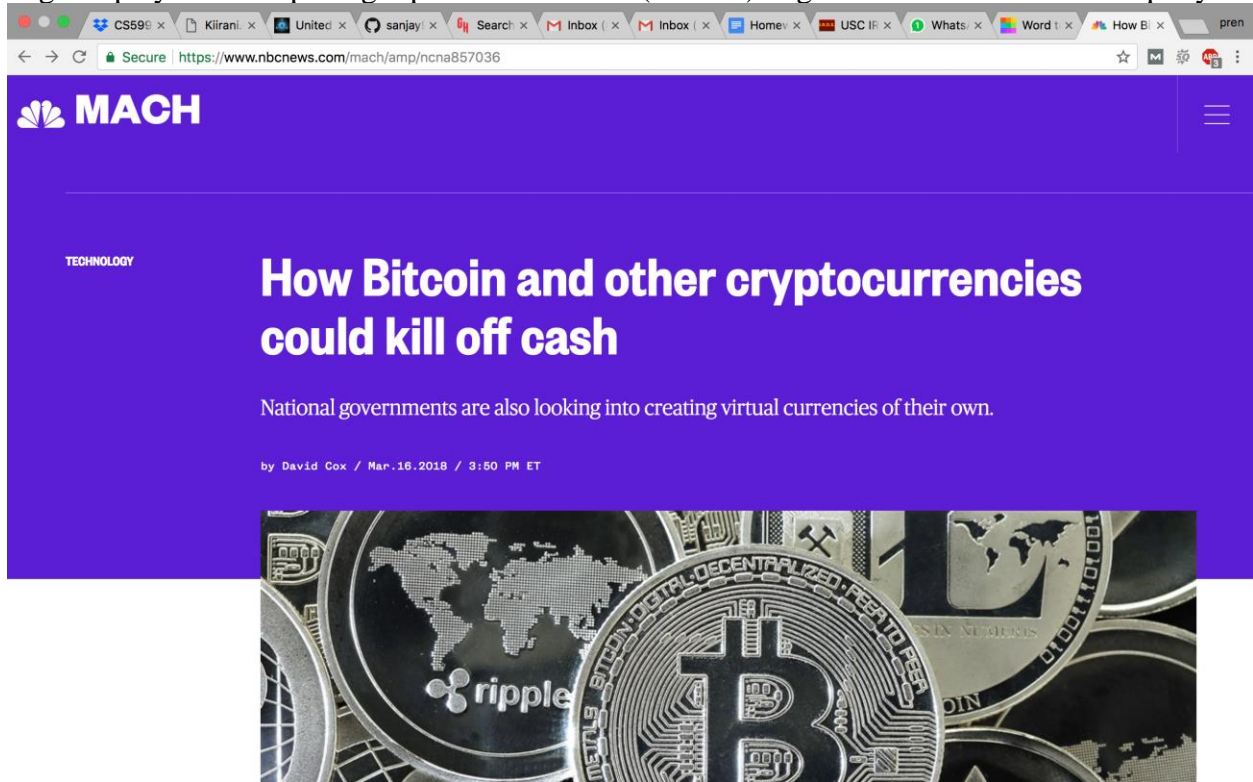
TITLE:	How Bitcoin and other cryptocurrencies could kill off cash
FILE ID:	27640bb3f3bb0345f023330f021248e0.html
URL:	https://www.nbcnews.com/mach/amp/ncna857036
DESCRIPTION:	National governments are also looking into creating virtual currencies of their own.

TITLE:	The hidden dangers of buying virtual currency go beyond a simple hack
FILE ID:	68fac26a85fda5d7984e109781f38e81.html
URL:	https://www.nbcnews.com/news/amp/ncna852706
DESCRIPTION:	Cyberthieves can break into crypto exchanges, drain crypto wallets, and infect computers with malware.

TITLE:	What is Bitcoin? And should you invest in it?
FILE ID:	eea75aac08fe9029b9cdfd6f1721d149.html
URL:	https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106
DESCRIPTION:	Only 24 percent of Americans are familiar with Bitcoin. Here's what you need to know about this "cryptocurrency."

TITLE:	Bitcoin, other cryptocurrencies plunge as waves of bad news
--------	---

Page displayed after opening top link for Lucene (Default) Algo for Bitcoin Blockchain query:



4. Results:

Bitcoin Blockchain	
Lucene	PageRank
https://www.nbcnews.com/mach/amp/ncna857036	https://www.nbcnews.com/news/world/bitcoin-heist-600-powerful-computers-stolen-iceland-n852811
https://www.nbcnews.com/news/amp/ncna826776	https://www.nbcnews.com/mach/amp/ncna857036
https://www.nbcnews.com/better/amp/ncna844106	https://www.nbcnews.com/news/amp/ncna852706
https://www.nbcnews.com/news/amp/ncna789011	https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106

https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-ncna857036	https://www.nbcnews.com/news/amp/ncna856821
https://www.nbcnews.com/news/amp/ncna852706	https://www.nbcnews.com/tech/tech-news/bitcoin-other-cryptocurrencies-plunge-wave-bad-news-n856821
https://www.nbcnews.com/tech/innovation/bitcoin-pays-entire-wedding-its-price-soars-past-16-000-n826776	https://www.nbcnews.com/news/amp/ncna826776
https://www.nbcnews.com/business/economy/middle-america-crazy-love-bitcoin-n789011	https://www.nbcnews.com/news/amp/ncna813351
https://www.nbcnews.com/better/business/what-bitcoin-should-you-invest-it-ncna844106	https://www.nbcnews.com/better/amp/ncna844106
https://www.nbcnews.com/business/business-news/hidden-dangers-buying-virtual-currency-go-beyond-simple-hack-n852706	https://www.nbcnews.com/mach/science/how-bitcoin-other-cryptocurrencies-could-kill-cash-ncna857036

Stephen Hawking	
Lucene	PageRank
https://www.nbcnews.com/mach/science/brief-history-stephen-hawking-cambridge-cultural-icon-ncsl856456	https://www.nbcnews.com/mach
https://www.nbcnews.com/news/world/british-physicist-stephen-hawking-among-world-s-greatest-minds-science-n702301	https://www.nbcnews.com/think
https://www.nbcnews.com/think/opinion/stephen-hawking-s-life-was-made-possible-access-great-health-ncna856971	https://www.nbcnews.com/mach/video/this-futuristic-jet-pack-that-lets-users-fly-up-to-10-000-feet-in-the-sky-1189436995946
https://www.nbcnews.com/science/space/im-atheist-stephen-hawking-god-space-travel-n210076	https://www.nbcnews.com/mach/video/from-space-dream-to-space-junk-the-vanguard-1-satellite-completes-60-years-in-space-today-1188094531644

https://www.nbcnews.com/science/space/im-atheist-stephen-hawking-god-space-travel-n210076	https://www.nbcnews.com/mach/video/why-youtube-s-algorithm-boosted-conspiracy-theory-videos-in-the-wake-of-the-parkland-shooting-1187660355671
https://www.nbcnews.com/news/world/amp/british-physicist-stephen-hawking-among-world-s-greatest-minds-science-n702301	https://www.nbcnews.com/mach/video/pi-in-the-sky-how-a-nasa-scientist-uses-pi-to-look-for-life-beyond-earth-1185847363999
https://www.nbcnews.com/mach/amp/ncna818926	https://www.nbcnews.com/mach/science/alien-satellites-might-offer-new-way-find-e-t-ncna857841
https://www.nbcnews.com/science/science-news/six-years-after-zero-g-flight-stephen-hawking-still-space-f6C9703084	https://www.nbcnews.com/mach/space
https://www.nbcnews.com/mach/science/what-stephen-hawking-taught-us-about-black-holes-ncna857471	https://www.nbcnews.com/mach/science
https://www.nbcnews.com/mach/science/stephen-hawking-says-humans-must-leave-earth-within-600-years-ncna818926?cid=sm_npd_nn_fb_mc	https://www.nbcnews.com/mach/video/a-wrinkle-in-time-is-probably-right-about-multiple-dimensions-1182239299525

Falcon Heavy	
Lucene	PageRank
https://www.nbcnews.com/mach/amp/ncna845256	https://www.nbcnews.com/mach/video/this-futuristic-jet-pack-that-lets-users-fly-up-to-10-000-feet-in-the-sky-1189436995946
https://www.nbcnews.com/mach/amp/ncna845096	https://www.nbcnews.com/mach/video/from-space-dream-to-space-junk-the-vanguard-1-satellite-completes-60-years-in-space-today-1188094531644
https://www.nbcnews.com/mach/science/watch-spacex-s-falcon-heavy-	https://www.nbcnews.com/mach/video/why-youtube-s-algorithm-boosted-conspiracy-theory-

rocket-make-its-maiden-flight-ncna845096	videos-in-the-wake-of-the-parkland-shooting-1187660355671
https://www.nbcnews.com/mach/science/spacex-s-falcon-heavy-launch-gives-big-boost-elon-musk-ncna845256	https://www.nbcnews.com/mach/video/pi-in-the-sky-how-a-nasa-scientist-uses-pi-to-look-for-life-beyond-earth-1185847363999
https://www.nbcnews.com/mach/science/spacex-s-falcon-feat-may-be-just-prelude-main-event-ncna846446	https://www.nbcnews.com/mach/video/-a-wrinkle-in-time-is-probably-right-about-multiple-dimensions-1182239299525
https://www.nbcnews.com/mach/science/here-s-what-will-happen-tesla-spacex-shot-space-ncna845566	https://www.nbcnews.com/mach/video/astronaut-scott-kelly-s-genes-show-long-term-changes-after-a-year-in-space-1186826819794
https://www.nbcnews.com/mach/science/month-space-pictures-cosmic-penguin-playing-mars-ncsl852081	https://www.nbcnews.com/mach/video/this-futuristic-jet-pack-that-lets-users-fly-up-to-10-000-feet-in-the-sky-1189436995946?playlist=mmlsnnnd_mach
https://www.nbcnews.com/mach/amp/ncna846446	https://www.nbcnews.com/mach/video/driverless-cars-can-now-be-tested-in-california-without-a-safety-driver-1172721219646
https://www.nbcnews.com/mach/science/how-spacex-aims-bring-high-speed-internet-whole-world-ncna853766	https://www.nbcnews.com/news/world/unapproves-watered-down-new-sanctions-against-north-korea-n800511
https://www.nbcnews.com/mach/science/talking-robotic-head-will-join-astronauts-iss-ncna854821	https://www.nbcnews.com/news/world/prince-mohammed-bin-salman-aims-rebrand-saudi-arabia-n817201

Syrian War	
Lucene	PageRank
https://www.nbcnews.com/storyline/syrias-children/syrian-civil-war-photo-photographer-s-rescue-goes-viral-n748146	https://www.nbcnews.com/news/mideast/syria-s-eastern-ghouta-president-assad-visits-troops-front-line-n857801

https://www.nbcnews.com/news/mideast/syria-s-civil-war-has-been-raging-7-years-what-n849851	https://www.nbcnews.com/politics/white-house/saudi-crown-prince-if-iran-makes-nuclear-bomb-so-will-n857921
https://www.nbcnews.com/news/mideast/syrias-civil-war-has-been-raging-7-years-whats-behind-n849851	https://www.nbcnews.com/storyline/aleppos-children/aleppo-s-bana-al-abed-mother-take-questions-facebook-n700336
https://www.nbcnews.com/news/other/barrel-bombs-just-another-symbol-brutality-syrian-civil-war-f2D11765640	https://www.nbcnews.com/news/north-korea/north-korea-uses-pirate-tactics-earn-millions-sanction-busting-n856921
https://www.nbcnews.com/slideshow/children-among-injured-besieged-damascus-suburb-n834526	https://www.nbcnews.com/storyline/isis-terror/isis-will-remain-threat-2018-experts-warn-n828146
https://www.nbcnews.com/slideshow/air-raids-cause-devastation-eastern-ghouta-region-syria-n834526	https://www.nbcnews.com/news/world/analysis-trump-s-america-first-foreign-policy-isolates-u-s-n833046
https://www.nbcnews.com/news/mideast/amp/syria-s-civil-war-has-been-raging-7-years-what-n849851	https://www.nbcnews.com/news/world/north-korea-s-kim-trump-meet-andrea-mitchell-weighs-risks-n855111
https://www.nbcnews.com/storyline/syrias-children/amp/syrian-civil-war-photo-photographer-s-rescue-goes-viral-n748146	https://www.nbcnews.com/news/world/putin-s-bravado-over-russian-nuclear-weapons-emboldened-trump-analysts-n852301
https://www.nbcnews.com/storyline/aleppos-children/aleppo-siege-bombs-hit-hospital-rebel-held-syrian-city-nurse-n657841	http://www.nbcnews.com/id/51444762/ns/meet_the_press-transcripts/t/april-lindsey-graham-bill-richardson-michele-flournoy-mike-murphy-maggie-haberman-andrea-mitchell/
https://www.nbcnews.com/storyline/aleppos-children/aleppo-siege-syrian-kids-find-childhood-deep-underground-n622036	http://www.nbcnews.com/id/52898657/ns/meet_the_press-transcripts/t/august-john-lewis-cory-booker-bobby-jindal-raul-labrador-sheryl-wudunn-al-sharpton-david-brooks-doris-kearns-goodwin/

LA Rams	
Lucene	PageRank
https://www.nbcnews.com/news/sports/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366	https://www.nbcnews.com/news/sports/sports-broadcasting-legend-dick-enberg-dies-82-n832076
https://www.nbcnews.com/news/sports/amp/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366	https://www.nbcnews.com/news/sports/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366
https://www.nbcnews.com/news/sports/amp/sports-broadcasting-legend-dick-enberg-dies-82-n832076	https://www.nbcnews.com/news/sports/amp/sports-broadcasting-legend-dick-enberg-dies-82-n832076
https://www.nbcnews.com/news/sports/sports-broadcasting-legend-dick-enberg-dies-82-n832076	https://www.nbcnews.com/news/sports/amp/how-history-informs-colin-kaepernick-s-national-anthem-protest-n639366

Black Panther	
Lucene	PageRank
https://www.nbcnews.com/news/nbcblk/black-panther-s-wakanda-sheds-light-black-excellence-n844801	https://www.nbcnews.com/think
https://www.nbcnews.com/news/nbcblk/black-panther-sets-fandango-record-top-pre-selling-marvel-movie-n836631	https://www.nbcnews.com/better/pop-culture/girl-scout-cookies-ranked-better-worst-ncna737381
https://www.nbcnews.com/news/nbcblk/black-panther-novelist-journalism-sci-fi-bring-hidden-light-n848011	https://www.nbcnews.com/pop-culture/movies/black-panther-tops-box-office-fifth-straight-weekend-equaling-avatar-n857781
https://www.nbcnews.com/tech/internet/plot-destroy-black-panther-ratings-thwarted-facebook-n844226	https://www.nbcnews.com/think/hot-take
https://www.nbcnews.com/pop-culture/movies/black-panther-blows-away-box-office-192-million-weekend-n849246	https://www.nbcnews.com/think/thought-experiment

https://www.nbcnews.com/news/nbcblk/black-panther-black-comic-creators-spent-years-breaking-down-barriers-n848461	https://www.nbcnews.com/pop-culture/celebrity/celebrity-deaths-2017-looking-back-famous-figures-we-lost-n831216
https://www.nbcnews.com/news/nbcblk/black-panther-actress-janeshia-adams-ginyard-talks-living-stuntwoman-life-n847646	https://www.nbcnews.com/better/pop-culture/pros-cons-america-s-favorite-diet-plans-ncna854396
https://www.nbcnews.com/think/opinion/defeat-killmonger-black-panther-holds-lessons-israel-ncna850831	https://www.nbcnews.com/pop-culture/awards/golden-globes-2018-metoo-casts-shadow-over-celebration-movies-tv-n835451
https://www.nbcnews.com/news/amp/ncna848011	https://www.nbcnews.com/better/pop-culture/anti-diet-movement-leading-us-astray-ncna854856
https://www.nbcnews.com/news/amp/ncna847706	https://www.nbcnews.com/storyline/confederate-flag-furor/house-would-ban-confederate-flags-va-cemetery-flagpoles-n576906

Austin Bombings	
Lucene	PageRank
https://www.nbcnews.com/news/us-news/police-identify-teen-killed-austin-package-bombing-investigation-grows-n856371	https://www.nbcnews.com/better
https://www.nbcnews.com/news/us-news/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786	https://www.nbcnews.com/mach
https://www.nbcnews.com/news/us-news/amp/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786	https://www.nbcnews.com/leftfield
https://www.nbcnews.com/widget/video-embed/1184733251954	https://www.nbcnews.com/news/us-news/nbc-affiliates-n19981

https://www.nbcnews.com/news/us-news/amp/police-identify-teen-killed-austin-package-bombing-investigation-grows-n856371	https://www.nbcnews.com/think
https://www.nbcnews.com/news/us-news/austin-texas-edge-after-spate-package-bombs-kill-two-n855956	https://www.nbcnews.com/news/us-news/deadly-austin-bombings-were-meant-send-message-police-chief-says-n857786
https://www.nbcnews.com/news/us-news/amp/austin-texas-edge-after-spate-package-bombs-kill-two-n855956	https://www.nbcnews.com/news/us-news/veterans-hospital-killings-spotlight-problem-violence-against-health-care-workers-n855931
http://www.nbcnews.com/search/	https://www.nbcnews.com/news/us-news/cirque-du-soleil-star-yann-arnaud-plunges-death-during-florida-n857761
https://www.nbcnews.com/news/latino/celia-cruz-be-remembered-grammy-lifetime-achievement-award-n516956	https://www.nbcnews.com/news/latino/maduro-challenger-henri-falcon-unsettles-venezuela-s-presidential-vote-n857836
https://www.nbcnews.com/news/latino/california-affirmative-action-bill-sparks-fierce-debate-n87426	https://www.nbcnews.com/better/wellness

Paris Climate Deal	
Lucene	PageRank
https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-declines-endorsing-paris-climate-change-deal-g7-summit-will-n765406	https://www.nbcnews.com/tech/social-media/tech-companies-struggle-human-side-fake-news-n857426
https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/amp/trump-declines-endorsing-paris-climate-change-deal-g7-summit-will-n765406	https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-message-during-overseas-trip-now-he-s-home-n765136

https://www.nbcnews.com/politics/white-house/trump-pulls-u-s-out-paris-climate-agreement-n767066	https://www.nbcnews.com/news/world/analysis-trump-s-america-first-foreign-policy-isolates-u-s-n833046
https://www.nbcnews.com/politics/white-house/here-s-what-happens-if-trump-leaves-paris-climate-agreement-n766761	https://www.nbcnews.com/news/us-news/trump-un-north-korean-leader-suicide-mission-n802596
https://www.nbcnews.com/politics/politics-news/forget-trump-u-s-storms-ahead-climate-change-never-n801826	https://www.nbcnews.com/feature/meet-the-press-24-7/meet-press-august-6-2017-n790001
https://www.nbcnews.com/meet-the-press/meet-press-june-4-2017-n768046	https://www.nbcnews.com/news/us-news/trump-un-north-korean-leader-suicide-mission-n802596?cid=eml_nbn_20170919
https://www.nbcnews.com/news/world/analysis-trump-s-america-first-foreign-policy-isolates-u-s-n833046	https://www.nbcnews.com/news/china/china-backs-boasts-new-era-patriotic-movies-rap-music-n843686
https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/trump-end-first-foreign-trip-pledges-unwavering-support-u-s-n765446	http://www.nbcnews.com/id/53771922/ns/meet_the_press-transcripts/t/dec-charlayne-hunter-gault-tom-brokaw-jesse-jackson-richard-stengel-charles-ogletree-al-sharpton-katty-kay-paul-gigot-harry-smith/
https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/g-7-summit-leaders-brace-clash-trump-trade-climate-n764971	https://www.nbcnews.com/mach/science/michio-kaku-sees-amazing-things-our-future-except-those-scary-ncna851226
https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/amp/trump-end-first-foreign-trip-pledges-unwavering-support-u-s-n765446	https://www.nbcnews.com/storyline/trump-s-first-foreign-trip/donald-trump-nato-why-his-silence-article-5-big-deal-n765011

5. Overlap Table:

Query	Number of Overlaps
Bitcoin Blockchain	5
Stephen hawking	0
Falcon Heavy	0
Syrian War	0
LA Rams	4
Black Panther	0
Austin Bombings	1
Paris Climate Deal	1

Overlap Chart:

