CSCI 572 Assignment 4

Comparing Search Engine Ranking Algorithms

Steps for the assignment:

Step 1: Preparation:

- Downloaded and Installed Virtual Box and Ubuntu on my Windows machine
- Downloaded and Installed Apache Solr and start Solr by “bin/solr start”
- Downloaded and extracted all the LA times files from the given google drive.

Step 2: Indexing the given HTML files in Solr:

- I created a core myexample using the command bin/solr create –c myexample.
- I changed and saved the managed-schema file in the conf folder as follow:

  <field name="id" type="string" indexed="true" stored="true" required="true" multiValued="false" />

  <field name="_version_" type="long" indexed="true" stored="true"/>

  <field name="_root_" type="string" indexed="true" stored="false"/>

  <field name="_text_" type="text_general" indexed="true" stored="false" multiValued="true"/>

  <copyField source="*" dest="_text_"/>
- then indexing HTML files in Solr, run command “bin/post -c myexample -filetypes html latimes”
- I opened the browser and went to http://localhost:8983/solr/. Select the core “myexample” from the dropdown to see if the files have gotten indexed.

Step 3: Build the PHP run environment:

- I download and installed XAMPP for running PHP in Ubuntu and tested a sample PHP code.
- Then, I cloned the GitHub repository of solr-php-client client APIS and configure environment for the index.php as follow:
  - Create a 572HW4 folder under XAMPP/htdocs
  - Copy solr-php-client to this 572HW4 folder

- Create a new php file named **index.php**
- Modify the php code provided in assignment description so that we have two algorithm options and a submit button.

• Then I test if it works. Go to http://localhost/572, choosing **Lucene(default)** as algorithm to send a query and then return expected results

Step 4: Add PageRank algorithm for indexing:

• I wrote my Java code with the import of JSoup library. By running the Java code, I got all the outgoing links in HTML pages, and formed an edgelist file named result.tex.

• Then I wrote python code to compute PageRank scores:

- I Installed python and pip to my computer
- Then I installed Networx and nose for later use
- Wrote the Python code and run it to get pagerank values from graph. Saved the returned result as a file named **external_pageRankFile**.

• I use the following parameter to configure:

alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None.

• I placed pagerank file in the data folder of my core myexample.

• The next step is to add the field in the managed-schema which refers to this score.

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

• Then, I made modifications required in the solrconfig.xml file so that when the index is reloaded, it can access the rank file.

• I reloaded the core on the Solr admin page to check if choosing PageRank Algorithm as option works.

Why some pages have higher PageRank values than others?

Based on query results , it is easy to find that the pages with more incoming links receives a higher PageRank value than others. That is when a page is pointed more by other pages, it will have a higher PageRank value.

Tables containing list of URLs of the top ten results produced by two ranking algorithm.

1.Donald Trump

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/nation/politics/trailguide/la-na-trailguide-updates-trump-release-new-doctor-s-letter-1473951652-htmlstory.html	http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html
2	http://www.latimes.com/nation/politics/trailguide/la-na-trailguide-updates-donald-trump-open-to-allowing-those-in-1473296491-htmlstory.html	http://www.latimes.com/
3	http://www.latimes.com/politics/la-pol-ca-donald-trump-california-races-downticket-20160527-snap-htmlstory.html	http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html
4	http://www.latimes.com/opinion/op-ed/la-oe-johnston-trump-cons-and-cheats-20181004-story.html	http://www.latimes.com/politics/la-na-pol-kavanaugh-ford-hearing-htmlstory.html
5	http://www.latimes.com/sports/olympics/la-sp-olympics-live-updates-the-moment-when-the-donald-trump-and-kim-1518187562-htmlstory.html	http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html
6	http://www.latimes.com/opinion/la-ol-enter-the-fray-whom-to-believe-michael-cohen-or-donald-1532702692-htmlstory.html	http://www.latimes.com/
7	http://www.latimes.com/espanol/politica/la-es-dos-funerales-y-una-boda-el-rechazo-a-donald-trump-20180828-story.html	http://www.latimes.com/entertainment/la-et-entertainment-news-updates-2018-htmlstory.html
8	http://www.latimes.com/nation/politics/trailguide/la-na-live-updates-trailguide-1475872277-htmlstory.html	http://www.latimes.com/
9	http://www.latimes.com/opinion/la-ol-enter-the-fray-donald-trump-might-actually-have-to-1535040294-htmlstory.html	http://www.latimes.com/
10	http://www.latimes.com/entertainment/tv/la-et-st-donald-trump-60-minutes-20161110-story.html	http://www.latimes.com/politics/essential/la-pol-ca-essential-politics-may-2018-htmlstory.html

2. LA Lakers

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/	http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-southern-california-2017-htmlstory.html

2	http://www.latimes.com/	http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html
3	http://www.latimes.com/sports/lakers/la-sp-lakers-plaschke-20180924-story.html	http://www.latimes.com/sports/dodgers/la-sp-live-dodgers-brewers-nlcs-htmlstory.html
4	http://www.latimes.com/sports/lakers/la-sp-lakers-plaschke-20180924-story.html	http://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html
5	http://www.latimes.com/	http://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html
6	http://www.latimes.com/	http://www.latimes.com/
7	http://www.latimes.com/sports/lakers/la-sp-lakers-media-day-20180924-story.html	http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html
8	http://www.latimes.com/	http://www.latimes.com/politics/la-na-pol-kavanaugh-ford-hearing-htmlstory.html
9	http://www.latimes.com/sports/lakers/la-sp-lakers-sidebar-20180924-story.html	http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html
10	http://www.latimes.com/sports/lakers/la-sp-lakers-5-questions-20181013-story.html	http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-get-the-eric-sondheimer-s-latest-1490657918-htmlstory.html

3.Star Wars

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/entertainment/movies/la-ca-mn-star-wars-timeline-20180516-story.html	http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html
2	http://www.latimes.com/entertainment/movies/la-et-mn-star-wars-last-jedi-characters-20171214-htmlstory.html	http://www.latimes.com/entertainment/la-et-entertainment-news-updates-2018-htmlstory.html
3	http://www.latimes.com/travel/themeparks/la-trb-star-wars-land-disneyland-20150817-story.html	http://www.latimes.com/
4	http://www.latimes.com/entertainment/movies/la-et-mn-ron-howard-star-wars-han-solo-20170622-story.html	http://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html
5	http://www.latimes.com/fashion/la-ig-wwd-rag-and-bones-star-wars-20171106-story.html	http://www.latimes.com/about/la-privacypolicy-20180703-story.html
6	http://www.latimes.com/fashion/la-ig-wwd-rag-and-bones-star-wars-20171106-story.html	http://www.latimes.com/
7	http://www.latimes.com/entertainment/herocomplex/la-et-hc-jon-favreau-star-wars-mandalorian-20181003-htmlstory.html	http://www.latimes.com/
8	http://www.latimes.com/entertainment/herocomplex/la-et-hc-comic-con-star-wars-fandom-20180723-htmlstory.html	http://www.latimes.com/

9	http://www.latimes.com/brandpublishing/travelplus/lasvegasguide/features/la-ss-vegas-the-hot-list-burlesque-star-wars-speakeasies-20180620dto-story.html	http://www.latimes.com/travel/la-tr-california-bucket-list-updates-2017-htmlstory.html
10	http://www.latimes.com/entertainment/music/la-et-ms-star-wars-secrets-empire-20180116-story.html	http://www.latimes.com/

4. Lebron James

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-diet-20181004-story.html	http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html
2	http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-20181013-story.html	http://www.latimes.com/
3	http://www.latimes.com/sports/lakers/la-sp-lakers-nuggets-20181002-story.html	http://www.latimes.com/
4	http://www.latimes.com/sports/lakers/la-sp-lakers-nuggets-20181002-story.html	http://www.latimes.com/
5	http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-rondo-defense-20180930-story.html	http://www.latimes.com/
6	http://www.latimes.com/sports/nba/la-sp-lebron-james-20160620-snap-story.html	http://www.latimes.com/
7	http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-20181013-story.html	http://www.latimes.com/entertainment/la-et-tiff-2018-toronto-film-festival-updates-htmlstory.html
8	http://www.latimes.com/sports/lakers/la-sp-lakers-lebron-james-20181013-story.html	http://www.latimes.com/
9	http://www.latimes.com/	http://www.latimes.com/
10	http://www.latimes.com/	http://www.latimes.com/

5. 2018 World Cup

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/sports/la-sp-42nd-ryder-cup-pictures-2018-photogallery.html	http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-southern-california-2017-htmlstory.html
2	http://www.latimes.com/sports/la-lb-816-45174-la-me-ln-mexico-korea-soccer-viewing-20180622-htmlstory.html	http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html
3	http://www.latimes.com/sports/la-lb-816-45174-la-me-ln-mexico-korea-soccer-viewing-20180622-htmlstory.html	http://www.latimes.com/sports/dodgers/la-sp-live-dodgers-brewers-nlcs-htmlstory.html

4	http://www.latimes.com/sports/soccer/worldcup/la-sp-mexico-korea-1998-world-cup-2018-world-cup-20180622-story.html	http://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html
5	http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-b-lgica-detiene-en-seco-a-brasil-1530907910-htmlstory.html	http://www.latimes.com/sports/la-sp-live-dodgers-braves-nlds-htmlstory.html
6	http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-empresario-que-fue-secuestrado-en-m-xico-1530960855-htmlstory.html	http://www.latimes.com/
7	http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-video-brasil-anota-un-autogol-y-b-lgica-1530901739-htmlstory.html	http://www.latimes.com/opinion/la-ol-enter-the-fray-htmlstory.html
8	http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-video-grave-error-de-muslera-le-regala-1530891077-htmlstory.html	http://www.latimes.com/politics/la-na-pol-kavanaugh-ford-hearing-htmlstory.html
9	http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-video-francia-acaricia-las-semifinales-1530889005-htmlstory.html	http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html
10	http://www.latimes.com/espanol/deportes/la-world-cup-espanol-updates-2018-aficionados-de-rusia-agradecen-a-1530473387-htmlstory.html	http://www.latimes.com/sports/highschool/varsity-times/la-sp-high-school-sports-updates-get-the-eric-sondheimer-s-latest-1490657918-htmlstory.html

6. North Korea

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-will-president-trump-bring-up-north-1528706649-htmlstory.html	http://www.latimes.com/politics/washington/la-na-pol-essential-washington-updates-2018-htmlstory.html
2	http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-trump-becomes-first-u-s-president-to-1528766878-htmlstory.html	http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html
3	http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-kimchi-flavored-ice-cream-a-small-1528697996-htmlstory.html	http://www.latimes.com/
4	http://www.latimes.com/world/la-fg-trump-kim-north-korea-summit-swift-north-korean-coverage-of-kim-jong-1528782028-htmlstory.html	http://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html

5	http://www.latimes.com/world/la-fg-north-korea-nuclear-20180420-story.html	http://www.latimes.com/about/la-privacypolicy-20180703-story.html
6	http://www.latimes.com/nation/la-na-pol-pompeo-north-korea-20181007-story.html	http://www.latimes.com/
7	http://www.latimes.com/world/la-fg-north-korea-environment-20171006-story.html	http://www.latimes.com/
8	http://www.latimes.com/world/asia/la-fg-north-korea-war-threat-20170925-story.html	http://www.latimes.com/
9	http://www.latimes.com/nation/la-fg-north-korea-foreign-minister-20170923-story.html	http://www.latimes.com/travel/la-tr-california-bucket-list-updates-2017-htmlstory.html
10	http://www.latimes.com/world/asia/la-fg-trump-north-korea-analysis-20180309-story.html	http://www.latimes.com/entertainment/la-et-tiff-2018-toronto-film-festival-updates-htmlstory.html

7. Hurricane Florence

	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/nation/la-na-hurricane-florence-20180913-story.html	http://www.latimes.com/about/la-privacypolicy-20180703-story.html
2	http://www.latimes.com/nation/la-na-hurricane-florence-20180913-story.html	http://www.latimes.com/
3	http://www.latimes.com/nation/la-na-hurricane-florence-carolinas-virginia-20180912-story.html	http://www.latimes.com/
4	http://www.latimes.com/nation/la-na-hurricane-florence-live-updates-hurricane-florence-meanders-through-1536975893-htmlstory.html	http://www.latimes.com/
5	http://www.latimes.com/nation/la-na-hurricane-florence-20180914-story.html	http://www.latimes.com/
6	http://www.latimes.com/nation/la-lb-838-47191-la-na-pol-trump-hurricane-florence-20180915-htmlstory.html	http://www.latimes.com/opinion/editorials/la-ed-scotus-wedding-cake-20180605-story.html
7	http://www.latimes.com/nation/la-na-hurricane-florence-live-updates-htmlstory.html	http://www.latimes.com/
8	http://www.latimes.com/nation/la-na-hurricane-florence-20180915-story.html	http://www.latimes.com/
9	http://www.latimes.com/nation/la-lb-838-47190-la-na-hurricane-florence-20180915-htmlstory.html	http://www.latimes.com/entertainment/la-et-tony-awards-2018-winners-htmlstory.html

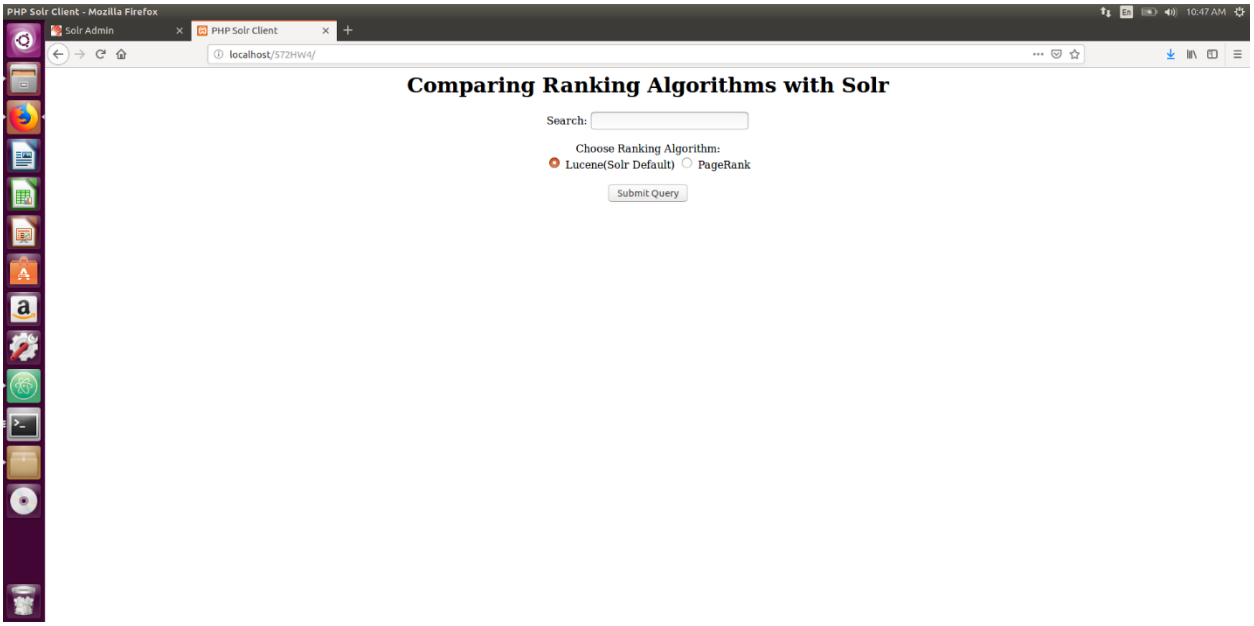
10	http://www.latimes.com/nation/la-lb-838-47192-la-na-florence-new-bern-20180915-htmlstory.html	http://www.latimes.com/

8. Paul Allen

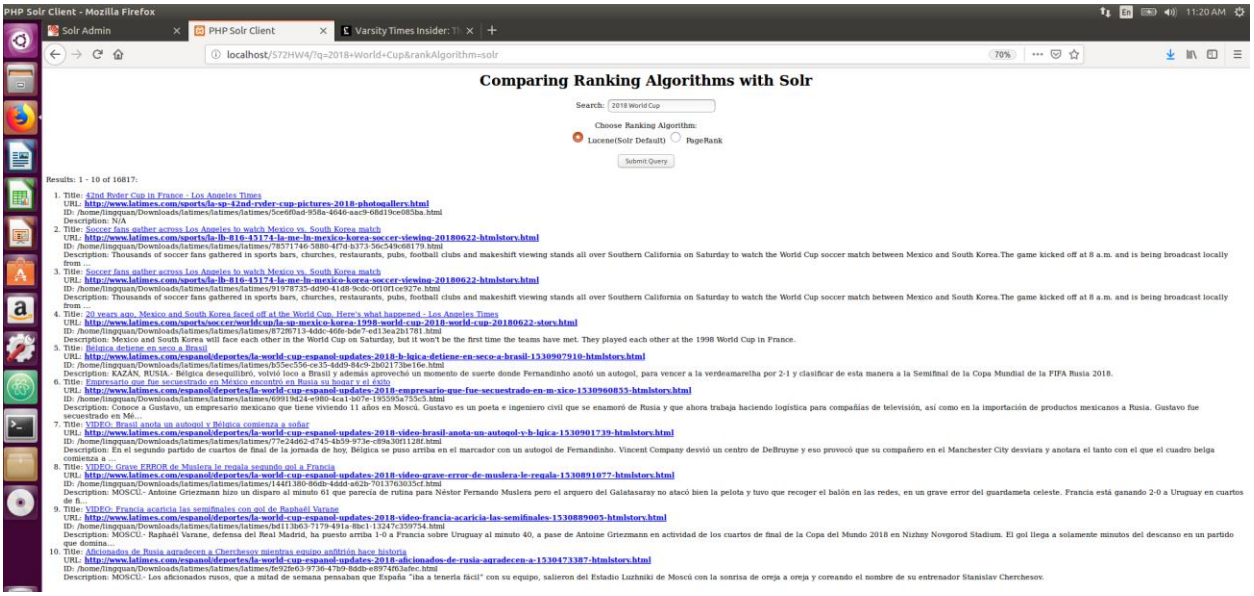
	Lucene(Solr Default)	PageRank
1	http://www.latimes.com/local/obituaries/la-fi-paul-allen-pictures-20181015-photogallery.html	http://www.latimes.com/
2	http://www.latimes.com/local/obituaries/la-fi-tn-paul-allen-obituary-20181015-story.html	http://www.latimes.com/politics/la-politics-newsletter-signup-2018-story.html
3	http://www.latimes.com/local/obituaries/la-fi-tn-paul-allen-obituary-20181015-story.html	http://www.latimes.com/
4	http://www.latimes.com/business/realestate/hot-property/la-fi-hotprop-paul-allen-beverly-hills-20180710-story.html	http://www.latimes.com/entertainment/la-et-2018-emmys-70th-emmy-awards-live-updates-htmlstory.html
5	http://www.latimes.com/business/realestate/hot-property/la-fi-hotprop-paul-allen-beverly-hills-20180710-story.html	http://www.latimes.com/politics/la-na-pol-kavanaugh-hearing-20180927-story.html
6	http://www.latimes.com/local/obituaries/la-fi-tn-paul-allen-obituary-20181015-story.html	http://www.latimes.com/
7	http://www.latimes.com/	http://www.latimes.com/
8	http://www.latimes.com/	http://www.latimes.com/
9	http://www.latimes.com/local/obituaries/la-me-maureen-paul-turlish-20180802-story.html	http://www.latimes.com/
10	http://www.latimes.com/local/obituaries/la-na-paul-laxalt-20180806-story.html	http://www.latimes.com/projects/la-fi-disney-anaheim-city-council/

Screenshots that describes the whole flow:

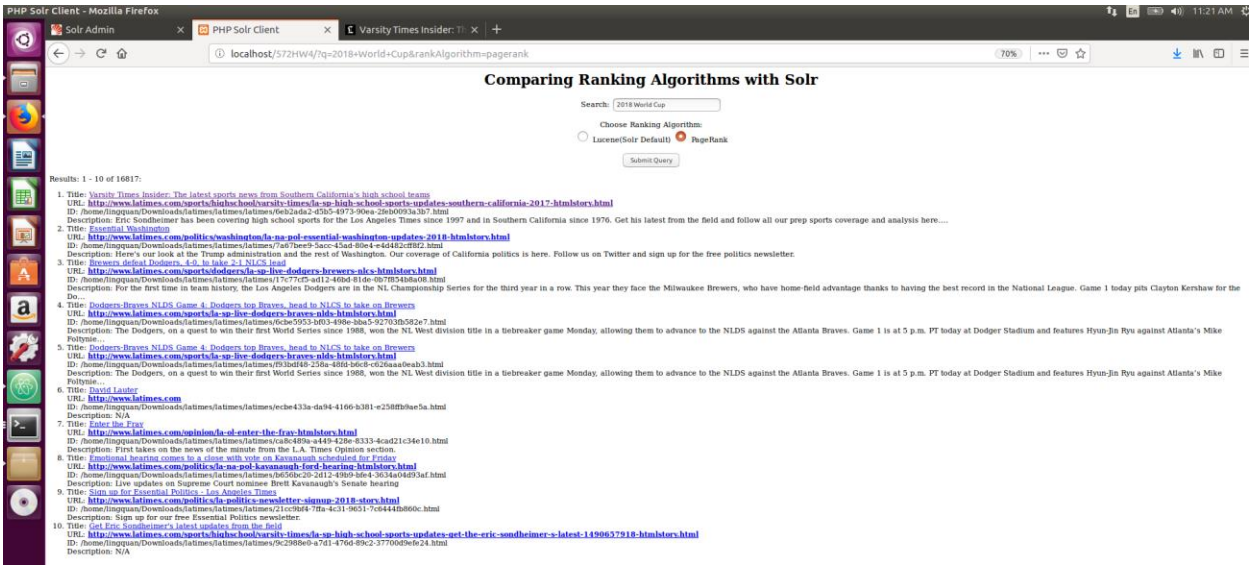
1.initial page



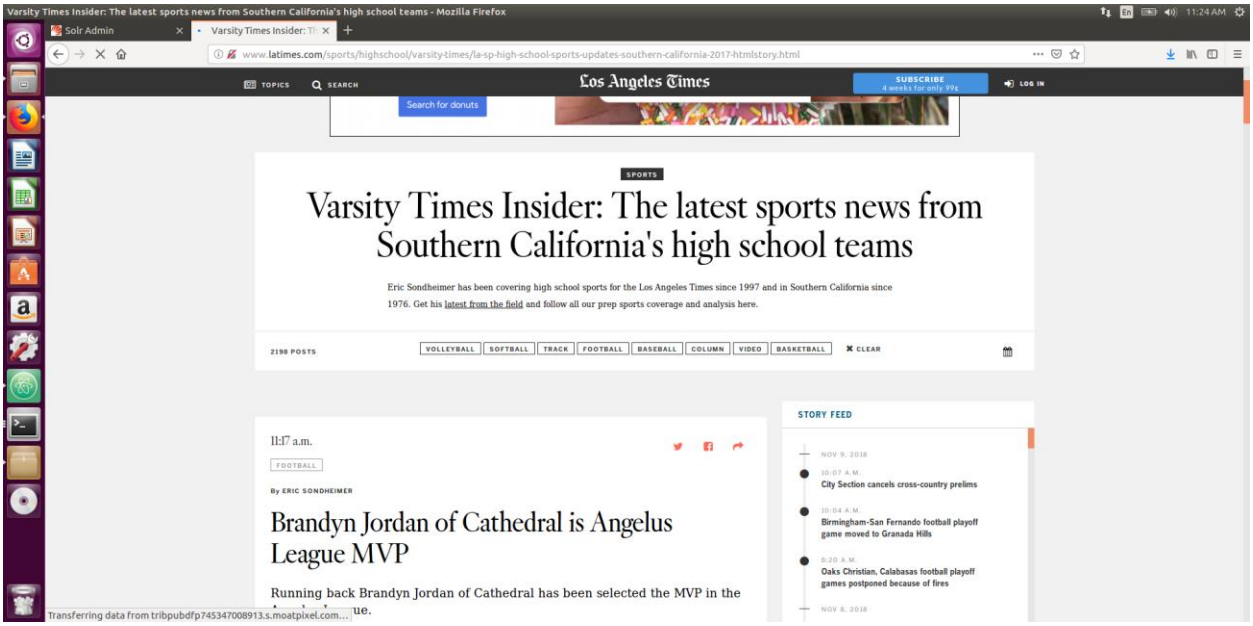
2. the results for Lucene(Default)



3. the results for PageRank



4. the actual webpage



Overlap Graph:

Query	Overlaps
Donald Trump	0
LA Lakers	1
Star Wars	0

Lebron James	1
2018 World Cup	0
North Korea	0
Hurricane Florence	0
Paul Allen	1