Ehsan Hosseinzadeh Khaligh Student ID 1112811250

HW4 Report:

1) Steps you followed to complete this assignment.

- a. Since I am using a macbook, there was no need to install Ubuntu.
- b. Both Java and php were already installed.

\$ java -version java version "1.8.0_291" Java(TM) SE Runtime Environment (build 1.8.0_291-b10) Java HotSpot(TM) 64-Bit Server VM (build 25.291-b10, mixed mode)

\$ php -v

WARNING: PHP is not recommended

PHP is included in macOS for compatibility with legacy software.

Future versions of macOS will not include PHP.

PHP 7.3.24-(to be removed in future macOS) (cli) (built: Jun 17 2021 21:41:15) (NTS)

Copyright (c) 1997-2018 The PHP Group

Zend Engine v3.3.24, Copyright (c) 1998-2018 Zend Technologies

- c. Downloaded Solr 7.7.2 and placed in /Desktop
- d. Opened a terminal and exported

export JAVA_HOME="/Library/Java/JavaVirtualMachines/jdk1.8.0_291.jdk/Contents/Home/" export PATH=\$JAVA HOME/bin:\$PATH

e. Started solr and created a core called 'myexample'

\$cd /Users/Ehsan/Desktop/solr-7.7.2 \$bin/solr start \$bin/solr create –c myexample \$ pwd /Users/Ehsan/Desktop/solr-7.7.2/server/solr/myexample

/Osers/Ensan/Desktop/son-7.7.2/server/son/myexample

f. Modified managed-schema to comment out

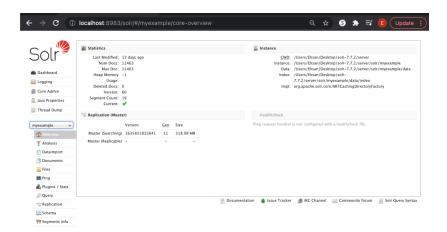
ehsans-MBP-2:conf Ehsan\$ pwd
/Users/Ehsan/Desktop/solr-7.7.2/server/solr/myexample/conf
Managed-schema file:
..
<copyField source="*" dest="_text_"/>

g. Used post command to post download FoxNews crawled data.

\$ pwd

/Users/Ehsan/Desktop/solr-7.7.2

Ehsan\$ bin/post -c myexample -filetypes html FOXNEWS/foxnews/



I tried using the 'q' box to test query output results with Solr UI, but realized there is a bug in the version I am using and it is mentioned also in Piazza.

Next Following the instructions, I commented out below line on solrconfig.xml

/Users/Ehsan/Desktop/solr-7.7.2/server/solr/myexample/conf/solrconfig.xml

Lastly reloaded the new config from UI.

Solr result URL:

```
**Tesponseleader*(* | "responseleader*(* | "respons
```

h. Cloned the solr-php-client repo and used this example as a base line to develop the search UI (Radio Buttons + Display result)

Placed the index.php inside the cloned repo.

ehsans-MBP-2:solr-php-client Ehsan\$ pwd /Users/Ehsan/Desktop/solr-php-client ehsans-MBP-2:solr-php-client Ehsan\$ Is

Apache COPYING README.md composer.json index.php my.php tests

Once the development of php script finished and lucene was selected as a default, I ran the the local server on localhost:9000

ehsans-MBP-2:solr-php-client Ehsan\$ pwd /Users/Ehsan/Desktop/solr-php-client ehsans-MBP-2:solr-php-client Ehsan\$ php -S localhost:9000

There were some additional changes inside php such as:

\$URLtoHTMLcsvFile = file('/Users/Ehsan/Desktop/solr-7.7.2/FOXNEWS/URLtoHTML_fox_news.csv'); \$solr = new Apache_Solr_Service('localhost', 8983, '/solr/myexample');

By the end of this step, I was able to run queries and get results back using Lucene.

<u>Side note: Since I have Astigmatism and waiting for my new glasses, I added a logic to highlight Title in yellow for testing and compreting results (below screenshots).</u>

i. Next, I downloaded https://jsoup.org/download jar file and created a project in Eclipse to generate the edgelist.txt given the code template.

Maily two hashmaps have been used to map URLs and HTML files when reading the csv doc.

j. Once edgelist.txt generated, I installed networkx using python pip to generate external_pageRankFile.txt using DiGraph().

PageRank algorithm Configuration:

nx.pagerank(self.mygraph, alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None)

k. Once external_pageRankFile.txt is generated, I placed a copy of the file on /Users/Ehsan/Desktop/solr-7.7.2/server/solr/myexample/data.

Next, I modified managed-schema to add:

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

Finally the listener is added to solarconfig.xml and the core is reloaded.

I. Finally, the search UI also returned documents with PageRank and I started running the query terms to get top 10 search results shown below and have the overlap graph.

pageRankFile desc is used on php as both descending and ascending were options.

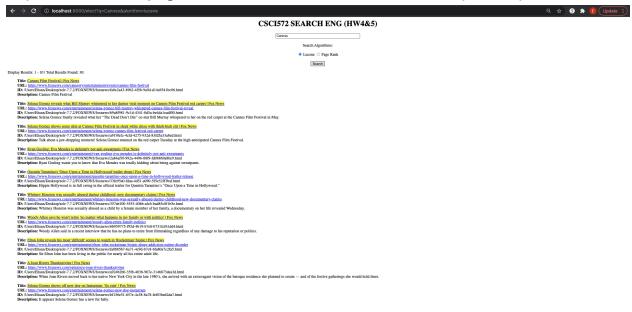
2) Explanation regarding why some pages have higher PageRank values than others.

PageRank works by counting the number and quality of links to a page to determine how important the website is. For this reason, mostly home or parent pages are shown on top search results despite not being the most relevant item to the query term(s). As an example, if we look at the 'Senate' search we can see that Lucne returns more specific pages/articles and page rank returns Fox news Parent pages. The underlying assumption is that more important websites are likely to receive more links from other websites in page rank. In contrast, Lucne uses a combination of TF-IDF & vector closeness and to get a better result. If a combination of a page rank alongside TF-IDF and/or Vector space is used, better results can be seen from the search engine.

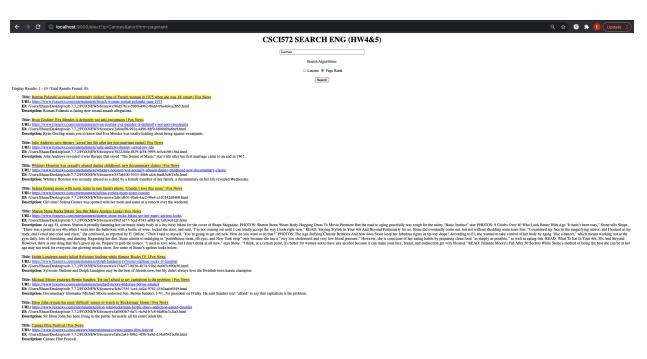
3.1) Screenshot of the initial page where you enter the query in the input box.



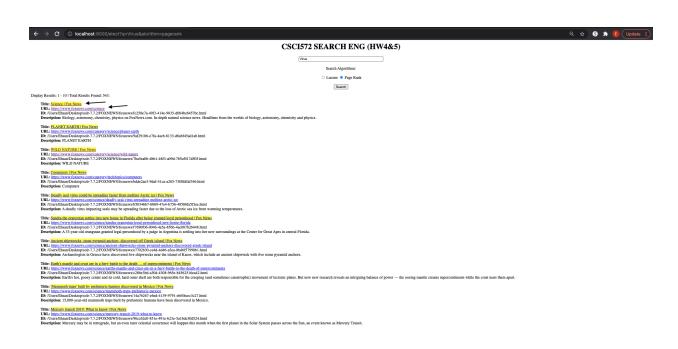
3.2) Screenshot of page which shows the results for Lucene(Default).

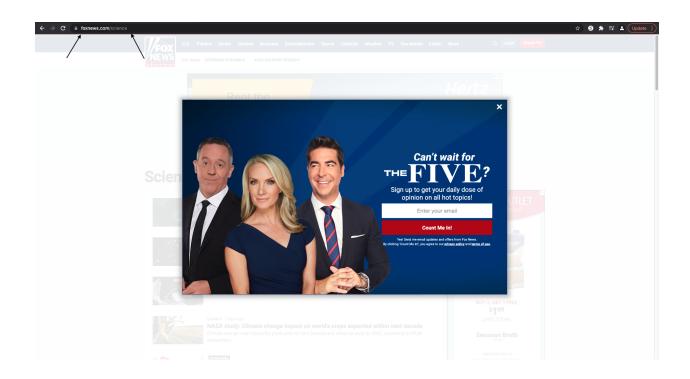


3.3) Screenshot of page which shows the results for PageRank.



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





4) A table containing a list of URLs of the top ten results produced by both ranking methods for each query.

Cannes:

Lucene	PageRank
https://www.foxnews.com/category/entertain ment/events/cannes-film-festival	https://www.foxnews.com/entertainment/frenc h-woman-roman-polanski-rape-1975
https://www.foxnews.com/entertainment/selen a-gomez-bill-murray-whispered-cannes-film-f estival-reveal	https://www.foxnews.com/entertainment/ryan- gosling-eva-mendes-is-definitely-not-anti-swe atpants
https://www.foxnews.com/entertainment/selen a-gomez-cannes-film-festival-red-carpet	https://www.foxnews.com/entertainment/julie- andrews-therapy-saved-my-life
https://www.foxnews.com/entertainment/ryan- gosling-eva-mendes-is-definitely-not-anti-swe atpants	https://www.foxnews.com/entertainment/whitney-houston-was-sexually-abused-during-childhood-new-documentary-claims
https://www.foxnews.com/entertainment/quentin-tarantino-once-upon-a-time-in-hollywood-trailer-release	https://www.foxnews.com/entertainment/selen a-gomez-mom-sister-concert
https://www.foxnews.com/entertainment/whitney-houston-was-sexually-abused-during-child-hood-new-documentary-claims	https://www.foxnews.com/entertainment/shar on-stone-rocks-bikini-see-her-many-ageless-l ooks
https://www.foxnews.com/entertainment/wood y-allen-retire-family-politics	https://www.foxnews.com/entertainment/dolp h-lundgren-sylvester-stallone-rocky-iv-hospita l
https://www.foxnews.com/entertainment/elton-john-rocketman-biopic-drugs-addiction-eating-disorder	https://www.foxnews.com/entertainment/mich ael-moore-endorses-bernie-sanders
https://www.foxnews.com/opinion/a-joan-rivers-thanksgiving	https://www.foxnews.com/entertainment/elton-john-rocketman-biopic-drugs-addiction-eating-disorder
https://www.foxnews.com/entertainment/selen a-gomez-new-dog-instagram	https://www.foxnews.com/category/entertain ment/events/cannes-film-festival

Congress:

Lucene	PageRank
https://www.foxnews.com/us/navy-vet-donate s-gold-star-families-stories-to-library-of-congress	https://www.foxnews.com/opinion
https://www.foxnews.com/opinion/judge-andrew-napolitano-congress-constitutional-power-declare-war	https://www.foxnews.com/tech
https://www.foxnews.com/opinion/rep-andy-biggs-congress-debt-ceiling-budget-deal	https://www.foxnews.com/category/politics/foreign-policy
https://www.foxnews.com/opinion/gutfeld-congress-attacking-zuckerberg	https://www.foxnews.com/category/world/conf licts
https://www.foxnews.com/politics/jon-stewart-blasts-congress-over-9-11-fund-hearing-drove-me-nuts-there-were-empty-seats	https://www.foxnews.com/official-polls
https://www.foxnews.com/category/us/congress	https://www.foxnews.com/category/politics/executive
https://www.foxnews.com/faith-values/bible-s chool-drew-brees-day-congress	https://www.foxnews.com/category/tech/topic s/security
https://www.foxnews.com/opinion/andy-biggs-congress-judges-immigration-laws	https://www.foxnews.com/opinion/mike-hucka bee-lets-fight-for-fabricated-structural-steel-m anufacturing-jobs
https://www.foxnews.com/opinion/bradley-bla keman-trade-pact-with-canada-and-mexico-of fers-incredible-benefits-congress-should-appr ove-it	http://www.foxnews.com/trending
https://www.foxnews.com/politics/trump-tells- congress-to-immediately-approve-mexico-can ada-trade-deal-says-us-moving-past-stupid-y ears-of-trade-policy	https://www.foxnews.com/media/elijah-cummings-seat-kimberly-klacik-baltimore-election-republican

Democrats:

Lucene	PageRank
https://www.foxnews.com/opinion/rnc-chair-mcdaniel-democrats-impeachment-vote-is-baseless-political-vendetta-trump-did-nothing-wrong	https://www.foxnews.com/opinion
https://www.foxnews.com/politics/federal-judg e-house-democrats-mazars-subpoena-trump- financial-record	https://www.foxnews.com/us
https://www.foxnews.com/politics/democrats- arent-actually-impeaching-the-president-yet-b ut-they-are-digging-deep	https://www.foxnews.com
https://www.foxnews.com/media/democrats-disdain-working-class-trump-jr-fox-friends	https://www.foxnews.com/politics
https://www.foxnews.com/opinion/david-bossi e-baseless-trump-impeachment-inquiry-by-de sperate-democrats-makes-dramatic-flip-flop	https://www.foxnews.com/category/us/personal-freedoms
https://www.foxnews.com/opinion/rep-andy-biggs-lipstick-impeachment	https://www.foxnews.com/category/politics/house-of-representatives
https://www.foxnews.com/media/sen-lindsey-graham-every-american-should-be-bothered-by-what-democrats-are-doing-to-trump	https://www.foxnews.com/category/politics/senate
https://www.foxnews.com/opinion/david-bossi e-trump-impeachment-vote-is-democratic-dec laration-of-war-republicans-must-declare-war- on-dems	https://www.foxnews.com/category/entertain ment/music
https://www.foxnews.com/tech/google-facebook-and-twitter-staffs-splurge-on-democrats-ahead-of-midterms	https://www.foxnews.com/category/politics/foreign-policy
https://www.foxnews.com/media/democrats-impeachment-abuse-ken-starr	https://www.foxnews.com/category/world/conf licts

Patriot Movement:

Lucene	PageRank
https://www.foxnews.com/politics/trump-antifa -terrorist-portland-protests	https://www.foxnews.com/science
https://www.foxnews.com/world/iranian-cruise -missiles-drones-saudi-oil	https://www.foxnews.com
https://www.foxnews.com/transcript/gutfeld-whats-a-liberal-to-do-when-a-republican-turns-out-to-be-better-at-compassion	https://www.foxnews.com/tech
https://www.foxnews.com/media/fox-nation-patriot-awards-american-heroes-st-petersburg	https://www.foxnews.com/category/science/pl anet-earth
https://www.foxnews.com/media/marine-vet- most-valuable-patriot-awards-congress	https://www.foxnews.com/category/entertain ment/movies
https://www.foxnews.com/media/fox-nation-patriot-awards	https://www.foxnews.com/entertainment/gwyn eth-paltrow-relationship-harvey-weinstein
https://www.foxnews.com/media/tomi-lahren- patriot-awards-fox-nation	https://www.foxnews.com/politics/aoc-bloomb erg-purchase-election
https://www.foxnews.com/faith-values/veterans-s-serve-disaster-relief	https://www.foxnews.com/world/ukraine-russi a-separatists-pullback
https://www.foxnews.com/entertainment/netflix-pulls-patriot-act-with-hasan-minhaj-episode-after-getting-valid-legal-complaint-from-saudiarabia	https://www.foxnews.com/science/earths-man tle-and-crust-are-in-a-fiery-battle-to-the-death -of-supercontinents
https://www.foxnews.com/tech/netflix-defends -pulling-patriot-act-episode-in-saudi-arabia	https://www.foxnews.com/auto/2020-ram-150 0-built-to-serve-pickup-salutes-the-u-s-militar y

Republicans:

Lucene	PageRank
https://www.foxnews.com/politics/house-republicans-storm-impeachment-inquiry-deposition-what-went-down	https://www.foxnews.com/opinion
https://www.foxnews.com/opinion/tom-del-beccaro-republicans-must-win-the-impeachment-trial-and-they-can-by-following-these-five-steps	https://www.foxnews.com
https://www.foxnews.com/politics/republicans- hunter-biden-ukraine-whisteleblower-impeach ment-witness	https://www.foxnews.com/politics
https://www.foxnews.com/politics/bill-weld-thri lled-other-republicans-are-challeneging-trump -in-primary	https://www.foxnews.com/category/politics/house-of-representatives
https://www.foxnews.com/politics/house-republicans-plan-to-call-adam-schiff-to-testify-in-impeachment-inquiry-call-him-a-fact-witness	https://www.foxnews.com/category/politics/senate
https://www.foxnews.com/politics/republicans-impeachment-hearings-lineup	https://www.foxnews.com/category/world/conf licts
https://www.foxnews.com/politics/trump-settle s-score-with-republicans-who-didnt-embrace- him-in-midterms	https://www.foxnews.com/official-polls
https://www.foxnews.com/opinion/newt-gingrich-house-republicans-2020-retirements	https://www.foxnews.com/category/world/sca ndals
https://www.foxnews.com/opinion/matt-gorma n-kavanaugh-republicans	https://www.foxnews.com/category/politics/elections
https://www.foxnews.com/media/joe-biden-republicans-epiphany-trump	https://www.foxnews.com/politics

Senate:

Lucene	PageRank
https://www.foxnews.com/politics/trump-senat e-trial-impeachment-inquiry-reporters-notebo ok	https://www.foxnews.com/opinion
https://www.foxnews.com/category/politics/20 20-senate-races	https://www.foxnews.com/travel
https://www.foxnews.com/media/jeff-sessions -former-embattled-trump-attorney-general-an nounces-bid-for-old-us-senate-seat	https://www.foxnews.com/us
https://www.foxnews.com/opinion/newt-gingri ch-left-viciously-attacks-mcconnell-because-h es-successful-senate-leader-and-trump-ally	https://www.foxnews.com/health
https://www.foxnews.com/faith-values/pastor- andrew-brunson-senate-trump-prayer	https://www.foxnews.com/shows
https://www.foxnews.com/politics/jeff-session s-alabama-senate-campaign-expected	https://www.foxnews.com/entertainment
https://www.foxnews.com/category/politics/elections/senate	https://www.foxnews.com/world
https://www.foxnews.com/politics/trump-sessions-against-senate-run-alabama	https://www.foxnews.com/science
https://www.foxnews.com/media/judge-pirro-jeff-sessions-senate-run-fox-friends	https://www.foxnews.com
https://www.foxnews.com/politics/virginia-dem ocrat-accused-of-sex-with-teen-secretary-win s-state-senate-seat	https://www.foxnews.com/politics

Olympics 2020:

Lucene	PageRank
https://www.foxnews.com/sports/alex-morgan -2020-olympics-child-birth	https://www.foxnews.com/opinion
https://www.foxnews.com/sports/naomi-osaka -us-citizenship-japan-2020-olympics	https://www.foxnews.com/entertainment
https://www.foxnews.com/category/sports/olympics	https://www.foxnews.com
https://www.foxnews.com/tech/russian-hacker s-attack-possible-olympic-ban	https://www.foxnews.com/politics
https://www.foxnews.com/category/sports/soccer	https://www.foxnews.com/tech
https://www.foxnews.com/person/c/barnini-ch akraborty	https://www.foxnews.com/category/health/heart-health
https://www.foxnews.com/category/sports/tennis	https://www.foxnews.com/lifestyle
https://www.foxnews.com/category/person/ale x-morgan	https://www.foxnews.com/category/us/personal-freedoms
https://www.foxnews.com/sports/kikkan-randa ll-new-york-city-marathon-skiing-cancer	https://www.foxnews.com/category/entertain ment/style
https://www.foxnews.com/sports	https://www.foxnews.com/category/entertain ment/tv

Stock:

Lucene	PageRank
https://www.foxnews.com/tech/jeff-bezos-to-be-no-2-richest-behind-bill-gates-after-amazone-stock-drop	https://www.foxnews.com
https://www.foxnews.com/transcript/cavuto-looking-back-at-the-stock-market-crash-30-years-ago	https://www.foxnews.com/category/us/econo my
https://www.foxnews.com/transcript/cavuto-the-e-markets-are-easier-to-take-stock-of-than-life	https://www.foxnews.com/science/revolutiona ry-war-rifle-stolen-50-years-ago-on-display
https://www.foxnews.com/category/shows/your-world/transcript/common-sense	https://www.foxnews.com/entertainment/demi -lovato-first-interview-year
https://www.foxnews.com/category/tech/comp anies/microsoft	https://www.foxnews.com/media/chris-wallace -elizabeth-warren-medicare-for-all-plan
https://www.foxnews.com/person/c/gordon-ch ang	https://www.foxnews.com/tech/hp-recalls-mor e-laptop-lithium-ion-batteries-after-reports-of- overheating-melting-and-charring
https://www.foxnews.com/category/tech/companies/amazon	https://www.foxnews.com/tech/heres-how-the -trump-tariffs-hike-may-affect-prices-for-pc-pa rts
https://www.foxnews.com/sports/royals-hire-ex-cardinals-skipper-mike-matheny-as-manager	https://www.foxnews.com/category/tech/companies/amazon
https://www.foxnews.com/category/person/sheryl-sandberg	https://www.foxnews.com/opinion/max-lucado -fearless-parenting
https://www.foxnews.com/auto/fiat-chrysler-sa les-boss-reid-bigland-sues-company-for-violat ing-whistleblower-protections	https://www.foxnews.com/auto/electric-ford-mustang-lithium

Virus:

Lucene	PageRank
https://www.foxnews.com/tech/computer-virus -cripples-tribune-publishing-los-angeles-times	https://www.foxnews.com/science
https://www.foxnews.com/science/deadly-seal- -virus-spreading-melting-arctic-ice	https://www.foxnews.com/category/science/pl anet-earth
https://www.foxnews.com/health/connecticut-t hird-mosquito-eee-death	https://www.foxnews.com/category/science/wild-nature
https://www.foxnews.com/category/tech/topics/virus	https://www.foxnews.com/category/tech/topic s/computers
https://www.foxnews.com/health/eee-eastern-equine-encephalitis-massachusetts-man-killed	https://www.foxnews.com/science/deadly-seal- -virus-spreading-melting-arctic-ice
https://www.foxnews.com/health/what-is-norovirus-3-things-to-know-about-the-vomit-inducing-bug	https://www.foxnews.com/science/sandra-ora ngutan-legal-personhood-new-home-florida
https://www.foxnews.com/health/pregnant-women-who-get-flu-shot-are-less-likely-to-be-hospitalized-from-virus-cdc-says	https://www.foxnews.com/science/ancient-shipwrecks-stone-pyramid-anchors-discovered-greek-island
https://www.foxnews.com/health/flu-season-2 019-vaccine	https://www.foxnews.com/science/earths-man tle-and-crust-are-in-a-fiery-battle-to-the-death -of-supercontinents
https://www.foxnews.com/category/health/infectious-disease/zika	https://www.foxnews.com/science/mammoth-traps-prehistoric-mexico
https://www.foxnews.com/health/british-mom-son-flu	https://www.foxnews.com/science/mercury-transit-2019-what-to-know

5) An overlap graph

