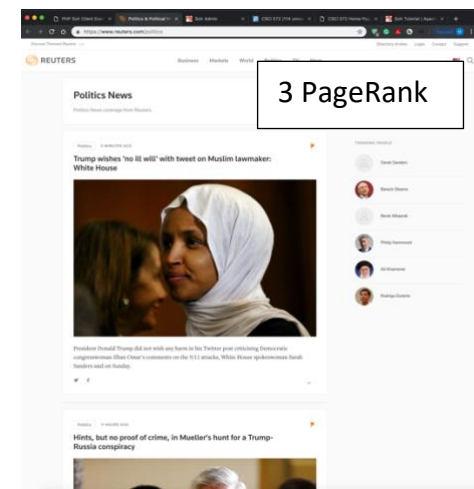
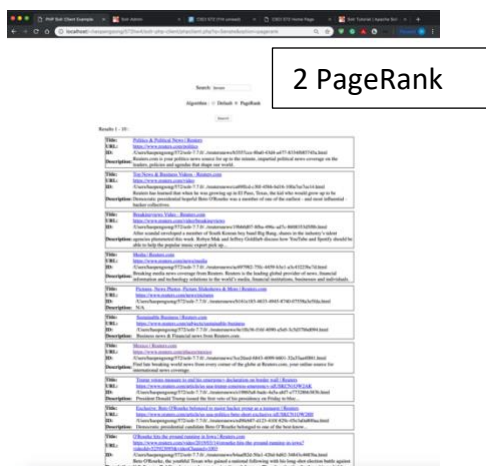
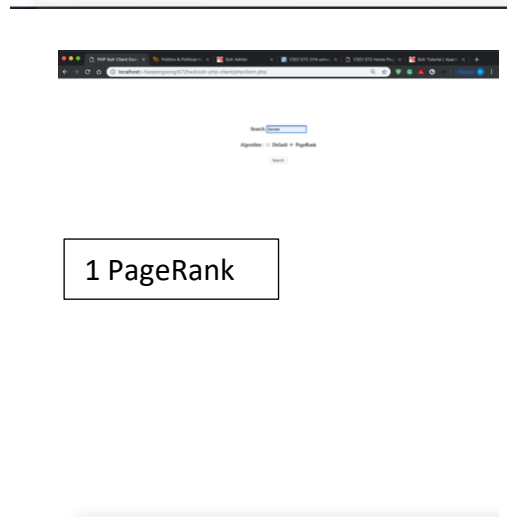
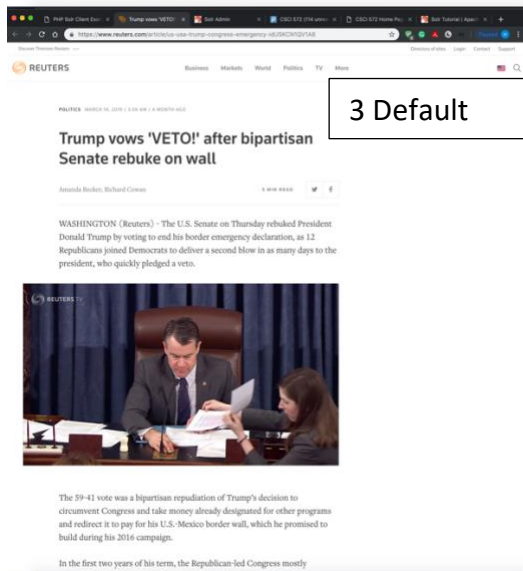


1. Steps that I've followed to complete this assignment
 - a. Read through the tutorial about Lucene on apache.org.
 - b. Followed the tutorial about indexing HTML files in Solr where I set up the Solr server and configured the server.
 - c. Used Tika to parse and index all the html files.
 - d. Used Jsoup in Java to extract links from downloaded webpages and generated a directed graph. Vertices in the graph are webpages and edges between vertices are links pointing to or from them.
 - e. Created an adjacency list for the graph using networkx library in Python with input generated from step d.
 - f. Computed the PageRank for the downloaded files using networkx in Python with input made from step e. The parameters I used are: $\alpha=0.85$, $\text{personalization}=\text{None}$, $\text{max_iter}=100$, $\text{tol}=1e-06$, $\text{nstart}=\text{None}$, $\text{weight}=\text{'weight'}$, $\text{dangling}=\text{None}$. Output the result in a txt file.
 - g. Placed the PageRank score file (external_pageRankFile.txt) inside the data folder of the core in Solr.
 - h. Edited the managed-schema files to refer to this file.
 - i. Added two eventListeners within solrconfig.xml file to enable core to access the score files each time the index is reloaded.
 - j. Set up and configured the apache server on the laptop.
 - k. Started the server and hosted a PHP script. The script embedded with html content with a query box where it allows users to enter a query. The query will be processed by the PHP program at the server which eventually sends a Http GET request to Solr.
 - l. Implemented a program to format the JSON response from Solr server and displayed the results to the user. If id field doesn't contain any data or there is no id field at all, the program will fetch the corresponding URL from the provided map file. If the description field is missing or no value, it will show as "N/A".
 - m. Issued queries as described on the assignment instruction pdf file.
2. Explanation regarding why some pages have higher PageRank values than others
Answer: Page with a lot of other pages point to it tend to have higher score. Page with a link from an important page (page with high PageRank score) tends to have higher score as well.
3. Document the flow for any single query using screenshots



4. A table containing list of URLs of the top ten results produced by both ranking methods

Venezuela	Lucene Default	PageRank	OverLap
1	https://www.reuters.com/investigates/special-report/venezuela-russia-rosneft/	https://www.reuters.com/	0
2	https://www.reuters.com/article/us-venezuela-russia-rosneft-special-repo-idUSKCN1QV1HN	https://www.reuters.com/finance/markets	
3	https://www.reuters.com/article/us-usa-venezuela-sanctions-idUSKCN1QV2VS	https://www.reuters.com/politics	
4	https://www.reuters.com/article/us-venezuela-politics-usa-abrams-idUSKCN1QW2HM	https://www.reuters.com/news/world	
5	https://www.reuters.com/article/us-venezuela-politics-tankers-idUSKCN1QW2L6	https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE2SD	
6	https://www.reuters.com/article/us-venezuela-russia-rosneft-money-idUSKCN1QV1HS	https://www.reuters.com/video	
7	https://www.reuters.com/journalists/nelson-bocanegra	https://www.reuters.com/video/breakingviews	
8	https://www.reuters.com/journalists/lesley-wroughton	https://www.reuters.com/investigates/section/homepage/	
9	https://www.reuters.com/journalists/roberta-rampton	https://www.reuters.com/subjects/euro-zone	
10	https://www.reuters.com/article/us-venezuela-politics-usa-banks-idUSKCN1QN2QE	https://www.reuters.com/places/india	
Senate	Lucene Default	PageRank	
1	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/politics	2
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	
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/breakingviews	
4	https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N	https://www.reuters.com/news/media	
5	https://www.reuters.com/article/us-usa-politics-beto-orourke-special-rep-idUSKCN1QW26N	https://www.reuters.com/news/pictures	
6	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	https://www.reuters.com/subjects/sustainable-business	
7	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	https://www.reuters.com/places/mexico	
8	https://www.reuters.com/politics	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	
9	https://www.reuters.com/article/us-congo-politics-idUSKCN1QW28R	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	
10	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV00E	https://www.reuters.com/video/2019/03/14/orourke-hits-the-ground-running-in-iowa?videoid=525923095&videoChannel=1003	
Democrats	Lucene Default	PageRank	
1	https://www.reuters.com/article/us-usa-election-medicare-idUSKBN1QU189	https://www.reuters.com/	2
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/subjects/supreme-court	
4	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoid=520293110&videoChannel=13976	https://www.reuters.com/breakingviews	
5	https://www.reuters.com/article/us-usa-election-biden-idUSKCN1QH1EX	https://www.reuters.com/news/archive/politicsNews	
6	https://www.reuters.com/article/us-usa-trump-budget-idUSKBN1QS12D	https://www.reuters.com/article/us-usa-election-sanders-idUSKCN1QW2ZJ	
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-trump-congress-emergency-idUSKCN1QW2AK	
8	https://www.reuters.com/article/us-usa-trump-russia-congress-idUSKCN1QV23E	https://www.reuters.com/article/us-usa-politics-beto-short-exclusive-idUSKCN1QW28H	
9	https://www.reuters.com/politics	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	
10	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-election-booker-idUSKCN1QW2P4	
Republicans	Lucene Default	PageRank	
1	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/politics	3
2	https://www.reuters.com/video/2019/02/28/republicans-bash-cohen-democrats-at-cpac?videoid=520293110&videoChannel=13976	https://www.reuters.com/news/us	
3	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/news/archive/politicsNews	
4	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	
5	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	
6	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QW13C	
7	https://www.reuters.com/article/us-california-fires-trump-idUSKCN1P31ND	https://www.reuters.com/article/us-usa-immigration-wall-insight-idUSKCN1QW1GR	
8	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	https://www.reuters.com/article/us-usa-trump-russia-congress-idUSKCN1QV23E	
9	https://www.reuters.com/article/us-usa-tax-survey-idUSKCN1QW1BY	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	
10	https://www.reuters.com/article/us-usa-politics-trump-idUSKCN1QH29L	https://www.reuters.com/article/us-usa-court-census-idUSKCN1QW2MC	
Patriot Movement	Lucene Default	PageRank	
1	https://www.reuters.com/article/us-usa-trump-turkey-idUSKCN1QV328	https://www.reuters.com/markets/bonds	3
2	https://www.reuters.com/article/us-france-protests-idUSKCN1Q325A	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C	
3	https://www.reuters.com/article/uk-people-rkelly-metoo-idUKKCN1QC0V7	https://www.reuters.com/article/canada-forex-idUSL1N21210B	
4	https://www.reuters.com/article/us-britain-climatechange-education-idUSKCN1QW2QU	https://www.reuters.com/article/us-climate-change-youth-usa-idUSKCN1QW1BE	
5	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/investigates/special-report/myanmar-suukyi-history/	
6	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	
7	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	https://www.reuters.com/article/us-slovakia-election-president-idUSKCN1QW2Z3	

8	https://www.reuters.com/article/us-italy-china-analysis-idUSKCN1QW1E2	https://www.reuters.com/article/us-climate-change-youth-usa-idUSKCN1QW1BE	
9	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C	https://www.reuters.com/article/us-climate-change-youth-idUSKCN1QW01S	
10	https://www.reuters.com/article/us-usa-election-orourke-idUSKCN1QV13C	https://www.reuters.com/article/us-auction-christies-georgemichael-art-idUSKCN1QP1EY	
Oscar 2019	Lucene Default	PageRank	
1	https://www.reuters.com/article/us-awards-oscars-women-idUSKCN1Q7130	https://www.reuters.com/	0
2	https://www.reuters.com/article/us-awards-oscars-diversity-idUSKCN1QE0N4	https://www.reuters.com/tools/mobile	
3	https://www.reuters.com/article/us-aviancaholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/finance/markets	
4	https://www.reuters.com/article/us-aviancaholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/politics	
5	https://www.reuters.com/article/us-aviancaholdings-fleet-idUSKCN1QW2XT	https://www.reuters.com/news/world	
6	https://www.reuters.com/journalists/pei-li	https://www.reuters.com/finance	
7	https://www.reuters.com/article/us-britain-lgbt-celebrities-idUSKCN1QW2KY	https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE2SD	
8	https://www.reuters.com/theWire	https://www.reuters.com/video	
9	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI	https://www.reuters.com/video/breakingviews	
10	https://www.reuters.com/article/us-britain-lgbt-rights-idUSKCN1QW2QI	https://www.reuters.com/investigates/section/homepage/	
Channel	Lucene Default	PageRank	
1	https://www.reuters.com/video/2019/03/11/in-brexits-next-act-theatrics-are-only-c?videoid=524541171&videoChannel=117760	https://www.reuters.com/politics	0
2	https://www.reuters.com/video/2019/03/08/uk-japan-scientists-study-radioactive-fu?videoid=523497511&videoChannel=-10464	https://www.reuters.com/article/us-reuters-editorial-leadership-idUSKBN1KE2SD	
3	https://www.reuters.com/video/2019/03/13/woods-neck-not-painful-ahead-of-players?videoid=525458826&videoChannel=79	https://www.reuters.com/video	
4	https://www.reuters.com/video/2019/01/16/saudi-arabia-set-for-regions-largest-win?videoid=504158943&videoChannel=118285	https://www.reuters.com/video/breakingviews	
5	https://www.reuters.com/video/2019/03/11/tanzanian-mother-finds-fame-with-soccer?videoid=524707453&videoChannel=73	https://www.reuters.com/markets/bonds	
6	https://www.reuters.com/video/2019/03/13/malawi-floods-wreak-havoc-cyclone-idai-e?videoid=525916362&videoChannel=73	https://www.reuters.com/news/science	
7	https://www.reuters.com/video/2019/03/13/guardiolas-city-demolish-schalke-to-crui?videoid=525247223&videoChannel=79	https://www.reuters.com/places/china	
8	https://www.reuters.com/video/2019/03/04/breakingviews-tv-lyft-the-veil?videoid=521916948&videoChannel=117766	https://www.reuters.com/news/entertainment/arts	
9	https://www.reuters.com/video/2019/02/08/amazon-rethinks-ny-hq-over-opposition-re?videoid=512785961&videoChannel=5	https://www.reuters.com/news/us	
10	https://www.reuters.com/video/2019/02/01/jailed-reuters-reporters-appeal-to-myanmar?videoid=510223083&videoChannel=-13280	https://www.reuters.com/subjects/supreme-court	
Wall	Lucene Default	PageRank	
1	http://www.reuters.com/investigates/section/wall-streets-way/	https://www.reuters.com/	1
2	https://www.reuters.com/article/us-usa-trump-budget-wall-exclusive-idUSKBN1QR0CW	https://www.reuters.com/finance/markets	
3	https://www.reuters.com/article/us-usa-immigration-wall-insight-idUSKCN1QW1GR	https://www.reuters.com/politics	
4	https://www.reuters.com/politics	https://www.reuters.com/finance	
5	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/video	
6	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/video/breakingviews	
7	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/investigates/section/homepage/	
8	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QW2AK	https://www.reuters.com/places/china	
9	https://www.reuters.com/article/us-usa-trump-congress-emergency-idUSKCN1QV1A8	https://www.reuters.com/news/us	
10	https://www.reuters.com/journalists/noel-randewich	https://www.reuters.com/news/media	

5. An overlap graph showing the amount of overlap between the two ranking strategies for all queries. By overlap we mean the number of results that appear in the top ten under both ranking methods. Axes should be labelled correctly to indicate query name and overlap number. If no overlaps at all, do not plot an empty graph, but you must mention it in your report.

Query	OverLap
Venezuela	0
Senate	2
Democrats	2
Republicans	3
Patriot Movement	3
Oscar 2019	0
Channel	0
Wall	1

