CSCI-572 Information Retrieval and Web Search Engines

Spring 2019 Assignment 4

Indexing the Web Using Solr

Approach

This assignment focuses on investigation of ranking strategies of two different algorithms used by a custom search engine. For building the search engine, an underlying opensource enterprise search platform called "Solr" was used. As of part of an earlier assignment a bunch of web pages were crawled from a popular news site (The Guardian) and they were ingested into "Solr". By default, Solr uses Lucene to facilitate ranking and determine how relevant a document is to a users' query. The default behavior of Lucene, which uses a combination of Vector Space and Boolean Models, served as the first algorithm to rank documents based on user queries. The default behavior of Solr can be changed by introducing a custom ranking scheme. For achieving this, an external file containing a mapping from a key field to a score should be provided to Solr. The page rank algorithm was used to generate the score corresponding to each key in this external file.

NetworkX library in python was used to generate the page rank scores. A NetworkX graph was created for this purspose. The directed graph was obtained by extracting all the outgoing links from each of the crawled web pages using JSoup library and creating a directed edge list for each of the links. The edge list was provided as an input to the page rank function. The following configuration was used for creating the directed graph in NetworkX:

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

Post generation of page-rank based scoring scheme, the file was placed in relevant location in Solr, and corresponding configuration changes were made in managed-schema.xml and solrcongig.xml, to enable Solr to refer to this scoring scheme. The Solr core was reloaded for the new changes to reflect.

A presentation layer was designed using "nodejs" to enable communication between the user and the solr framework. Restful web services were exposed and a single page "angularjs" application was created. A search box on the front end enabled a user to enter a search term and nodejs webservice forwarded the user request to solr. The response from solr was collected and presented back to the user in a presentable format. The user has the provision of viewing the ranking scheme either based on the default behavior of Lucene or based on the page rank results. The presentation layer provides the comparison results in two separate tables and overlaps are highlighted using the same colored marker. If the two schemes have overlaps, the overlap count is displayed on the top.

Why some pages have higher page ranks?

Page rank to a great extent is influenced by the number of webpages or links that are pointing to the webpage under consideration. Some pages have a higher page rank than the others, because they have greater number of in-links pointing to them. Each in-link to a web page shifts a portion of its own page rank probability mass to the page whose page rank is being computed.

Flow of a single query term is depicted below

1. This is the initial screen to enter text

2. Screen on clicking "Lucene" button displays result in default Lucene based search order

Q Bond

| Lucene | Page Rank | Clear | Results |

Term: **Bond**   *Result:* **1 - 10 of 316**

| LUCENE BASED SEARCH RESULTS |

1 **TV made America's bail system famous. Now reformers want to end it | US news | The Guardian**
http://www.theguardian.com/us-news/2017/aug/30/new-jersey-bail-reform-criminal-justice-bond-money
/home/koustav/shared/guardiannews/58780614-108e-4f09-bd34-f2ce690e456a.html
Bail money has trapped poor Americans for decades. But bondsmen are hitting back at attempts to reform the system

2 **Judge orders Florida shooting suspect to be held in jail without bond – video | US news | The Guardian**
http://www.theguardian.com/us-news/video/2018/feb/15/judge-orders-florida-shooting-suspect-to-be-held-in-jail-without-bond-video
/home/koustav/shared/guardiannews/283e90ef-c59b-4406-bd3a-613c2ce88993.html
Nikolas Cruz, 19, has appeared in court charged with the murder of 17 people and will be held in jail without bond, a Broward County judge has ordered

3 **Pennsylvania recount: Jill Stein to take fight to federal court after $1m bond set | US news | The Guardian**
http://www.theguardian.com/us-news/2016/dec/03/jill-stein-pennsylvania-recount-green-party
/home/koustav/shared/guardiannews/0cd64206-6eef-431f-b9a1-716bd9dd5bc6.html
Green party candidate vows to counter judge's 'outrageous demand' for $1m bond by seeking emergency relief in federal court

4 **Movies | The Guardian**
http://www.theguardian.com/us/film
/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42b2ea1.html
Latest Film news, comment and analysis from the Guardian, the world's leading liberal voice

5 **Sean Spicer | Page 3 of 11 | Us-news | The Guardian**
http://www.theguardian.com/us-news/sean-spicer
/home/koustav/shared/guardiannews/0d167158-16dc-4401-9e93-849859fac0d3.html
N/A

6 **US immigration | Us-news | The Guardian**
http://www.theguardian.com/us-news/usimmigration
/home/koustav/shared/guardiannews/ff85e14e-e0f4-4a9b-8f51-5cd3c3caea24.html
N/A

7 **John F Kennedy | Page 4 of 12 | Us-news | The Guardian**
http://www.theguardian.com/us-news/john-f-kennedy
/home/koustav/shared/guardiannews/edb299f6-e3e0-4223-8b7e-7fda7c49069f.html
N/A

8 **Walter Scott shooting | Us-news | The Guardian**
http://www.theguardian.com/us-news/walter-scott-shooting
/home/koustav/shared/guardiannews/b653978a-6491-4671-b98e-7a1720e6cfa6.html
N/A

9 **Paul Manafort in court: a not guilty plea, $10m bond – and no sign of repentance | US news | The Guardian**
http://www.theguardian.com/us-news/2017/oct/30/paul-manafort-court-not-guilty-plea-charges
/home/koustav/shared/guardiannews/b88c41ef-a380-45c2-b759-44e76fec2051.html
On Monday afternoon, Trump's ex-campaign chair walked into a Washington courtroom – and the ensuing hour made it clear how steeply he has fallen

10 **Chicago's 'Skullcap Crew': band of police accused of brutality evade discipline | US news | The Guardian**
http://www.theguardian.com/us-news/2016/aug/03/chicago-skullcap-crew-police-brutality
/home/koustav/shared/guardiannews/7ca30cb9-6730-4ec1-aa39-81eaf956260f.html
Dogged by allegations of abuse, members of the group have been named in more than 20 federal lawsuits – yet have won repeated praise from department

3. Screen on clicking "Page Rank" button displays result in default Page Rank based search order

Q Bond

| Lucene | Page Rank | Clear | Results |

Result: **1 - 10 of 316**

PAGE RANK BASED SEARCH RESULTS

⊕ 1 **Movies | The Guardian**
http://www.theguardian.com/us/film
/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42b2ea1.html
Latest Film news, comment and analysis from the Guardian, the world's leading liberal voice

⊕ 2 **Felicity Huffman among dozens charged over admissions fraud at top US schools | US news | The Guardian**
http://www.theguardian.com/us-news/2019/mar/12/us-college-admissions-fraud-scheme-charges-georgetown-southern-california-universities
/home/koustav/shared/guardiannews/2fbc519f-cc92-4a6f-9a3e-edd7272d61f9.html
Scheme helped wealthy Americans buy their children's way into elite universities including Yale, Georgetown and Stanford

⊕ 3 **What is life really like in border country, where Trump wants his wall? | US news | The Guardian**
http://www.theguardian.com/us-news/2019/jan/15/us-mexico-border-trump-wall-dispatch
/home/koustav/shared/guardiannews/1504a582-5edf-4b17-bc6c-1688190444af.html
The Guardian travelled to five locations on the US-Mexico border to discover how Trump's rhetoric jars with the reality on the ground

⊕ 4 **Trump ordered officials to give Kushner security clearance despite concerns, report says - live | US news | The Guardian**
http://www.theguardian.com/us-news/live/2019/feb/28/cohen-testimony-closed-doors-congress-latest-trump-kim-summit-us-politics
/home/koustav/shared/guardiannews/a3087987-803a-40c3-8035-ac91ad18b481.html
Despite pushback from intelligence officials and top administration members, the president ordered top-secret clearance for his son-in-law

⊕ 5 **Trump agrees to temporarily reopen government after America's longest shutdown – as it happened | US news | The Guardian**
http://www.theguardian.com/us-news/live/2019/jan/25/roger-stone-trump-russia-mueller-investigation-latest-live
/home/koustav/shared/guardiannews/d1c0375f-56cb-4b7a-a2c6-c0930fc5513f.html
Move comes as president faces intense scrutiny after the arrest of key ally Roger Stone in Mueller's Russia inquiry

⊕ 6 **Trump agrees to temporarily reopen government after America's longest shutdown – as it happened | US news | The Guardian**
http://www.theguardian.com/us-news/live/2019/jan/25/roger-stone-trump-russia-mueller-investigation-latest-live
/home/koustav/shared/guardiannews/413a6bbc-7cc1-4c51-89f1-96089cd5a900.html
Move comes as president faces intense scrutiny after the arrest of key ally Roger Stone in Mueller's Russia inquiry

⊕ 7 **Roger Stone pleads not guilty to lying to Congress and witness tampering | US news | The Guardian**
http://www.theguardian.com/us-news/2019/jan/29/roger-stone-arraignment-trump-russia-mueller-investigation
/home/koustav/shared/guardiannews/4c4aaacb-10c3-4b33-a270-cf63d619a7ab.html
Stone pleaded not guilty to charges in a seven-count indictment filed against him when he was arrested in Florida last Friday

⊕ 8 **Florida shooting: suspect escaped scene by hiding among students as they fled | US news | The Guardian**
http://www.theguardian.com/us-news/2018/feb/15/florida-shooting-suspect-charged-questions-nikolas-cruz
/home/koustav/shared/guardiannews/78c98502-1d07-4422-975b-ada93193b3e7.html
Nikolas Cruz, 19, charged with 17 counts of murder as officials confirm the AR-15 rifle used to commit massacre was purchased legally

⊕ 9 **US immigration | Page 3 of 122 | Us-news | The Guardian**
http://www.theguardian.com/us-news/usimmigration
/home/koustav/shared/guardiannews/c426854b-32e7-407a-8903-7573d8487250.html
N/A

⊕ 10 **Roger Stone ordered to appear in court over Instagram post targeting judge | US news | The Guardian**
http://www.theguardian.com/us-news/2019/feb/19/roger-stone-court-instagram-post-judge
/home/koustav/shared/guardiannews/fe60abf5-a5a1-4f76-b70d-a57c4dc63809.html
Stone has formally apologized for the post, which showed the judge with what appeared to be crosshairs near her head

4. Final Comparison Result screen is displayed when both the buttons (Lucene or Page Rank) are clicked in any order or when enter key is pressed after entering the search term. Please note that the *URL* and *Title* are both *hyperlinks* and clickable. The color and decoration style have been changed to make them look consistent and for demarcating between the default "Lucene" and "Page Rank" based ordering.

Overlaps: **2**   🔍 Bond

<div align="center">Lucene | Page Rank | Clear | Results</div>

Term: **Bond**	*Result:* **1 - 10 of 316**	*Term:* **Bond**	*Result:* **1 - 10 of 316**
LUCENE BASED SEARCH RESULTS		**PAGE RANK BASED SEARCH RESULTS**	

LUCENE BASED SEARCH RESULTS

➕ 1 **TV made America's bail system famous. Now reformers want to end it | US news | The Guardian**
http://www.theguardian.com/us-news/2017/aug/30/new-jersey-bail-reform-criminal-justice-bond-money
/home/koustav/shared/guardiannews/58780614-108e-4f09-bd34-f2ce690e456a.html
Bail money has trapped poor Americans for decades. But bondsmen are hitting back at attempts to reform the system

➕ 2 **Judge orders Florida shooting suspect to be held in jail without bond – video | US news | The Guardian**
http://www.theguardian.com/us-news/video/2018/feb/15/judge-orders-florida-shooting-suspect-to-be-held-in-jail-without-bond-video
/home/koustav/shared/guardiannews/283e90ef-c59b-4406-bd3a-613c2ce88993.html
Nikolas Cruz, 19, has appeared in court charged with the murder of 17 people and will be held in jail without bond, a Broward County judge has ordered

➕ 3 **Pennsylvania recount: Jill Stein to take fight to federal court after $1m bond set | US news | The Guardian**
http://www.theguardian.com/us-news/2016/dec/03/jill-stein-pennsylvania-recount-green-party
/home/koustav/shared/guardiannews/0cd64206-6eef-431f-b9a1-716bd9dd5bc6.html
Green party candidate vows to counter judge's 'outrageous demand' for $1m bond by seeking emergency relief in federal court

➕ 4 **Movies | The Guardian**
http://www.theguardian.com/us/film
/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42b2ea1.html
Latest Film news, comment and analysis from the Guardian, the world's leading liberal voice

➕ 5 **Sean Spicer | Page 3 of 11 | Us-news | The Guardian**
http://www.theguardian.com/us-news/sean-spicer
/home/koustav/shared/guardiannews/0d167158-16dc-4401-9e93-849859fac0d3.html
N/A

➕ 6 **US immigration | Us-news | The Guardian**
http://www.theguardian.com/us-news/usimmigration
/home/koustav/shared/guardiannews/ff85e14e-e0f4-4a9b-8f51-5cd3c3caea24.html
N/A

➕ 7 **John F Kennedy | Page 4 of 12 | Us-news | The Guardian**
http://www.theguardian.com/us-news/john-f-kennedy
/home/koustav/shared/guardiannews/edb299f6-e3e0-4223-8b7e-7fda7c49069f.html
N/A

➕ 8 **Walter Scott shooting | Us-news | The Guardian**
http://www.theguardian.com/us-news/walter-scott-shooting
/home/koustav/shared/guardiannews/b653978a-6491-4671-b98e-7a1720e6cfa6.html
N/A

➕ 9 **Paul Manafort in court: a not guilty plea, $10m bond – and no sign of repentance | US news | The Guardian**
http://www.theguardian.com/us-news/2017/oct/30/paul-manafort-court-not-guilty-plea-charges
/home/koustav/shared/guardiannews/b88c41ef-a380-45c2-b759-44e76fec2051.html
On Monday afternoon, Trump's ex-campaign chair walked into a Washington courtroom – and the ensuing hour made it clear how steeply he has fallen

➕ 10 **Chicago's 'Skullcap Crew': band of police accused of brutality evade discipline | US news | The Guardian**
http://www.theguardian.com/us-news/2016/aug/03/chicago-skullcap-crew-police-brutality
/home/koustav/shared/guardiannews/7ca30cb9-6730-4ec1-aa39-81eaf956260f.html
Dogged by allegations of abuse, members of the group have been named in more than 20 federal lawsuits – yet have won repeated praise from department

PAGE RANK BASED SEARCH RESULTS

➕ 1 **Movies | The Guardian**
http://www.theguardian.com/us/film
/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42b2ea1.html
Latest Film news, comment and analysis from the Guardian, the world's leading liberal voice

➕ 2 **Felicity Huffman among dozens charged over admissions fraud at top US schools | US news | The Guardian**
http://www.theguardian.com/us-news/2019/mar/12/us-college-admissions-fraud-scheme-charges-georgetown-southern-california-universities
/home/koustav/shared/guardiannews/2fbc519f-cc92-4a6f-9a3e-edd7272d61f9.html
Scheme helped wealthy Americans buy their children's way into elite universities including Yale, Georgetown and Stanford

➕ 3 **What is life really like in border country, where Trump wants his wall? | US news | The Guardian**
http://www.theguardian.com/us-news/2019/jan/15/us-mexico-border-trump-wall-dispatch
/home/koustav/shared/guardiannews/1504a582-5edf-4b17-bc6c-1688190444af.html
The Guardian travelled to five locations on the US-Mexico border to discover how Trump's rhetoric jars with the reality on the ground

➕ 4 **Trump ordered officials to give Kushner security clearance despite concerns, report says - live | US news | The Guardian**
http://www.theguardian.com/us-news/live/2019/feb/28/cohen-testimony-closed-doors-congress-latest-trump-kim-summit-us-politics
/home/koustav/shared/guardiannews/a3087987-803a-40c3-8035-ac91ad18b481.html
Despite pushback from intelligence officials and top administration members, the president ordered top-secret clearance for his son-in-law

➕ 5 **Trump agrees to temporarily reopen government after America's longest shutdown – as it happened | US news | The Guardian**
http://www.theguardian.com/us-news/live/2019/jan/25/roger-stone-trump-russia-mueller-investigation-latest-live
/home/koustav/shared/guardiannews/d1c0375f-56cb-4b7a-a2c6-c0930fc5513f.html
Move comes as president faces intense scrutiny after the arrest of key ally Roger Stone in Mueller's Russia inquiry

➕ 6 **Trump agrees to temporarily reopen government after America's longest shutdown – as it happened | US news | The Guardian**
http://www.theguardian.com/us-news/live/2019/jan/25/roger-stone-trump-russia-mueller-investigation-latest-live
/home/koustav/shared/guardiannews/413a6bbc-7cc1-4c51-89f1-96089cd5a900.html
Move comes as president faces intense scrutiny after the arrest of key ally Roger Stone in Mueller's Russia inquiry

➕ 7 **Roger Stone pleads not guilty to lying to Congress and witness tampering | US news | The Guardian**
http://www.theguardian.com/us-news/2019/jan/29/roger-stone-arraignment-trump-russia-mueller-investigation
/home/koustav/shared/guardiannews/4c4aaacb-10c3-4b33-a270-cf63d619a7ab.html
Stone pleaded not guilty to charges in a seven-count indictment filed against him when he was arrested in Florida last Friday

➕ 8 **Florida shooting: suspect escaped scene by hiding among students as they fled | US news | The Guardian**
http://www.theguardian.com/us-news/2018/feb/15/florida-shooting-suspect-charged-questions-nikolas-cruz
/home/koustav/shared/guardiannews/78c98502-1d07-4422-975b-ada93193b3e7.html
Nikolas Cruz, 19, charged with 17 counts of murder as officials confirm the AR-15 rifle used to commit massacre was purchased legally

➕ 9 **US immigration | Page 3 of 122 | Us-news | The Guardian**
http://www.theguardian.com/us-news/usimmigration
/home/koustav/shared/guardiannews/c426854b-32e7-407a-8903-7573d8487250.html
N/A

➕ 10 **Roger Stone ordered to appear in court over Instagram post targeting judge | US news | The Guardian**
http://www.theguardian.com/us-news/2019/feb/19/roger-stone-court-instagram-post-judge
/home/koustav/shared/guardiannews/fe60abf5-a5a1-4f76-b70d-a57c4dc63809.html
Stone has formally apologized for the post, which showed the judge with what appeared to be crosshairs near her head

5. On clicking any of the links on the search results

6. Result metadata can be obtained by clicking the "+" button on the left side of any of the results

Overlaps: **2**

🔍 Bond

Lucene | Page Rank | Clear | Results

Term: **Bond**   *Result:* **1 - 10 of 316**   *Term:* **Bond**   *Result:* **1 - 10 of 316**

LUCENE BASED SEARCH RESULTS

⊕ 1 **TV made America's bail system famous. Now reformers want to end it | US news | The Guardian**
http://www.theguardian.com/us-news/2017/aug/30/new-jersey-bail-reform-criminal-justice-bond-money
/home/koustav/shared/guardiannews/58780614-108e-4f09-bd34-f2ce690e456a.html
Bail money has trapped poor Americans for decades. But bondsmen are hitting back at attempts to re
form the system

⊖ 2 **Judge orders Florida shooting suspect to be held in jail without bond – video | US news | The G
uardian**
*http://www.theguardian.com/us-news/video/2018/feb/15/judge-orders-florida-shooting-suspect-to-be-held-in-jail-wit
hout-bond-video*
/home/koustav/shared/guardiannews/283e90ef-c59b-4406-bd3a-613c2ce88993.html
Nikolas Cruz, 19, has appeared in court charged with the murder of 17 people and will be held in jail
without bond, a Broward County judge has ordered

id	/home/koustav/shared/guardiannews/283e90ef-c59b-4406-bd3a-613c2c e88993.html
og_image	https://i.guim.co.uk/img/media/e135ae117c38020ea5116ac7e2e4322a2a4 2ed76/0_62_3143_1885/master/3143.jpg?width=1200&height=630&quali ty=85&auto=format&fit=crop&overlay-align=bottom%2Cleft&overlay-wi dth=100p&overlay-base64=L2ltZy9zdGF0aWMvb3ZlcmxheXMvdGctZGV mYXVsdC5wbmc&s=ec4039b613c5bc8da7e9e21ab4b463f2
og_image_width	1200
article_published_time	2018-02-15T21:35:38Z
twitter_card	summary_large_image
stream_content_type	text/html
og_site_name	the Guardian
keywords	US news,Parkland, Florida school shooting
apple_mobile_web_app_title	Guardian
twitter_app_url_ipad	gnmguardian://us-news/video/2018/feb/15/judge-orders-florida-shootin g-suspect-to-be-held-in-jail-without-bond-video?contenttype=Article&s ource=twitter
og_description	Nikolas Cruz, 19, has appeared in court charged with the murder of 17 pe ople and will be held in jail without bond, a Broward County judge has or dered
article_modified_time	2018-02-16T09:24:46Z
dc_title	Judge orders Florida shooting suspect to be held in jail without bond – vi deo	US news	The Guardian
content_encoding	UTF-8
application_name	The Guardian
thumbnail	https://i.guim.co.uk/img/media/e135ae117c38020ea5116ac7e2e4322a2a4 2ed76/0_62_3143_1885/master/3143.jpg?width=620&quality=85&auto=f ormat&fit=max&s=8f8623cfb07604583d8e4b788166bbf7
al_ios_app_name	The Guardian
twitter_app_id_ipad	409128287
resourcename	/home/koustav/shared/guardiannews/283e90ef-c59b-4406-bd3a-613c2c e88993.html
fb_pages	10513336322
article_tag	US news,Parkland, Florida school shooting
twitter_app_name_ipad	The Guardian
content_language	en
twitter_app_name_iphone	The Guardian
format_detection	telephone=no
msapplication_tileimage	https://assets.guim.co.uk/images/favicons/023dafadbf5ef53e0865e4baaaa 32b3b/windows_tile_144_b.png
apple_itunes_app	app-id=409128287, app-argument=https://www.theguardian.com/us-ne ws/video/2018/feb/15/judge-orders-florida-shooting-suspect-to-be-held -in-jail-without-bond-video, affiliate-data=ct=newsmartappbanner&pt=3 04191
twitter_app_id_googleplay	com.guardian
twitter_app_url_iphone	gnmguardian://us-news/video/2018/feb/15/judge-orders-florida-shootin g-suspect-to-be-held-in-jail-without-bond-video?contenttype=Article&s ource=twitter
theme_color	#052962
handheldfriendly	true
description	Nikolas Cruz, 19, has appeared in court charged with the murder of 17 pe ople and will be held in jail without bond, a Broward County judge has or dered
twitter_app_name_googleplay	The Guardian
title	Judge orders Florida shooting suspect to be held in jail without bond – vi

PAGE RANK BASED SEARCH RESULTS

⊖ 1 **Movies | The Guardian**
http://www.theguardian.com/us/film
/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42b2ea1.html
Latest Film news, comment and analysis from the Guardian, the world's leading liberal voice

id	/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42 b2ea1.html
apple_itunes_app	app-id=409128287, app-argument=https://www.theguardian.com/us/fil m, affiliate-data=ct=newsmartappbanner&pt=304191
og_image	https://assets.guim.co.uk/images/eada8aa27c12fe2d5afa3a89d3fbae0d/fa llback-logo.png
twitter_app_id_googleplay	com.guardian
twitter_app_url_iphone	gnmguardian://us/film?contenttype=front&source=twitter
theme_color	#052962
twitter_card	summary
stream_content_type	text/html
og_site_name	the Guardian
twitter_url	https://www.theguardian.com/us/film
handheldfriendly	true
apple_mobile_web_app_title	Guardian
twitter_app_url_ipad	gnmguardian://us/film?contenttype=front&source=twitter
description	Latest Film news, comment and analysis from the Guardian, the world's le ading liberal voice
twitter_app_name_googleplay	The Guardian
title	Movies	The Guardian
al_ios_url	gnmguardian://us/film?contenttype=front&source=applinks
twitter_app_id_iphone	409128287
og_description	Latest Film news, comment and analysis from the Guardian, the world's le ading liberal voice
al_ios_app_store_id	409128287
twitter_dnt	on
twitter_site	@guardian
dc_title	Movies	The Guardian
content_encoding	UTF-8
content_type	text/html; charset=UTF-8
stream_size	978403
application_name	The Guardian
x_parsed_by	org.apache.tika.parser.DefaultParser
og_type	website
al_ios_app_name	The Guardian
og_title	Movies	The Guardian
twitter_app_id_ipad	409128287
resourcename	/home/koustav/shared/guardiannews/6b7af0c2-07bc-4245-ae8f-e628f42 b2ea1.html
fb_pages	10513336322
msapplication_tilecolor	#052962
x_ua_compatible	IE=Edge
fb_app_id	180444840287
twitter_app_name_ipad	The Guardian
viewport	width=device-width,minimum-scale=1,initial-scale=1
og_url	http://www.theguardian.com/us/film
content_language	en
twitter_app_name_iphone	The Guardian
format_detection	telephone=no
msapplication_tileimage	https://assets.guim.co.uk/images/favicons/023dafadbf5ef53e0865e4baaaa 32b3b/windows_tile_144_b.png
version	1629625511559823400

7. Screen to view just URLs of the returned results: The "Results / URLs" toggle button just beside the "Clear" button can be used to switch the view between the Results screen and the URLs screen

Overlaps: **2**

Q Bond

Lucene	Page Rank	Clear	URLs

Term: **Bond**   *Result:* **1 - 10 of 316**   *Term:* **Bond**   *Result:* **1 - 10 of 316**

#	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/2017/aug/30/new-jersey-bail-reform-criminal-justice-bond-money	http://www.theguardian.com/us/film
2	http://www.theguardian.com/us-news/video/2018/feb/15/judge-orders-florida-shooting-suspect-to-be-held-in-jail-without-bond-video	http://www.theguardian.com/us-news/2019/mar/12/us-college-admissions-fraud-scheme-charges-georgetown-southern-california-universities
3	http://www.theguardian.com/us-news/2016/dec/03/jill-stein-pennsylvania-recount-green-party	http://www.theguardian.com/us-news/2019/jan/15/us-mexico-border-trump-wall-dispatch
4	http://www.theguardian.com/us/film	http://www.theguardian.com/us-news/live/2019/feb/28/cohen-testimony-closed-doors-congress-latest-trump-kim-summit-us-politics
5	http://www.theguardian.com/us-news/sean-spicer	http://www.theguardian.com/us-news/live/2019/jan/25/roger-stone-trump-russia-mueller-investigation-latest-live
6	http://www.theguardian.com/us-news/usimmigration	http://www.theguardian.com/us-news/live/2019/jan/25/roger-stone-trump-russia-mueller-investigation-latest-live
7	http://www.theguardian.com/us-news/john-f-kennedy	http://www.theguardian.com/us-news/2019/jan/29/roger-stone-arraignment-trump-russia-mueller-investigation
8	http://www.theguardian.com/us-news/walter-scott-shooting	http://www.theguardian.com/us-news/2018/feb/15/florida-shooting-suspect-charged-questions-nikolas-cruz
9	http://www.theguardian.com/us-news/2017/oct/30/paul-manafort-court-not-guilty-plea-charges	http://www.theguardian.com/us-news/usimmigration
10	http://www.theguardian.com/us-news/2016/aug/03/chicago-skullcap-crew-police-brutality	http://www.theguardian.com/us-news/2019/feb/19/roger-stone-court-instagram-post-judge

Results

1. Search Term: Venezuela

Term: **Venezuela**   *Result:* **1 - 10 of 431**   *Term:* **Venezuela**   *Result:* **1 - 10 of 431**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/us-politics
2	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/donaldtrump
3	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/2019/mar/14/beto-orourke-announces-2020-run-for-presidency
4	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/us-elections-2016
5	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/video/2019/feb/13/ilhan-omar-elliot-abrams-venezuela-video
6	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/florida
7	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/donaldtrump
8	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/us-foreign-policy

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
9	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/mike-pence
10	http://www.theguardian.com/us-news/live/2019/jan/30/trump-politics-venezuela-congress-shutdown-latest-news-updates	http://www.theguardian.com/us-news/live/2019/mar/08/donald-trump-latest-news-live-alabama-paul-manafort-russia-hr1-bill-vote

2. Search Term: Senate

Term: **Senate** Result: **1 - 10 of 5746** Term: **Senate** Result: **1 - 10 of 5746**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news
2	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/us-politics
3	http://www.theguardian.com/us-news/live/2014/dec/09/cia-torture-report-released-senate	http://www.theguardian.com/us
4	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/donaldtrump
5	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/beto-o-rourke
6	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/ng-interactive/2019/jan/17/2020-election-candidates-list-democrats-primaries-trump-bernie-sanders-joe-biden-elizabeth-warren-kamala-harris-beto-o-rourke-running
7	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/2019/mar/14/beto-orourke-announces-2020-run-for-presidency
8	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/live/2019/mar/14/trump-news-today-latest-live-updates-beto-o-rourke-boeing-flights-us-politics-ireland
9	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/2019/mar/14/senate-national-emergency-declaration-vote-trump
10	http://www.theguardian.com/us-news/live/2018/sep/28/kavanaugh-senate-committee-vote-live	http://www.theguardian.com/us-news/video/2019/mar/14/who-is-2020-presidential-candidate-beto-orourke-video-profile

3. Search Term: Democrats

Term: **Democrats** Result: **1 - 10 of 7214** Term: **Democrats** Result: **1 - 10 of 7214**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news
2	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/us-politics
3	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us
4	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us/commentisfree
5	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/donaldtrump
6	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/democrats
7	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/beto-o-rourke
8	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/ng-interactive/2019/jan/17/2020-election-candidates-list-democrats-primaries-trump-bernie-sanders-joe-biden-elizabeth-warren-kamala-harris-beto-o-rourke-running

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
9	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/2019/mar/14/beto-orourke-announces-2020-run-for-presidency
10	http://www.theguardian.com/us-news/live/2018/nov/06/us-midterms-elections-2018-latest-live-polls-news-updates-donald-trump-republicans-democrats	http://www.theguardian.com/us-news/live/2019/mar/14/trump-news-today-latest-live-updates-beto-o-rourke-boeing-flights-us-politics-ireland

4. Search Term: Republicans

	Term: **Republicans**	Result: **1 - 10 of 7449**	Term: **Republicans**	Result: **1 - 10 of 7449**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news
2	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/us-politics
3	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us/commentisfree
4	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/donaldtrump
5	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/democrats
6	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/beto-o-rourke
7	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/ng-interactive/2019/jan/17/2020-election-candidates-list-democrats-primaries-trump-bernie-sanders-joe-biden-elizabeth-warren-kamala-harris-beto-o-rourke-running
8	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/2019/mar/14/beto-orourke-announces-2020-run-for-presidency
9	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/live/2019/mar/14/trump-news-today-latest-live-updates-beto-o-rourke-boeing-flights-us-politics-ireland
10	http://www.theguardian.com/us-news/live/2018/nov/07/midterm-elections-democrats-trump-republicans-house-latest-live	http://www.theguardian.com/us-news/2019/mar/14/senate-national-emergency-declaration-vote-trump

5. Search Term: Patriot Movement

	Term: **Patriot Movement**	Result: **1 - 10 of 2128**	Term: **Patriot Movement**	Result: **1 - 10 of 2128**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/2016/nov/07/oregon-county-sheriff-glenn-palmer-todd-mckinley-reelection	http://www.theguardian.com/us/environment
2	http://www.theguardian.com/us-news/oregon-militia-standoff	http://www.theguardian.com/us
3	http://www.theguardian.com/us-news/2018/nov/18/arrests-and-clashes-as-rival-right-and-left-rallies-descend-on-portland-again	http://www.theguardian.com/us/lifeandstyle
4	http://www.theguardian.com/us-news/2018/aug/03/portland-patriot-prayer-rally-fears-violence-charlottesville	http://www.theguardian.com/us/commentisfree
5	http://www.theguardian.com/us-news/portland	http://werw.theguardian.com/us-news/ng-interactive/2019/jan/17/2020-election-candidates-list-democrats-primaries-trump-bernie-sanders-joe-biden-elizabeth-warren-kamala-harris-beto-o-rourke-running
6	http://www.theguardian.com/us-news/2018/aug/19/rightwing-protest-seattle-march-against-alleged-leftwing-violence	http://www.theguardian.com/us-news/2019/mar/14/beto-orourke-announces-2020-run-for-presidency
7	http://www.theguardian.com/us-news/oregon	http://www.theguardian.com/us-news/2019/mar/14/trump-brexit-theresa-may-uk-criticism-talks-latest-news
8	http://www.theguardian.com/us-news/2018/aug/04/patriot-prayer-to-carry-guns-at-portland-rally-as-fears-of-violence-rise	http://www.theguardian.com/us-news/2019/mar/13/rich-kids-top-college-admissions
9	http://www.theguardian.com/us-news/2018/aug/06/portland-protest-rightwing-patriot-prayer-police	http://www.theguardian.com/us-news/2019/mar/12/monumental-step-backwards-the-1bn-gas-pipeline-project-dividing-new-york
10	http://www.theguardian.com/us-news/2015/may/21/fbi-patriot-act-doj-report	http://www.theguardian.com/us-news/2019/mar/12/climate-change-student-activists-environment

6. Search Term: Oscar 2019

Term: **Oscar 2019**	Result: **1 - 10 of 15399**	Term: **Oscar 2019**	Result: **1 - 10 of 15399**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/2017/jan/17/barack-obama-commutes-sentence-oscar-lopez-rivera-puerto-rico-activist	http://www.theguardian.com/us-news
2	http://www.theguardian.com/us-news/2016/oct/01/richard-trentlage-oscar-mayer-wiener-song-dies-hot-dogs	http://www.theguardian.com/us-news/us-politics
3	http://www.theguardian.com/us/film	http://www.theguardian.com/us/environment
4	http://www.theguardian.com/us-news/dick-cheney	http://www.theguardian.com/us
5	http://www.theguardian.com/us-news/2018/oct/14/zero-tolerance-family-separation-indefinite-detention-karnes-texas	http://www.theguardian.com/us/lifeandstyle
6	http://www.theguardian.com/us-news/2018/oct/15/tolerancia-cero-familias-separadas-karnes-texas-detencion-infinita	http://www.theguardian.com/us/business
7	http://www.theguardian.com/us-news/california	http://www.theguardian.com/us/sport
8	http://www.theguardian.com/us-news/nate-silver	http://www.theguardian.com/money
9	http://www.theguardian.com/us-news/black-lives-matter-movement	http://www.theguardian.com/us/travel
10	http://www.theguardian.com/us-news/2019/feb/25/us-briefing-oscars-shocks-pro-voter-bill-and-venezuela-meeting	http://www.theguardian.com/us/technology

7. Search Term: Channel

Term: **Channel**	Result: **1 - 10 of 477**	Term: **Channel**	Result: **1 - 10 of 477**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/2018/aug/17/stormy-daniels-blames-celebrity-big-brother-withdrawal-on-custody-issue	http://www.theguardian.com/us/tv-and-radio
2	http://www.theguardian.com/us-news/2017/may/28/trump-white-house-jared-kushner-russia	http://www.theguardian.com/us/commentisfree
3	http://www.theguardian.com/us-news/2017/mar/05/roger-stone-trump-adviser-julian-assange	http://www.theguardian.com/us-news/ng-interactive/2018/dec/05/robert-mueller-who-charged-list-trump-russia-inquiry-latest-indictments-flynn-manafort-cohen-what-happens-next
4	http://www.theguardian.com/us-news/2017/apr/07/us-russia-relations-syria-military-strikes-putin-trump	http://www.theguardian.com/us-news/2019/feb/01/cory-booker-announces-run-for-presidency-in-2020
5	http://www.theguardian.com/us-news/2017/may/29/vladimir-putin-is-bigger-threat-than-isis-john-mccain-says	http://www.theguardian.com/us-news/live/2019/mar/01/trump-kushner-news-latest-live-updates-us-politics-today-democrats-republicans
6	http://www.theguardian.com/us-news/ronpaul	http://www.theguardian.com/us-news/2019/mar/04/fox-news-stormy-daniels-rupert-murdoch-trump-win
7	http://www.theguardian.com/us-news/california	http://www.theguardian.com/us-news/series/the-campaign-minute-2016
8	http://www.theguardian.com/us-news/los-angeles	http://www.theguardian.com/us-news/2019/feb/12/parkland-shooting-survivor-gun-control-activism
9	http://www.theguardian.com/us-news/us-weather	http://www.theguardian.com/us-news/live/2019/jan/09/rod-rosenstein-donald-trump-latest-news-live
10	http://www.theguardian.com/us-news/nra	http://www.theguardian.com/us-news/commentisfree/2018/mar/23/parkland-students-manifesto-americas-gun-laws

8. Search Term: Wall

Term: **Wall**	Result: **1 - 10 of 3551**	Term: **Wall**	Result: **1 - 10 of 3551**

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
1	http://www.theguardian.com/us-news/live/2018/dec/21/trump-government-shutdown-border-wall-latest-live	http://www.theguardian.com/us-news
2	http://www.theguardian.com/us-news/live/2018/dec/21/trump-government-shutdown-border-wall-latest-live	http://www.theguardian.com/us-news/us-politics
3	http://www.theguardian.com/us-news/live/2018/dec/21/trump-government-shutdown-border-wall-latest-live	http://www.theguardian.com/us/business
4	http://www.theguardian.com/us-news/live/2018/dec/21/trump-government-shutdown-border-wall-latest-live	http://www.theguardian.com/us/culture
5	http://www.theguardian.com/us-news/live/2019/jan/10/trump-us-mexico-border-wall-shutdown-live-updates	http://www.theguardian.com/us-news/donaldtrump

	LUCENE BASED RESULTS URL	PAGE RANK BASED RESULTS URL
6	http://www.theguardian.com/us-news/live/2019/jan/10/trump-us-mexico-border-wall-shutdown-live-updates	http://www.theguardian.com/us-news/democrats
7	http://www.theguardian.com/us-news/live/2019/jan/10/trump-us-mexico-border-wall-shutdown-live-updates	http://www.theguardian.com/us-news/beto-o-rourke
8	http://www.theguardian.com/us-news/live/2019/jan/10/trump-us-mexico-border-wall-shutdown-live-updates	http://www.theguardian.com/us-news/ng-interactive/2019/jan/17/2020-election-candidates-list-democrats-primaries-trump-bernie-sanders-joe-biden-elizabeth-warren-kamala-harris-beto-o-rourke-running
9	http://www.theguardian.com/us-news/live/2019/jan/10/trump-us-mexico-border-wall-shutdown-live-updates	http://www.theguardian.com/us-news/2019/mar/14/beto-orourke-announces-2020-run-for-presidency
10	http://www.theguardian.com/us-news/live/2019/jan/10/trump-us-mexico-border-wall-shutdown-live-updates	http://www.theguardian.com/us-news/live/2019/mar/14/trump-news-today-latest-live-updates-beto-o-rourke-boeing-flights-us-politics-ireland

Overlap Table

No overlaps were found between the default "Lucene" based and "Page Rank" based algorithm (ordered in *descending* page rank scores) for the top 10 results returned by each of the queries. However, some overlap can be seen for queries like "Oscar 2019" if the SOLR_ROWS parameter in "run.sh" (refer Steps to Run) is increased to display more records. Please be informed that if two pages have same page rank score then the ordering between them is arbitrarily chosen by Solr and is not always consistent, it to an extent varies depending upon the version of Solr being used.

Steps to Run:

The "solr-nodejs-client" folder provided within CSCI572_HW4.zip is the "nodejs" project that communicates with "Solr" and returns the result back to the user. Nodejs and npm needs to be installed for running this application on a Windows or Linux based OS.

For ease of use a shell script by name "run.sh" has been provided for Linux based systems (script has only been tested on Ubuntu 64 bit OS) and it needs to be executed from its current location. The script installs "nodejs" and "npm", post which it downloads all the required dependencies using the "npm install" command and runs the project. The script provides the flexibility to change some of the "Solr" or "nodejs" related configurations before starting it. The configurations can be changed either in the shell script "run.sh" or "conf.json" file within the "solr-nodejs-client" folder. Please don't change any special characters in these files like quotation marks (") or slashes (/), just change the values. Changing anything else might have an impact on the application. Post configuration changes open a web browser and go to the location http://localhost:3000/, assuming the default port configuration is being used. Here is a comprehensive list of the configurations:

Property	Default Value	Description
SOLR_PROTOCOL	http	The protocol on which Solr is running for connecting to Solr
SOLR_HOST	localhost	The IP address of Solr instance
SOLR_PORT	8983	The port of Solr instance. It is just for the purposes of establishing a connection to Solr from "nodejs", not for launching a Solr instance on a different port.
SOLR_CORE_NAME	search_core	The Solr core to connect to from nodejs
SOLR_START	0	The starting offset of the result set, passed as a query parameter to Solr
SOLR_ROWS	10	The maximum number of records to fetch at a time, used for pagination and passed as a query parameter to Solr
PORT	3000	The port to launch "nodejs"
PAGE_RANK_FILE_NAME	pageRankFile	The filename that contains the page ranks for the "Page Rank" based sorting scheme
PAGE_RANK_ORDER	desc	Can either be asc or desc. The order of sorting the results either in ascending order or descending order of page rank values for the Page Rank based algotithm. For the set of queries in this assignment there are no overlaps while the results are sorted in descending order, however there are some overlaps when the results are sorted in ascending order