

ASSIGNMENT -3

REPORT by DEEPAK GARG

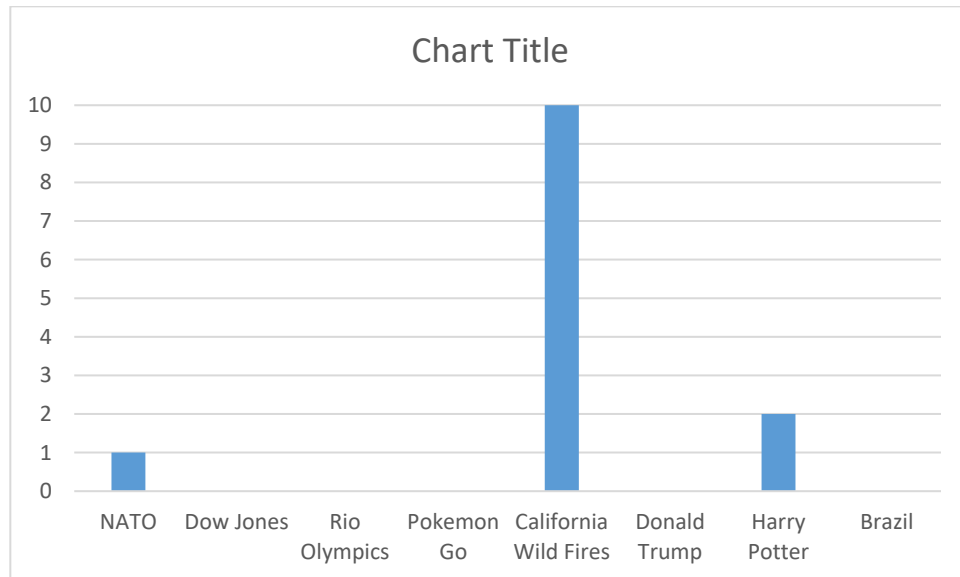
1. Given data :

- a. Used the downloaded .html web pages which were provided to us.
- b. Also used mapping files (map_nbcnews.csv file and map_wsj.csv file), provided to us. These files map the corresponding saved .html file with actual URL, which was used later to redirect to public URL rather than the local saved file as per the requirement.

2. Steps Followed

- Installed Ubuntu using Oracle VM Virtual Box , and in that I have installed the **solr – 6.2.0/**
- Created a core (named **myexample**) in solr and did the indexing using Apache Tika.
- The downloaded files are stored at /home/deepak/solr-6.2.0/crawl_data and the pages are indexed using solr
- I used NetworkX library, to compute page rank file of Page Ranking Algorithm. I created the NetworkX graph from **extract_links.txt** by drawing an edge from each downloaded file to its corresponding outgoing URLs. The page rank function then use this NetworkX graph as input and computes the page rank for each of the document using Python. I used alpha value = 0.85,max number of iterations =30 . I used default values for other parameters like error tolerance, etc.
- The output of Python program was **external_PageRank.txt**, which have key value as id of the file that is indexed and value as corresponding normalized page rank.
- After adding external_PageRank (external_PageRank.txt) file I have to modify few parameters in solr config.xml, schema.xml and managed-schema.xml as mentioned in the assignment. The core has to be reloaded for the changes to get effect.
- The external Page rank file is given as input to the solr using the steps mentioned in the assignment
-
- I used PHP programming language to write client (sample.php) and used XAMPP server to host the file. I have placed the sample.php file in htdocs folder of LAMP and I can see the result of PHP page at localhost/sample.php.

3. Overlap Graphs



Harry potter = 2 overlap

NATO = 1 overlap

California Wild Fires = 10 overlaps

Overlaps mean the results are same when using Pagerank approach and Default approach.

4. Why Some Pages Have Higher Page Rank

Page rank of each page depends on Page rank of other pages which are pointing to it. The reason why certain pages have higher Page rank can be because of three factors:

- A page is ranked higher as there are more links to it.
- Say a Page A has a higher PageRank than Page B, even though there are fewer links to Page A; the link to A comes from an important page and hence is of high value.
- Also, Well-structured website with a proper design strengthens the PageRank for some webpages.

5. *Screen Shots*

1) Default Algorithm(Lucence) – Queried for “NATO”

The screenshot shows a web browser at `localhost:3561/sample.php?q=Nato&algo=Default+Algorithm`. The search bar contains 'Nato'. Below the search bar, there are radio buttons for 'PageRank Algorithm' and 'Default Algorithm', with 'Default Algorithm' selected. A 'Submit Query' button is present. The results section shows 'Results 1 - 10 of 416:'. A list of 10 document results is displayed, each with a title, author, date created, and size. The titles are related to NATO's stance on Russia, the Ukraine crisis, and military operations.

Search:

☐ PageRank Algorithm ☒ Default Algorithm

Results 1 - 10 of 416:

- [Document](#)

Title : NATO Rejects Russian Air-Safety Proposal for Planes in Baltic Region - WSJ

Author : N/A

Date Created : N/A

Size : 33.351 KB
- [Document](#)

Title : Ukraine Crisis: U.S., Bulgaria to Begin NATO Military Drills - NBC News

Author : N/A

Date Created : N/A

Size : 117.744 KB
- [Document](#)

Title : NATO Will Set Up Commands to Soothe Nervous Eastern Europe - NBC News

Author : N/A

Date Created : N/A

Size : 116.679 KB
- [Document](#)

Title : Clinton: I'm proud to stand by our allies in NATO - NBC News

Author : N/A

Date Created : N/A

Size : 13.756 KB
- [Document](#)

Title : NATO Rejects Russian Air-Safety Proposal for Planes in Baltic Region - WSJ

Author : N/A

Date Created : N/A

Size : 126.914 KB
- [Document](#)

Title : 300 ISIS Fighters Killed in U.S.-Afghan Operation in Nangarhar: NATO - NBC News

Author : N/A

Date Created : N/A

Size : 127.616 KB
- [Document](#)

Title : NATO Summit: Obama Faces Crossroads in Fight Against ISIS - NBC News

Author : N/A

Date Created : N/A

Size : 129.418 KB
- [Document](#)

Title : Russia Launches Large-Scale Military Exercises in Disputed Territories - NBC News

Author : N/A

Date Created : N/A

Size : 127.273 KB
- [Document](#)

Title : The United States Must Be the World's Policeman - WSJ

Author : N/A

Date Created : N/A

Size : 125.512 KB
- [Document](#)

Title : Army Commander Ben Hodges: Ukraine Crisis Needs More Than Weapons - NBC News

Author : N/A

Date Created : N/A

Size : 130.748 KB

2) Page Rank Algorithm – Queried for “NATO”

The screenshot shows the same web browser at `localhost:3561/sample.php?q=Nato&algo=PageRank+Algorithm`. The search bar contains 'Nato'. Below the search bar, there are radio buttons for 'PageRank Algorithm' and 'Default Algorithm', with 'PageRank Algorithm' selected. A 'Submit Query' button is present. The results section shows 'Results 1 - 10 of 416:'. A list of 10 document results is displayed, each with a title, author, date created, and size. The titles are related to NATO's stance on Russia, the Ukraine crisis, and military operations.

Search:

☒ PageRank Algorithm ☐ Default Algorithm

Results 1 - 10 of 416:

- [Document](#)

Title : What Happens After ISIS Falls? - WSJ

Author : N/A

Date Created : N/A

Size : 266.852 KB
- [Document](#)

Title : Kerry Insists 'There is No Split' With Europe on Russia, Ukraine - NBC News

Author : N/A

Date Created : N/A

Size : 116.94 KB
- [Document](#)

Title : U.S. Sends A-10 Thunderbolt II Aircraft Back to Germany Amid Russian Threat - NBC News

Author : N/A

Date Created : N/A

Size : 133.661 KB
- [Document](#)

Title : 7 Things to Know About What's Happening in Ukraine - NBC News

Author : N/A

Date Created : N/A

Size : 152.213 KB
- [Document](#)

Title : Spain and Morocco Bust Suspected ISIS Recruiting Cell - NBC News

Author : N/A

Date Created : N/A

Size : 112.317 KB
- [Document](#)

Title : Palestinian Residents Fear ISIS Snipers, Starvation - NBC News

Author : N/A

Date Created : N/A

Size : 130.768 KB
- [Document](#)

Title : 300 ISIS Fighters Killed in U.S.-Afghan Operation in Nangarhar: NATO - NBC News

Author : N/A

Date Created : N/A

Size : 127.616 KB
- [Document](#)

Title : Trump Wants Border Wall, but Britain is Building One in France - NBC News

Author : N/A

Date Created : N/A

Size : 150.09 KB
- [Document](#)

Title : Vladimir Putin's Friend Sergei Roldugin Performs in Palmyra, Syria - NBC News

Author : N/A

Date Created : N/A

Size : 132.356 KB
- [Document](#)

Title : ISIS Trying to Topple Saudi King with Attack on Mohammed's Mosque - NBC News

Author : N/A

Date Created : N/A

Size : 135.689 KB

3) Default Algorithm(Lucence) – Queried for “Dow Jones”

localhost:3561/sample.php?q=Dow+Jones&algo=Default+Algorithm

Search: Dow Jones

☐ PageRank Algorithm ☒ Default Algorithm

Submit Query

Results 1 - 10 of 4998:

- [Document](#)
Title : Privacy Policy - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 88.587 KB
- [Document](#)
Title : Privacy Policy - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 88.661 KB
- [Document](#)
Title : Privacy Policy - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 88.682 KB
- [Document](#)
Title : Privacy Policy - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 88.651 KB
- [Document](#)
Title : Privacy Policy - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 88.672 KB
- [Document](#)
Title : Privacy Policy - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 88.682 KB
- [Document](#)
Title : Dow Jones Names Its Finance Chief - WSJ
Author : N/A
Date Created : N/A
Size : 121.971 KB
- [Document](#)
Title : Stand-Alone REIT Grouping Seen Unlikely to Whet Some Funds' Appetites - WSJ
Author : N/A
Date Created : N/A
Size : 36.477 KB
- [Document](#)
Title : Warren Buffett Could Lose an \$8-Per-Second Windfall on His Dow Chemical Stock - WSJ
Author : N/A
Date Created : N/A
Size : 41.386 KB
- [Document](#)
Title : Broad Fallout: Dow Continues Decline Down More Than 250 Points - NBC News
Author : N/A
Date Created : N/A
Size : 36.686 KB

4) Page Rank Algorithm – Queried for “Dow Jones”

localhost:3561/sample.php?q=Dow+Jones&algo=PageRank+Algorithm

Search: Dow Jones

☒ PageRank Algorithm ☐ Default Algorithm

Submit Query

Results 1 - 10 of 4998:

- [Document](#)
Title : After the Floods, an Outpouring of Help From the CEO - WSJ
Author : N/A
Date Created : N/A
Size : 132.023 KB
- [Document](#)
Title : Ahmad Khan Rahami Took Long Path to Radicalization - WSJ
Author : N/A
Date Created : N/A
Size : 40.537 KB
- [Document](#)
Title : Pressing Apple to Innovate Doesn't Play to Its Strengths - WSJ
Author : N/A
Date Created : N/A
Size : 132.715 KB
- [Document](#)
Title : CEOs See U.S. Economy Stuck in Slow-Growth Mode - WSJ
Author : N/A
Date Created : N/A
Size : 32.802 KB
- [Document](#)
Title : Valentino Designers Launch a Collection of Basics - WSJ
Author : N/A
Date Created : N/A
Size : 33.029 KB
- [Document](#)
Title : Russia Antitrust Body Opens Case Into Possible Price-Fixing Against Apple - WSJ
Author : N/A
Date Created : N/A
Size : 132.715 KB
- [Document](#)
Title : Electra Bikes Match Pantone's Colors of the Year - WSJ
Author : N/A
Date Created : N/A
Size : 124.469 KB
- [Document](#)
Title : Blackstone Beta Big on Logistics - WSJ
Author : N/A
Date Created : N/A
Size : 124.946 KB
- [Document](#)
Title : One Year, Three Conflicts: New Crash Courses in Global Business - WSJ
Author : N/A
Date Created : N/A
Size : 36.747 KB
- [Document](#)
Title : Federal Appeals Court to Hear Two Cases Challenging D.C. Gun Law - WSJ
Author : N/A
Date Created : N/A
Size : 38.184 KB

5) Default Algorithm – Queried for “RIO OLYMPICS”

The screenshot shows a web browser at the URL `localhost:3561/sample.php?q=Rio+Olympics&algo=Default+Algorithm`. The search bar contains "Rio Olympics". Below the search bar, there are two radio buttons: "PageRank Algorithm" (unselected) and "Default Algorithm" (selected). A "Submit Query" button is visible. The results section shows "Results 1 - 10 of 800:". The first 10 results are listed below:

- Document**
Title : More Than 1,000 U.S. Spies Protecting Rio Olympics - NBC News
Author : N/A
Date Created : N/A
Size : 121.391 KB
- Document**
Title : Feyissa Lilesa, Ethiopian Runner, Makes Defiant Protest Gesture at Rio Olympics - NBC News
Author : N/A
Date Created : N/A
Size : 116.923 KB
- Document**
Title : Dustin Johnson to Skip Rio Olympics Over Zika Fears - NBC News
Author : N/A
Date Created : N/A
Size : 112.319 KB
- Document**
Title : #Blackgirlmagic: African-American Women Take Spotlight at Rio Olympics - NBC News
Author : N/A
Date Created : N/A
Size : 111.799 KB
- Document**
Title : Rio Olympics: Shaunae Miller's 'Dive' That Denied Allyson Felix Gold Divides Twitter - NBC News
Author : N/A
Date Created : N/A
Size : 114.142 KB
- Document**
Title : Google Searches Reveal Your Burning Questions About the Olympics - NBC News
Author : N/A
Date Created : N/A
Size : 113.788 KB
- Document**
Title : Escaping the Olympics via Rio's Magical Urban Jungle - NBC News
Author : N/A
Date Created : N/A
Size : 114.059 KB
- Document**
Title : Zika Virus Won't Be a Big Problem at Olympics, Rio Officials Say - NBC News
Author : N/A
Date Created : N/A
Size : 132.283 KB
- Document**
Title : British Athlete Robbed at Olympics, Officials Say - NBC News
Author : N/A
Date Created : N/A
Size : 114.022 KB
- Document**
Title : Serena Williams, Zika and Team USA Top Rio Social Buzz - NBC News
Author : N/A
Date Created : N/A
Size : 105.515 KB

6) Page Rank Algorithm – Queried for “Dow Jones”

The screenshot shows a web browser at the URL `localhost:3561/sample.php?q=Rio+Olympics&algo=PageRank+Algorithm`. The search bar contains "Rio Olympics". Below the search bar, there are two radio buttons: "PageRank Algorithm" (selected) and "Default Algorithm" (unselected). A "Submit Query" button is visible. The results section shows "Results 1 - 10 of 800:". The first 10 results are listed below:

- Document**
Title : Online Drug Sales Triple After Silk Road Closure - NBC News
Author : N/A
Date Created : N/A
Size : 106.372 KB
- Document**
Title : For Most Olympians, Turning Gold Into Green is Difficult - NBC News
Author : N/A
Date Created : N/A
Size : 117.26 KB
- Document**
Title : Six Cases of Zika Virus-Linked Birth Defects in US, CDC Says - NBC News
Author : N/A
Date Created : N/A
Size : 129.545 KB
- Document**
Title : 4 Ways the Breast will Affect Americans - NBC News
Author : N/A
Date Created : N/A
Size : 304.889 KB
- Document**
Title : Trump in Praise of Dictators - NBC News
Author : N/A
Date Created : N/A
Size : 395.046 KB
- Document**
Title : Kanye West NYC Show Ends in Disappointment - NBC News
Author : N/A
Date Created : N/A
Size : 395.082 KB
- Document**
Title : Do You Know Your Miranda Rights? - NBC News
Author : N/A
Date Created : N/A
Size : 382.703 KB
- Document**
Title : WHO Warns Pregnant Women on Travel to Zika Zones - NBC News
Author : N/A
Date Created : N/A
Size : 125.974 KB
- Document**
Title : These Steps Can Turn Your Credit Around - NBC News
Author : N/A
Date Created : N/A
Size : 278.893 KB
- Document**
Title : Fighter and Thinker: the Two Sides of Muhammad Ali - NBC News
Author : N/A
Date Created : N/A
Size : 398.715 KB

7) Default Algorithm – Queried for “pokemon go”

The screenshot shows a web browser window with the address bar displaying 'localhost:3561/sample.php?q=Pokemon+Go&algo=Default+Algorithm'. The search bar contains 'Pokemon Go'. Below the search bar, there are two radio buttons: 'PageRank Algorithm' and 'Default Algorithm', with 'Default Algorithm' selected. A 'Submit Query' button is visible. The search results are listed as 'Results 1 - 10 of 819:'. The first 10 results are displayed, each with a 'Document' link, title, author, date created, and size.

Rank	Document	Title	Author	Date Created	Size
1	Document	How Pokemon Go's New 'Buddy System' Works - NBC News	N/A	N/A	34.669 KB
2	Document	How Pokemon Go's New 'Buddy System' Works - NBC News	N/A	N/A	100.188 KB
3	Document	Pokemon Go' Players Fuming After Update Resets Their Account to Level 1 - NBC News	N/A	N/A	104.213 KB
4	Document	'Pokemon Go' Eludes Cloning Attempts by Big Game Studios: Executives - NBC News	N/A	N/A	111.935 KB
5	Document	'Pokemon Go' Eludes Cloning Attempts by Big Game Studios: Executives - NBC News	N/A	N/A	39.551 KB
6	Document	Video Games - NBC News	N/A	N/A	194.256 KB
7	Document	'Pokemon Go' Hunters in South Korea Thwarted by War With North Korea - NBC News	N/A	N/A	41.849 KB
8	Document	Pokemon Go' Creators Working to Remove Problematic Pokestops - NBC News	N/A	N/A	100.073 KB
9	Document	Pokemon Go' Players Fuming After Update Resets Their Account to Level 1 - NBC News	N/A	N/A	35.703 KB
10	Document	Iran Bans 'Pokemon Go' Before Mobile Game Is Even Launched There - NBC News	N/A	N/A	36.077 KB

8) Page Rank Algorithm – Queried for “Pokemon go”

localhost:3561/sample.php?q=Pokemon+Go&algo=PageRank+Algorithm

Search: Pokemon Go

☒ PageRank Algorithm ☐ Default Algorithm

Submit Query

Results 1 - 10 of 819:

- [Document](#)
Title : Edible Marijuana Is Booming, But These Aren't Your Father's Pot Brownies - NBC News
Author : N/A
Date Created : N/A
Size : 158.123 KB
- [Document](#)
Title : Here's a Reason to Get a Puppy: Kids With Pets Have Less Anxiety - NBC News
Author : N/A
Date Created : N/A
Size : 105.83 KB
- [Document](#)
Title : Trump, Clinton and the Commander-In-Chief Forum
Author : N/A
Date Created : N/A
Size : 118.5 KB
- [Document](#)
Title : North Koreans Risk Imprisonment Calling Loved Ones Abroad: Amnesty Report - NBC News
Author : N/A
Date Created : N/A
Size : 111.568 KB
- [Document](#)
Title : Tesla Ramps Up War of Words Over Fatal Autopilot Crash - NBC News
Author : N/A
Date Created : N/A
Size : 107.402 KB
- [Document](#)
Title : Military Suicides: Most Attempts Come Before Soldiers Ever See Combat - NBC News
Author : N/A
Date Created : N/A
Size : 109.976 KB
- [Document](#)
Title : DOJ Says Apple Must Return to Court in Samsung Patent Lawsuit - NBC News
Author : N/A
Date Created : N/A
Size : 101.161 KB
- [Document](#)
Title : A Meeting More Important Than The Fed's
Author : N/A
Date Created : N/A
Size : 69.675 KB
- [Document](#)
Title : 40 Virginia Cases of Hepatitis A Now Linked to Smoothies - NBC News
Author : N/A
Date Created : N/A
Size : 99.811 KB
- [Document](#)
Title : Australia Police Raid Home of Man Said to Be Bitcoin Founder - NBC News
Author : N/A
Date Created : N/A
Size : 111.354 KB

9) Default Algorithm – Queried for “California wild fires”

localhost:3561/sample.php?q=California+Wild+Fires&algo=Default+Algorithm

Search: California Wild Fires

☐ PageRank Algorithm ☒ Default Algorithm

Submit Query

Results 1 - 10 of 10:

- [Document](#)
Title : How the Exxon Case Unraveled - WSJ
Author : N/A
Date Created : N/A
Size : 35.184 KB
- [Document](#)
Title : Low Gas Prices May Ease Economic Pinch, but Will They Last? - NBC News
Author : N/A
Date Created : N/A
Size : 127.878 KB
- [Document](#)
Title : 'Star Wars: The Force Awakens' World Premiere Wows Fans, Celebs - NBC News
Author : N/A
Date Created : N/A
Size : 135.822 KB
- [Document](#)
Title : Bringing in the Green: Seaweed Emerges as New Cash Crop in Maine - NBC News
Author : N/A
Date Created : N/A
Size : 116.855 KB
- [Document](#)
Title : How the Exxon Case Unraveled - WSJ
Author : N/A
Date Created : N/A
Size : 155.931 KB
- [Document](#)
Title : In Flood-Ravaged Louisiana, Small Businesses Struggle to Recover - WSJ
Author : N/A
Date Created : N/A
Size : 143.837 KB
- [Document](#)
Title : Science News and Analysis - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 448.142 KB
- [Document](#)
Title : The Wall Street Journal & Breaking News, Business, Financial and Economic News, World News and Video
Author : N/A
Date Created : N/A
Size : 883.355 KB
- [Document](#)
Title : The Wall Street Journal & Breaking News, Business, Financial and Economic News, World News and Video
Author : N/A
Date Created : N/A
Size : 824.096 KB
- [Document](#)
Title : The Wall Street Journal & Breaking News, Business, Financial and Economic News, World News and Video
Author : N/A
Date Created : N/A
Size : 810.089 KB

10) Page Rank Algorithm – Queried for “California wild fires”

The screenshot shows a web browser window with the address bar displaying `@localhost:3561/sample.php?q=California+Wild+Fires&algo=PageRank+Algorithm`. The search bar contains the text "California wild fires". Below the search bar, there are two radio buttons: "PageRank Algorithm" (selected) and "Default Algorithm". A "Submit Query" button is located below the radio buttons.

Results 1 - 10 of 10:

- [Document](#)
Title : How the Exxon Case Unraveled - WSJ
Author : N/A
Date Created : N/A
Size : 33.154 KB
- [Document](#)
Title : How the Exxon Case Unraveled - WSJ
Author : N/A
Date Created : N/A
Size : 155.931 KB
- [Document](#)
Title : The Wall Street Journal & Breaking News, Business, Financial and Economic News, World News and Video
Author : N/A
Date Created : N/A
Size : 824.096 KB
- [Document](#)
Title : Science News and Analysis - Wall Street Journal
Author : N/A
Date Created : N/A
Size : 448.142 KB
- [Document](#)
Title : Low Gas Prices May Ease Economic Pinch, but Will They Last? - NBC News
Author : N/A
Date Created : N/A
Size : 127.878 KB
- [Document](#)
Title : The Wall Street Journal & Breaking News, Business, Financial and Economic News, World News and Video
Author : N/A
Date Created : N/A
Size : 810.089 KB
- [Document](#)
Title : In Flood-Ravaged Louisiana, Small Businesses Struggle to Recover - WSJ
Author : N/A
Date Created : N/A
Size : 143.837 KB
- [Document](#)
Title : The Wall Street Journal & Breaking News, Business, Financial and Economic News, World News and Video
Author : N/A
Date Created : N/A
Size : 683.355 KB
- [Document](#)
Title : 'Star Wars: The Force Awakens' World Premiere Wows Fans, Celebs - NBC News
Author : N/A
Date Created : N/A
Size : 135.822 KB
- [Document](#)
Title : Bringing in the Green: Seaweed Emerges as New Cash Crop in Maine - NBC News
Author : N/A
Date Created : N/A
Size : 116.855 KB

11) Default Algorithm – Queried for “Harry potter”

localhost:3561/sample.php?q=Harry+Potter&algo=Default+Algorithm

Search: Harry Potter

☐ PageRank Algorithm ☒ Default Algorithm

Submit Query

Results 1 - 10 of 46:

- [Document](#)
Title : Scholastic's Revenue Blessed by 'Harry Potter and the Cursed Child' - WSJ
Author : N/A
Date Created : N/A
Size : 153.804 KB
- [Document](#)
Title : New 'Harry Potter' Book, Script of Upcoming Play, Gets Summer Release Date - NBC News
Author : N/A
Date Created : N/A
Size : 110.762 KB
- [Document](#)
Title : J.K. Rowling Begs Harry Potter Fans Not to Post 'Cursed Child' Spoilers - NBC News
Author : N/A
Date Created : N/A
Size : 106.891 KB
- [Document](#)
Title : Chair Used by 'Harry Potter' Author J.K. Rowling Up for Auction - NBC News
Author : N/A
Date Created : N/A
Size : 97.996 KB
- [Document](#)
Title : Finding the Spell of the Stage on the Page With 'Harry Potter and the Cursed Child' - WSJ
Author : N/A
Date Created : N/A
Size : 165.405 KB
- [Document](#)
Title : Cloaking Devices: Not Just For Harry Potter - NBC News
Author : N/A
Date Created : N/A
Size : 86.009 KB
- [Document](#)
Title : Books - NBC News
Author : N/A
Date Created : N/A
Size : 194.112 KB
- [Document](#)
Title : NBC Snags 'Harry Potter' Movies in Massive Deal
Author : N/A
Date Created : N/A
Size : 91.081 KB
- [Document](#)
Title : Harry Potter Makes a Magical Comeback
Author : N/A
Date Created : N/A
Size : 91.221 KB
- [Document](#)
Title : 'Unlucky Remarkable': Harry Potter Fans Rave About 'Cursed Child' - NBC News
Author : N/A
Date Created : N/A
Size : 298.522 KB

12) PageRank Algorithm – Queried for “Harry Potter”

localhost:3561/sample.php?q=Harry+Potter&algo=PageRank+Algorithm

Search: Harry Potter

☒ PageRank Algorithm ☐ Default Algorithm

Submit Query

Results 1 - 10 of 46:

- [Document](#)
Title : 'Slender Man' Suspect Not Competent to Stand Trial: Judge - NBC News
Author : N/A
Date Created : N/A
Size : 125.357 KB
- [Document](#)
Title : NBC Snags 'Harry Potter' Movies in Massive Deal
Author : N/A
Date Created : N/A
Size : 91.081 KB
- [Document](#)
Title : Finding the Spell of the Stage on the Page With 'Harry Potter and the Cursed Child' - WSJ
Author : N/A
Date Created : N/A
Size : 165.405 KB
- [Document](#)
Title : Meet the Press Transcript - September 20, 2015 - NBC News
Author : N/A
Date Created : N/A
Size : 224.559 KB
- [Document](#)
Title : Credit Card Contracts Literally Too Hard To Read For Most - NBC News
Author : N/A
Date Created : N/A
Size : 109.479 KB
- [Document](#)
Title : An Insider's Guide to Venice, Italy - WSJ
Author : N/A
Date Created : N/A
Size : 211.253 KB
- [Document](#)
Title : Green Party's Jill Stein Seeks to Woo Young Voters From Clinton - NBC News
Author : N/A
Date Created : N/A
Size : 140.862 KB
- [Document](#)
Title : Leisure & Arts - Wj.com
Author : N/A
Date Created : N/A
Size : 236.466 KB
- [Document](#)
Title : Russians Read Tolstoy's 'War and Peace' on TV in 4-Day Marathon - NBC News
Author : N/A
Date Created : N/A
Size : 97.39 KB
- [Document](#)
Title : Chelsea Handler's Favorite Gadgets - WSJ
Author : N/A
Date Created : N/A
Size : 157.864 KB

13) Default Algorithm – Queried for “Brazil”

The screenshot shows a web browser window with the address bar displaying `localhost:3561/sample.php?q=Brazil&algo=Default+Algorithm`. The search bar contains the text "Brazil". Below the search bar, there are two radio buttons: "PageRank Algorithm" (unselected) and "Default Algorithm" (selected). A "Submit Query" button is located below the radio buttons. The search results are displayed as a list of 10 documents, each with a title, author, date created, and size. The results are sorted by relevance.

Search: Brazil

☐ PageRank Algorithm ☒ Default Algorithm

Submit Query

Results 1 - 10 of 867:

- [Document](#)
Title : Ban in Talks to Buy Citigroup's Brazil Consumer-Banking Operations - WSJ
Author : N/A
Date Created : N/A
Size : 32,483 KB
- [Document](#)
Title : Ban in Talks to Buy Citigroup's Brazil Consumer-Banking Operations - WSJ
Author : N/A
Date Created : N/A
Size : 124,833 KB
- [Document](#)
Title : Scientists Test Vaccine Against Deadly Dengue Virus in Brazil - NBC News
Author : N/A
Date Created : N/A
Size : 108,305 KB
- [Document](#)
Title : Zika Doubled Birth Defect Rate in Brazil, Study Shows - NBC News
Author : N/A
Date Created : N/A
Size : 42,95 KB
- [Document](#)
Title : Brazil Plans Simpler Rules on Infrastructure Contracts - WSJ
Author : N/A
Date Created : N/A
Size : 124,237 KB
- [Document](#)
Title : Ex-Minister Solicited Bribes From Billionaire Elie Baibista, Brazilian Prosecutors Say - WSJ
Author : N/A
Date Created : N/A
Size : 38,122 KB
- [Document](#)
Title : Investors See Gold in Distressed Properties in Brazil - WSJ
Author : N/A
Date Created : N/A
Size : 124,82 KB
- [Document](#)
Title : Zika Has Been in Brazil Longer Than Anyone Thought: Study - NBC News
Author : N/A
Date Created : N/A
Size : 128,915 KB
- [Document](#)
Title : Judge Orders Ryan Lochte and Jimmy Feigen to Stay in Brazil During Robbery Probe - NBC News
Author : N/A
Date Created : N/A
Size : 120,896 KB
- [Document](#)
Title : Brazil Passes the Olympic Torch, Proud and Relieved - WSJ
Author : N/A
Date Created : N/A
Size : 166,368 KB

14) PageRank Algorithm – Queried for “Brazil”

The screenshot shows a web browser window with the address bar displaying `localhost:3561/sample.php?q=Brazil&algo=PageRank+Algorithm`. The search bar contains the text "Brazil". Below the search bar, there are two radio buttons: "PageRank Algorithm" (selected) and "Default Algorithm" (unselected). A "Submit Query" button is located below the radio buttons. The search results are displayed as a list of 10 documents, each with a title, author, date created, and size. The results are sorted by PageRank.

Search: Brazil

☒ PageRank Algorithm ☐ Default Algorithm

Submit Query

Results 1 - 10 of 867:

- [Document](#)
Title : How Bad is Zika for Babies? Study to Track 10,000 Pregnant Women - NBC News
Author : N/A
Date Created : N/A
Size : 45,086 KB
- [Document](#)
Title : For Most Olympians, Turning Gold Into Green is Difficult - NBC News
Author : N/A
Date Created : N/A
Size : 117,26 KB
- [Document](#)
Title : Women's Health - NBC News
Author : N/A
Date Created : N/A
Size : 192,328 KB
- [Document](#)
Title : Six Cases of Zika Virus-Linked Birth Defects in US, CDC Says - NBC News
Author : N/A
Date Created : N/A
Size : 128,545 KB
- [Document](#)
Title : 4 Ways the Breast will Affect Americans - NBC News
Author : N/A
Date Created : N/A
Size : 394,889 KB
- [Document](#)
Title : Delta's 'Innovation Class' Promises Interesting Seatmate - NBC News
Author : N/A
Date Created : N/A
Size : 99,571 KB
- [Document](#)
Title : Here's a Reason to Get a Puppy: Kids With Pets Have Less Anxiety - NBC News
Author : N/A
Date Created : N/A
Size : 105,83 KB
- [Document](#)
Title : Trump in Praise of Dictators - NBC News
Author : N/A
Date Created : N/A
Size : 395,046 KB
- [Document](#)
Title : Kanye West NYC Show Ends in Disappointment - NBC News
Author : N/A
Date Created : N/A
Size : 395,082 KB
- [Document](#)
Title : Do You Know Your Miranda Rights? - NBC News
Author : N/A
Date Created : N/A
Size : 382,703 KB

15) Default Algorithm – Queried for “Donald Trump”

The screenshot shows a web browser window with the address bar displaying `localhost:3561/sample.php?q=Donald+Trump&algo=Default+Algorithm`. The search bar contains the text "Donald Trump". Below the search bar, there are two radio buttons: "PageRank Algorithm" (unselected) and "Default Algorithm" (selected). A "Submit Query" button is located below the radio buttons. The search results are displayed as a list of 10 documents, each with a title, author, date created, and size.

Search: Donald Trump

☐ PageRank Algorithm ☒ Default Algorithm

Submit Query

Results 1 - 10 of 7251:

- Document**
Title : On the Stump: Donald Trump's Presidential Campaign - NBC News
Author : N/A
Date Created : N/A
Size : 127.94 KB
- Document**
Title : Donald Trump Expands 'Short List' of Potential Supreme Court Picks - WSJ
Author : N/A
Date Created : N/A
Size : 40 KB
- Document**
Title : Ted Cruz Endorses Donald Trump, Citing Concerns Over Hillary Clinton - WSJ
Author : N/A
Date Created : N/A
Size : 37.299 KB
- Document**
Title : Poll: Donald Trump Still Leads the GOP Field - NBC News
Author : N/A
Date Created : N/A
Size : 117.563 KB
- Document**
Title : Donald Trump Still in the Lead After Debates: New NBC News/SurveyMonkey Poll - NBC News
Author : N/A
Date Created : N/A
Size : 41.480 KB
- Document**
Title : Donald Trump to Meet With Mexican President Enrique Peña Nieto - NBC News
Author : N/A
Date Created : N/A
Size : 158.595 KB
- Document**
Title : Donald Trump Still in the Lead After Debates: New NBC News/SurveyMonkey Poll - NBC News
Author : N/A
Date Created : N/A
Size : 141.542 KB
- Document**
Title : Meet the Press Transcript - August 16, 2015 - NBC News
Author : N/A
Date Created : N/A
Size : 296.22 KB
- Document**
Title : Donald Trump Is No Laughing Matter to Some Late Night Comics - NBC News
Author : N/A
Date Created : N/A
Size : 114.208 KB
- Document**
Title : Donald Trump Criticized After He Appears to Mock Reporter Serge Kovaleski - NBC News
Author : N/A
Date Created : N/A
Size : 125.997 KB

16) PageRank Algorithm – Queried for “Donald Trump”

The screenshot shows a web browser window with the address bar displaying `localhost:3561/sample.php?q=Donald+Trump&algo=PageRank+Algorithm`. The search bar contains the text "Donald Trump". Below the search bar, there are two radio buttons: "PageRank Algorithm" (selected) and "Default Algorithm" (unselected). A "Submit Query" button is located below the radio buttons. The search results are displayed as a list of 10 documents, each with a title, author, date created, and size.

Search: Donald Trump

☒ PageRank Algorithm ☐ Default Algorithm

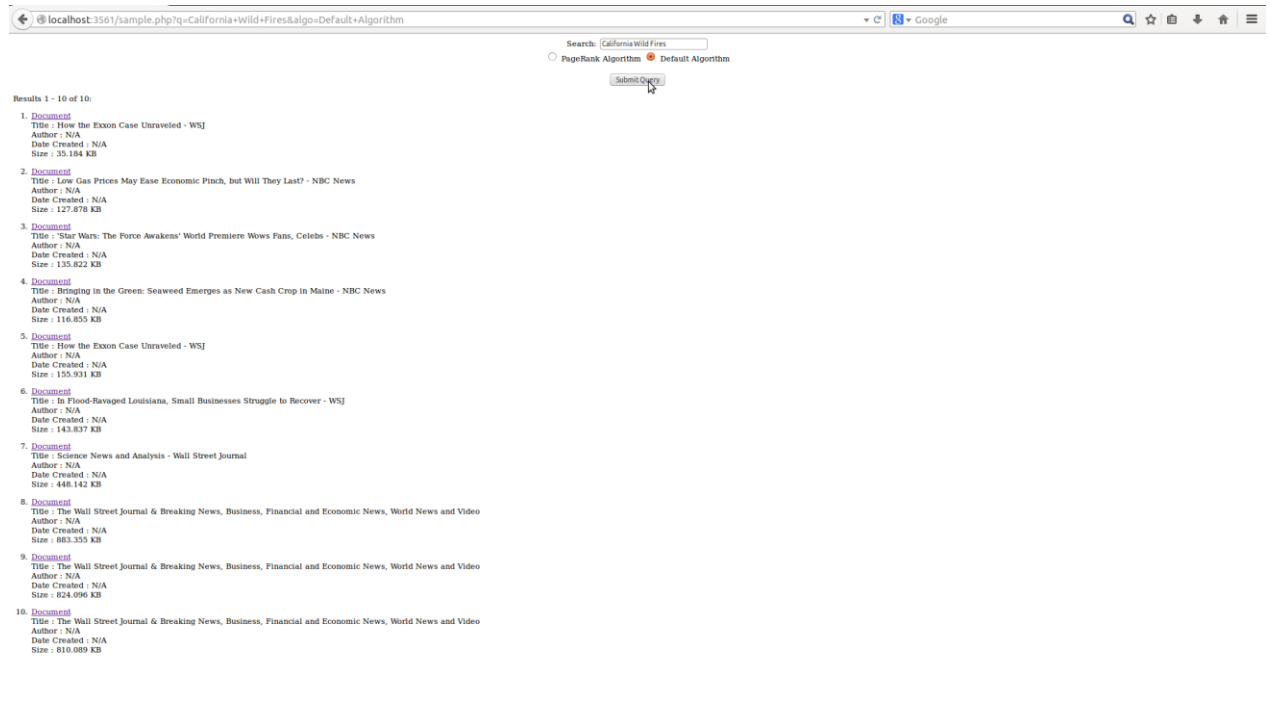
Submit Query

Results 1 - 10 of 7251:

- Document**
Title : After the Floods, an Outpouring of Help From the CEO - WSJ
Author : N/A
Date Created : N/A
Size : 132.023 KB
- Document**
Title : Pressing Apple to Innovate Doesn't Play to Its Strengths - WSJ
Author : N/A
Date Created : N/A
Size : 132.715 KB
- Document**
Title : CEOs See U.S. Economy Stuck in Slow-Growth Mode - WSJ
Author : N/A
Date Created : N/A
Size : 32.802 KB
- Document**
Title : Russia Antitrust Body Opens Case Into Possible Price-Fixing Against Apple - WSJ
Author : N/A
Date Created : N/A
Size : 132.715 KB
- Document**
Title : Syrian-American Volunteers Help New Arrivals Adapt to the U.S. - NBC News
Author : N/A
Date Created : N/A
Size : 299.39 KB
- Document**
Title : Electra Bikes Match Pantone's Colors of the Year - WSJ
Author : N/A
Date Created : N/A
Size : 124.469 KB
- Document**
Title : Russia's Putin Reaffirms Praise for Trump, Calling Him 'Bright' - NBC News
Author : N/A
Date Created : N/A
Size : 109.142 KB
- Document**
Title : Online Drug Sales Triple After Silk Road Closure - NBC News
Author : N/A
Date Created : N/A
Size : 100.372 KB
- Document**
Title : Blackstone Bets Big on Logistics - WSJ
Author : N/A
Date Created : N/A
Size : 124.946 KB
- Document**
Title : Matt Danzer and Ann Redding's Recipe for Salisbury Steak With Mushroom Gravy - WSJ
Author : N/A
Date Created : N/A
Size : 162.308 KB

FLOW FROM QUERY PAGE TO RESULT PAGE

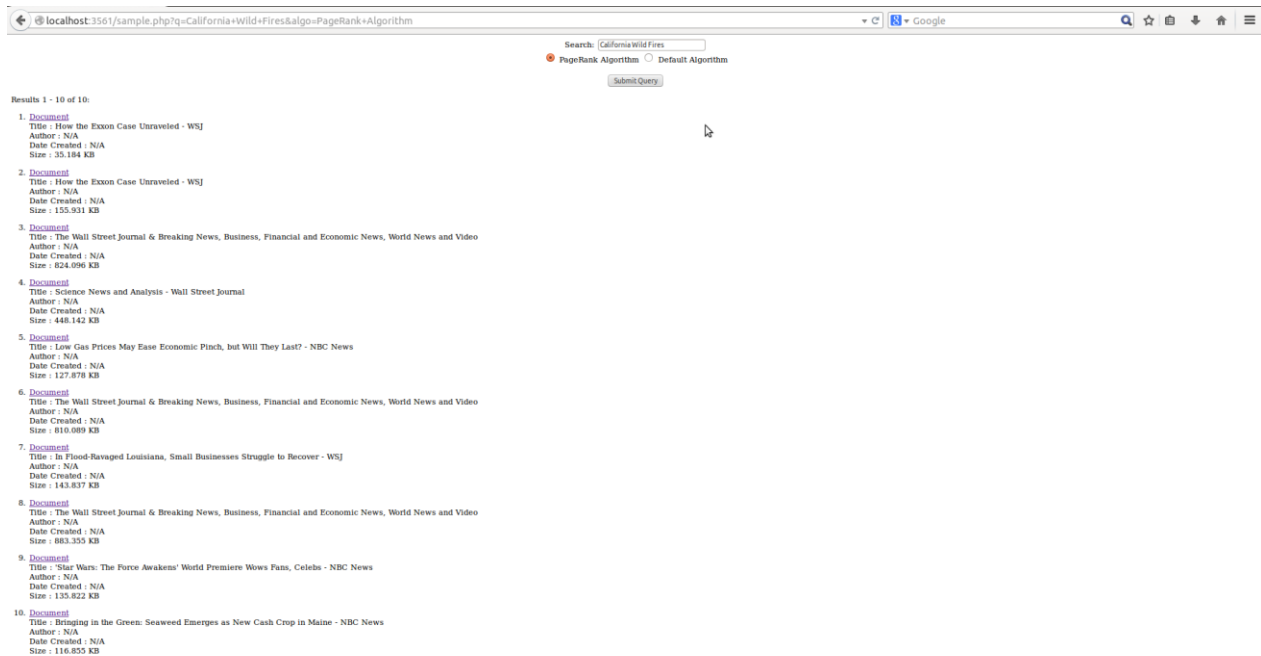
1. Default query page (California wild fire):



2. When clicked on first link, the following comes up(California wild fire) in Default: RESULT PAGE



3. PageRank query page (California wild fire)



4. When clicked on doc link, the following comes up(California wild fire) in pagerank : RESULT PAGE

The screenshot displays the homepage of The Wall Street Journal. At the top, there's a navigation bar with the site's name and various market indices like DJIA, S&P 500, Nasdaq, U.S. 10 Yr, Crude Oil, and Euro. Below this is a large banner for 'THE WALL STREET JOURNAL' with a search bar. The main content area is divided into several sections: 'What's News' featuring articles on generic-drug firms, Brexit, and the Bank of England; 'Markets' with a sidebar showing various indices and a main section for 'Google Hits Back at EU Over Antitrust Charges'; 'Opinion' with articles on Clinton and the Arab ally; and 'U.S. ELECTION 2016' with a focus on Clinton and Trump. The layout is professional and typical of a major news outlet's website.

Important Points

Clicking on URL should redirect to public URL.

Mapping files provided to us stores the mapping between the local file name to corresponding public URLs. In php file, the mapping file is used to redirect the page to actual URL rather than the local file path.

Radio Button:

I have used 2 radio button in UI, in which by default Algorithm is selected by default. I can switch to PageRank algorithm by clicking that respective radio button, accordingly the request is send to Solr in the PHP code.