

1.) OS: Mac OS X

❖ Indexing with TIKA

- bin/solr start
- bin/solr create -c myexample
- Merged the two news websites web pages to a single folder named “crawl”
- bin/post -c csci572 /Users/saigeetha/Documents/School\ Documents/Fall\ 2016/Assignments/CSCI_572_HW3/crawl
- Configured managed_schema and solr_config files for querying from the default “_text_”
- Based on warnings generated, changed the types of some of the fields to ‘strings’ in managed_schema accordingly
- Made sure all the web pages given were indexed

❖ Generation of edgelist.txt file

- Jsoup was used to create the edgelist.txt file that includes all the urls with their respective outgoing url links
- Merged both the mapping files and created the respective hash maps for generating the edgelist

❖ Generation of external_pageRankFile.txt

- The generated edgelist.txt file was given as the input to the python script
- The python script using networkx library creates digraph and calculates the pagerank values (external_PageRankFile.txt).
- The following parameters (alpha=0.85, personalization=None, max_iter=30, tol=1e-06, nstart=None, weight='weight', dangling=None) were passed to the pagerank function.
- The external_pageRankFile has all the urls with their assigned pageRank scores
- The pageRank score is calculated based on the number of incoming links

2.)

Query 1: NATO (Overlap: 0)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/ef88cc69-e232-4482-a76e-9111012ac4fc.html
2. /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/1b4ac96f-209f-4be5-b73d-a8bae7b054db.html

3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/be598af9-0a7a-450e-8cc2-4ec12bcd20ab.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a33c613e-c664-412f-a3df-90ceef275760.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/19db565e-7dbe-4d8f-8623-bfb0d983e4e5.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a4564bd4-4045-4443-9458-33abbe6e62c5.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/c23c5887-77bd-4be3-9758-b9025a495354.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b29e6cdc-d439-4556-8188-d4092cbbfd35.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/2592b561-93ec-4a6d-93b2-eba90e2ade2a.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/c8567c0b-3fde-409e-ab91-ba837aa7dcfd.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b7114274-086c-433f-9bad-950c8e85f678.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a7af769c-5ee9-4b9c-a507-4f6ed80d7d7e.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/19c7b568-d96a-418f-ace0-8f58b106fdb5.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/9fd3b1f6-3363-4c7f-8b9c-e3bff97b130a.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/aa38e496-ec53-4e46-923d-1d7e47064312.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/1c9e0e31-e5b5-4a72-8620-223517782add.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/cdfc7d92-cc6a-449f-a502-ba43f5c349a4.html

8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f1b6c951-146e-4663-b63c-2b24159940d6.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/48c5aa97-a283-4665-b095-bbc5b01a7f4c.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/23416c6e-4ecb-4d4a-b42e-ad789e1c2e30.html

Query 2: Dow Jones (Overlap: 2)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/4d0d947c-e076-47e8-9770-0085d2f36e0c.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/cc32db42-fad6-4093-bfe8-36dbde3caced.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/fed8416d-fa11-47f1-ab1a-585d78a7378d.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/444bcb14-2b62-4c06-82c6-cfb553890068.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/c1f2789c-4c38-4185-9a04-ec8c03e9dd4d.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a094029f-f4e4-4c60-b82b-d86061bb200e.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/ff59dfa4-fdba-4017-b51a-94ace9bc635c.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/8fd86283-d4f4-4145-a29d-65efc598c02b.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/1235c67c-5aad-41fc-96de-f7b338236f58.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f7979fda-f886-4c2b-b3ef-f5a8c74c0dad.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/8d259ce3-f267-4456-830d-4e0913a9edec.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/7301a5b8-6fbc-4bda-bbbd-98324149606a.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/de2e19e1-5cce-4162-96b5-402f286e8730.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f8a55e15-4708-408a-a383-aacae4d60e17.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f7979fda-f886-4c2b-b3ef-f5a8c74c0dad.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/fed8416d-fa11-47f1-ab1a-585d78a7378d.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/14070f88-d8f8-4083-8447-b0a41948b72d.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/2310aad1-0d51-4f84-9142-5516b7477701.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/adc1b054-9397-40bf-8eb7-ebc76af7e8b3.html

Query 3: Rio Olympics (Overlap: 0)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/e59dce9e-8459-4004-9b14-2ae3aea697b9.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/9ed0501b-99c3-4cb6-a790-8fb942f12cccd.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/d0d3de23-43ef-4dc1-8ade-8d0a2d5a8b89.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bd6e7f46-0c49-45a5-bdfb-06535b7fd59f.html

5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/03bb1480-7329-4fbf-8f12-9cd0e6d28e1c.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b3e25bb8-a2ed-49b2-8a0a-57208ea00bc0.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/3a28b358-242b-4605-bd38-03a2872451a4.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/5eaf7694-9685-4ade-8853-9d02427c30ef.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/69c091bc-4d55-4369-a905-bdddb0c69df1.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/d1c02513-93fb-4b20-b639-870ed2fbf1bf.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/7301a5b8-6fbc-4bda-bbbd-98324149606a.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/eb0dc35b-266e-42c9-9acc-e8ff3c644447.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/7ad991c7-8778-4bf4-964e-6cc9c7c6175b.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/4043d4fb-f571-4f1a-bf5c-db4827e9ec14.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b070a8b2-ff82-4c5f-b8cb-62e3cd64b1a1.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/18dce62a-7d0e-4ef2-845e-4e9ec31d4632.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/16d68465-5bcd-4ed2-8f6d-8d20eb5313e8.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/3231d2b6-f5c0-4ec9-8a5a-ccc1172c88f3.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/c60ffe49-e6d2-4389-9733-277ba6567992.html

10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/e7954911-017e-4e51-aeaab492c5a37b04.html

Query 4: Pokemon Go (Overlap:3)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/8dd0dc28-5358-496f-90fe-844f74b6310c.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/33a2875b-79b0-4997-83b4-717114d8f9eb.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/79870548-3a18-43c6-a2bb-31fff991c818.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/ba915459-dc48-4bef-ae1-cf561d657b70.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/dab6c82b-18a2-4342-ba42-9d9c13925994.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/9d836e82-465e-4545-a4a0-16c2b360bfd1.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f8bb1e15-e446-4d38-841a-bf75c5e24b93.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/350c8228-751a-4900-9fba-f5d3a9dc0b62.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/66b31fa1-49c6-4d7c-ad6e-06cbab2f5df5.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b300d1ff-1353-4be0-91d7-5e4caada56b5.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/9d836e82-465e-4545-a4a0-16c2b360bfd1.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a998a03c-29ee-4682-9870-249cc8569f51.html

3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b300d1ff-1353-4be0-91d7-5e4caada56b5.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/efc2bece-ec71-4d07-bcc2-c0997faf8556.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/8a5bc449-c23d-483e-998f-d9e4dd1b65a4.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/ba915459-dc48-4bef-aee1-cf561d657b70.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f4f4fcf9-4c0c-4ea0-b607-7faedd9e7c3d.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/3ffe8efe-444f-4028-9ee9-e985e6607bb9.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0edf2bf1-7b8e-4510-96ec-4bdd604a9943.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b691bd3c-5adb-4416-b297-a4f1ee5e9d10.html

Query 5: California Wild Fires (Overlap: 8)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/6161ce32-3903-454b-ba53-27f4d027cc97.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0d640342-a0fc-4d93-81d1-9b20f3faee17.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/720c608b-8301-4b9c-9487-e0215d078527.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/00a22bfe-524a-4fbf-a5d3-8a84975cd0a7.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/6b7539d0-8254-411b-be17-0aef372904a7.html

7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/5f2463c7-f75b-4696-a174-6226ebee6a4c.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/14070f88-d8f8-4083-8447-b0a41948b72d.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a4efd881-9f5b-4178-b6c1-1bd7099964da.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f9a9fa05-b6c1-4ab2-b2e2-1966d9cd46f5.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/720c608b-8301-4b9c-9487-e0215d078527.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a4efd881-9f5b-4178-b6c1-1bd7099964da.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f9a9fa05-b6c1-4ab2-b2e2-1966d9cd46f5.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/14070f88-d8f8-4083-8447-b0a41948b72d.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0d640342-a0fc-4d93-81d1-9b20f3faee17.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/6161ce32-3903-454b-ba53-27f4d027cc97.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f0dbb99f-8aef-4beb-8317-1ebbe54d8449.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/5f2463c7-f75b-4696-a174-6226ebee6a4c.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/54232f0d-32f1-4803-b1b0-593393aa6b68.html

Query 6: Donald Trump (Overlap: 0)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/c83367ce-93a8-48ef-90f2-f04023066977.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0e57ffed-1ea8-4d5f-8445-dcee42d638f0.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/ce384ef2-7c0d-420e-8868-f962413cec0d.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/9c16e628-1656-4178-92c2-3200f5dfa8d4.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/94c4e432-72b0-493c-a02e-e4559dfe852e.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/eb06e9ee-1437-4d09-bc19-bfa3a1303661.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/513708f8-281d-4eb5-89fd-11c9cdd973c8.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/1b63207d-5f11-4401-ae01-60decf799375.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/fbf2c2ac-1327-40bd-8352-37172455c6ba.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0dab20dd-3a36-40ff-9d30-75b355de84fe.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f49e83d5-beea-4b27-8cb3-0b3d4ed4d6d9.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0e0e0cb8-4116-4f8b-bdc8-8e725f545742.html

4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0e36b398-664e-472b-9d66-b99371c0d28b.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/0430a063-5676-4ddf-a142-f1f03af91221.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a21292f3-083a-4ceb-81fa-5df7440074a0.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/be7795a3-db46-41f1-a98d-1c8c5e8398ad.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f9480bc1-68ca-4937-b52b-29c3b511ef96.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/720c608b-8301-4b9c-9487-e0215d078527.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/d2d31706-8b85-4bbf-8246-b2e8de3b40d7.html

Query 7: Harry Potter (Overlap: 5)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/284d3f79-fca8-4f02-96a0-c9fdc7ebec5f.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a015289e-0924-4b2a-afea-c79fecb8ddbc.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/af423aaa-ecdd-45b3-94ab-983aedeff593.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/4d5f08b3-777b-4997-ae0f-9894a895f60b.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/fbcfe25e-69b6-48c6-90a3-fad438c19139.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/e5761493-439c-4072-8b21-bd8999bba50f.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bb9e4a61-8922-422f-9041-52ca3af8fc8d.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/ea9ae93b-c7a8-40dd-8886-b22e50b012c1.html

9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/82c3fb54-7397-42e7-8fac-ade3546a8b5c.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/97dd5c14-bf30-4e11-8489-eea05df1f061.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/ea9ae93b-c7a8-40dd-8886-b22e50b012c1.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/1e6b0cd0-cd0a-403a-827e-189e009366b4.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/82c3fb54-7397-42e7-8fac-ade3546a8b5c.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/aeeffa58-19d4-4f19-8c8c-41447ca83b49.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/f5129207-5e77-4f29-ba76-78913ad69088.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bb9e4a61-8922-422f-9041-52ca3af8fc8d.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/6856bebb-1904-4e99-81cf-a265a5b94be8.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/fb104bcd-83db-4b76-b779-afef57c79c59.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/284d3f79-fca8-4f02-96a0-c9fdc7ebec5f.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/e5761493-439c-4072-8b21-bd8999bba50f.html

Query 8: Brazil (Overlap: 0)

Solr

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/5eaf7694-9685-4ade-8853-9d02427c30ef.html

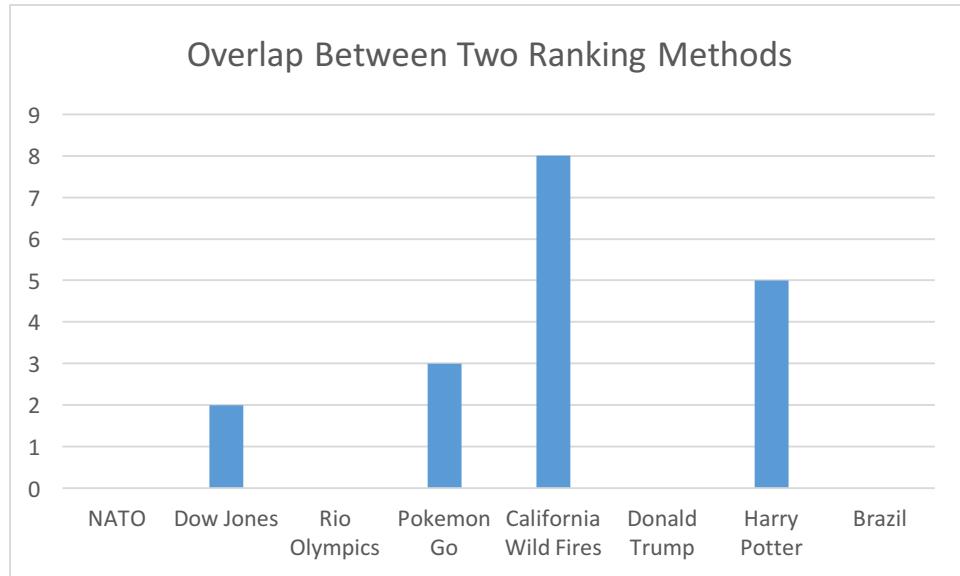
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/2bb3b7fd-63ef-405c-bf7e-7dc671c28987.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/38036043-160c-48f7-95ac-c86380787291.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/87c69936-abf8-429c-ae23-121616e984e8.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/581b2fd1-5289-418d-8920-e457af075e80.html
6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/8c91e06f-1f9e-47f8-aa7f-6df55d09e398.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/56225b6d-95f5-448f-96d9-0d2d0ce266bd.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/64484523-5a35-42f0-87ca-9bb772dd2e03.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/3c2d97a4-b6de-43b2-a86b-8ccee093b0ca.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/1cd43a1a-fc01-436a-91c6-eba7dec1db89.html

PageRank

1. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
2. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/7301a5b8-6fbc-4bda-bbbd-98324149606a.html
3. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a5f2fa3e-2ce6-4bd1-8abe-21250b451899.html
4. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/c08a6ff5-1447-445c-bf1f-25cc8c7245e5.html
5. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/201edd71-2149-40dd-b183-28fb3b5a671e.html

6. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/a0d3b3eb-3f50-4ee1-8753-cdb6ebe4b167.html
7. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/6bda7d38-0725-43f5-b551-278ff98664c5.html
8. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/aa885cd7-4b1f-4460-968b-d9c3cae8ad05.html
9. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/cc0b826b-570b-4c38-a8b0-1f38c76c7040.html
10. /Users/saigeetha/Documents/School Documents/Fall
2016/Assignments/CSCI_572_HW3/crawl/b32f29bb-955a-4d73-a810-fa39380d863a.html

3.) Overlap graph showing the amount of overlap between the two ranking methods



4.) PageRank says nothing about the content or size of a page, the text in the anchor of the link or the language it is written in. The PageRank of each page depends on the PageRank of the other pages. In general, PageRank provides a means to determine the value of a website for any given search query. The PageRank value is determined by how websites link together with the more popular sites receiving more links. It's because of these incoming links, some pages have a high PageRank value and thereby display higher up in search results.

5.)

a) Query Page



b) Enter a query



c) Click on Submit (For Solr Ranking method)

The screenshot shows a web browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Donald+Trump". The search bar contains "Donald Trump" with a "Submit" button and a "Use Page Rank" checkbox. Below the search bar, it says "Results 1 - 10 of 1983". The results are listed in three numbered sections:

- 1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/c83367ce-93a8-48ef-90f2-f04023066977.html
description	Recent and archived work by Maggie Haberman for The New York Times
title	Maggie Haberman
content_encoding	UTF-8
stream_size	111032
url:	http://www.nytimes.com/by/maggie-haberman
- 2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/0e57ffed-1ea8-4d5f-8445-dcee42d638f0.html
description	Recent and archived work by Alexander Burns for The New York Times
title	Alexander Burns
content_encoding	UTF-8
stream_size	110475
url:	http://www.nytimes.com/by/alexander-burns
- 3.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/ce384ef2-7c0d-420e-8868-f962413cc0d.html
-----------	-------------------------------------------------------------------------------------------------------------------------------

d) Click on any url (Clicked on first url <http://www.nytimes.com/by/maggie-haberman>)

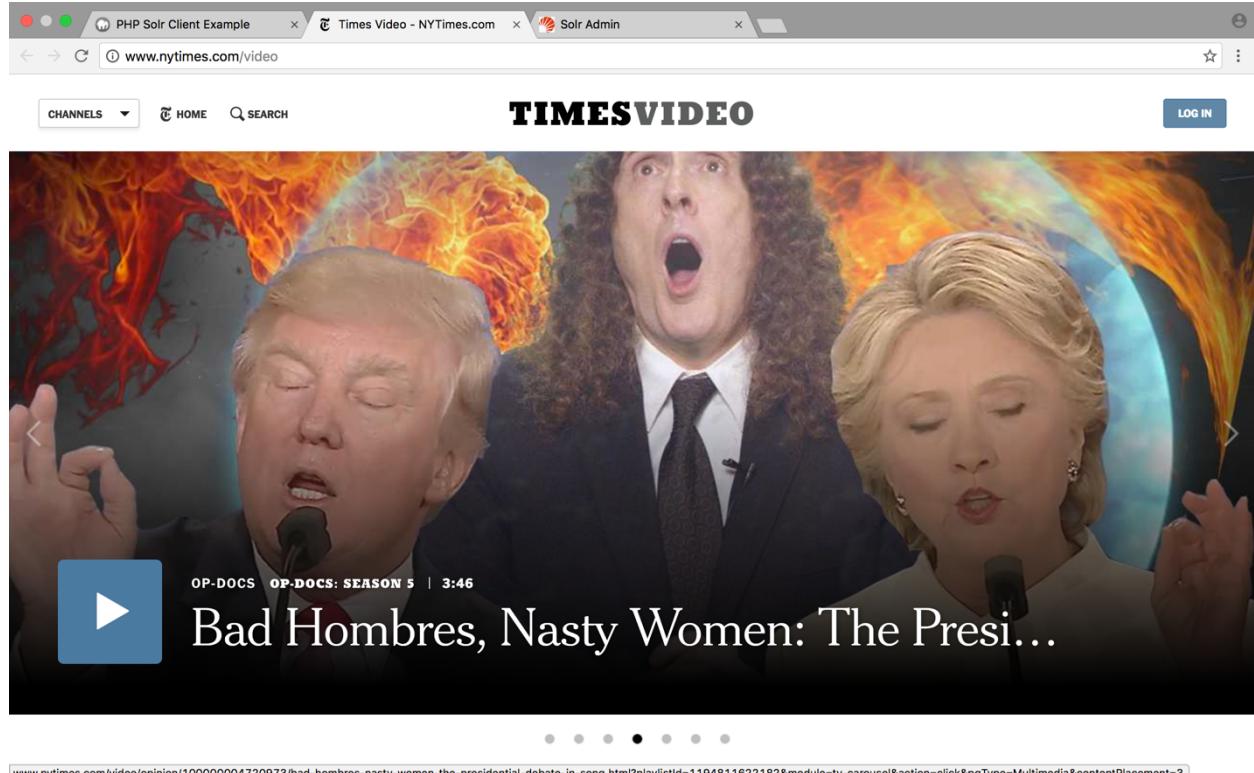
Screenshot of a web browser showing the New York Times homepage for Maggie Haberman. The page features a large advertisement for Spectrum, which is now Time Warner Cable. Below the ad, there is a headline about William Weld supporting Hillary Clinton. On the right side, there is a photo of two men in suits and a link to follow Maggie Haberman on Twitter.

- e) Now for the same query, select Use Page Rank checkbox for searching the results based on PageRank method

Screenshot of a web browser showing the Solr Admin interface. A search query "Donald Trump" is entered, and the "Use Page Rank" checkbox is checked. The results show two documents. The first document is a video about Donald Trump, and the second is an article about Today's Paper.

1.	id /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/f49e83d5-beea-4b27-8cb3-0b3d4ed4d6d9.html content_encoding UTF-8 description Whether it's reporting on conflicts abroad and political divisions at home, or covering the latest style trends and scientific developments, Times Video journalists provide a revealing and unforgettable view of the world. title Times Video - NYTimes.com stream_size 177846 url: http://www.nytimes.com/video
2.	id /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html stream_size 211360 description Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue. title Today's Paper - The New York Times content_encoding UTF-8 url: http://www.nytimes.com/pages/todayspaper/index.html

f) Click on any url (Clicked on first URL <http://www.nytimes.com/video>)



g) Query 1: NATO using Solr

PHP Solr Client Example CSCI 572 Home Page

localhost:8888/solr-php-client-master/?q=NATO

Search: NATO Submit Use Page Rank

Results 1 - 10 of 195:

1. [**id**](#) /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/ef88cc69-e232-4482-a76e-9111012ac4fc.html

content_encoding UTF-8
author Jochen Bittner

description The Atlantic alliance seems to have forgotten its purpose.
title Does NATO Still Exist? - The New York Times

stream_size 71190

url: <http://www.nytimes.com/2016/07/08/opinion/does-nato-still-exist.html>
2. [**id**](#) /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/1b4ac96f-209f-4be5-b73d-a8bae7b054db.html

description A top US commander says Russia can move large military forces to Eastern Europe faster than Nato.
title US Nato general fears rapid Russian troop deployments - BBC News

h) Query 1: NATO using PageRank

PHP Solr Client Example CSCI 572 Home Page

localhost:8888/solr-php-client-master/?q=NATO&pagerank=on

Search: NATO Submit Use Page Rank

Results 1 - 10 of 195:

1. [**id**](#) /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/b7114274-086c-433f-9bad-950c8e85f678.html

description A look at continuing efforts to reduce the water and energy needed to cool the world's data centres.
title Is the world wide web a waste of water? - BBC News

content_encoding UTF-8
stream_size 147132

url: <http://www.bbc.com/news/business-37471911>
2. [**id**](#) /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a7af769c-5ee9-4b9c-a507-4f6ed80d7d7e.html

description This election campaign has so far been driven more by personality clashes than differences in policy. But what are the key issues and where do Hillary Clinton and Donald Trump stand?
title US election: Where Clinton and Trump stand on key issues - BBC News

content_encoding UTF-8
stream_size 151392

url: <http://www.bbc.com/news/election-us-2016-37468751>
3. _____

i) Query 2: Dow Jones using Solr

The screenshot shows a browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Dow+Jones". The search bar contains "Dow Jones". The results page displays two search results, each with a numbered list item (1., 2.) followed by a table of document fields.

Result 1:

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/4d0d947c-e076-47e8-9770-0085d2f36e0c.html
content_encoding UTF-8	
author	The Associated Press
description	The highest reading on the Consumer Confidence Index since August 2007 fueled a broad rise led by tech and travel companies, including cruise lines.
title	Surge in Consumer Confidence Drives Rally After Days of Losses - The New York Times
stream_size	81830
url:	http://www.nytimes.com/2016/09/28/business/stock-markets.html?ref=todayspaper

Result 2:

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/cc32db42-fad6-4093-bfe8-36dbde3caced.html
content_encoding UTF-8	
author	The Associated Press

j) Query2: Dow Jones using PageRank

The screenshot shows a browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Dow+Jones&pagerank=on". The search bar contains "Dow Jones". The results page displays three search results, each with a numbered list item (1., 2., 3.) followed by a table of document fields.

Result 1:

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
stream_size	211360
description	Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.
title	Today's Paper - The New York Times
content_encoding UTF-8	
url:	http://www.nytimes.com/pages/todayspaper/index.html

Result 2:

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/8d259cc3-f267-4456-830d-4e0913a9edec.html
description	The latest Economy News from the BBC: breaking news on the global and UK economy and international investments including audio and video coverage.
title	Economy - BBC News
content_encoding UTF-8	
stream_size	177451
url:	http://www.bbc.com/news/business/economy

Result 3:

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/7301a5b8-6fbc-4bda-bbbd-98324149606a.html
description	The latest BBC Business News: breaking personal finance, company, financial and economic news, plus insight and analysis into UK and global markets.

k) Query 3: Rio Olympics using Solr

The screenshot shows a browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Rio+Olympics". The search bar contains "Rio Olympics". The results page displays 331 items, with the first few listed below:

Results 1 - 10 of 331:

1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/e59dce9e-8459-4004-9b14-2ae3aea697b9.html
description	Interactive Stories from the 2016 Rio Olympics Interactive Stories from the 2016 Rio Olympics
title	Interactive Stories - The New York Times
content_encoding	UTF-8
stream_size	111228
url:	http://www.nytimes.com/spotlight/rio-olympics-2016-interactive-stories
2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/9ed0501b-99c3-4cb6-a790-8fb942f12ccd.html
description	Complete coverage of the 2016 Olympic Games in Rio de Janeiro Complete coverage of the 2016 Olympic Games in Rio de Janeiro
title	Rio Olympics 2016 - The New York Times
content_encoding	UTF-8
stream_size	194470
url:	http://www.nytimes.com/news-event/rio-olympics-2016
3.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/d0d3de23-43ef-4dc1-8ade-8d0a2d5a8b89.html
-----------	--------------------------------------------------------------------------------------------------------------------------------

l) Query 3: Rio Olympics using PageRank

The screenshot shows a browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Rio+Olympics&pagerank=on". The search bar contains "Rio Olympics". The results page displays 331 items, with the first few listed below:

Results 1 - 10 of 331:

1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/7301a5b8-6fbc-4bda-bbbd-98324149606a.html
description	The latest BBC Business News: breaking personal finance, company, financial and economic news, plus insight and analysis into UK and global markets.
title	Business of Sport - BBC News
content_encoding	UTF-8
stream_size	186286
url:	http://www.bbc.com/news/business/business_of_sport
2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/eb0dc35b-266e-42c9-9acc-e8ff3c644447.html
title	Explainers - BBC News
content_encoding	UTF-8
stream_size	128166
url:	http://www.bbc.com/news/explainers
3.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/7ad991c7-8778-4bf4-964e-6cc9c7c6175b.html
stream_size	167628

m) Query 4: Pokemon Go using Solr

The screenshot shows a web browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Pokemon+Go". The search term "Pokemon Go" is entered in the search bar. Below the search bar, there is a checkbox labeled "Use Page Rank". The results section displays "Results 1 - 10 of 32:" followed by three numbered search results.

1. [redacted]
id /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/8dd0dc28-5358-496f-90fe-844f74b6310c.html

description New independent data suggests that the augmented reality craze is beginning to cool off.
title Why Pokemon Go may have passed its peak - BBC News

content_encoding UTF-8
stream_size 144288

url: http://www.bbc.com/news/technology-37176782?intlink_from_url&link_location=live-reporting-story

2. [redacted]
id /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/33a2875b-79b0-4997-83b4-717114d8f9eb.html

description BBC Click's Lara Lewington looks at some of the best of the week's technology news.
title Man jailed for Pokemon Go game in church and other news - BBC News

content_encoding UTF-8
stream_size 108788

url: <http://www.bbc.com/news/technology-37308759>

3. [redacted]

n) Query 4: Pokemon Go using PageRank

The screenshot shows a web browser window titled "PHP Solr Client Example" with the URL "localhost:8888/solr-php-client-master/?q=Pokemon+Go&pagerank=on". The search term "Pokemon Go" is entered in the search bar. Below the search bar, there is a checkbox labeled "Use Page Rank". The results section displays "Results 1 - 10 of 32:" followed by three numbered search results.

1. [redacted]
id /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/9d836e82-465e-4545-a4a0-16c2b360bfd1.html

description Candy Crush Saga launches its 2,000th level - but a decision to add adverts could prove controversial.
title Candy Crush Saga: Life beyond level 2,000 - BBC News

content_encoding UTF-8
stream_size 146157

url: <http://www.bbc.com/news/technology-37484114>

2. [redacted]
id /Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a998a03c-29ee-4682-9870-249cc8569f51.html

description BBC Click's Lara Lewington looks at some of the best of the week's technology news.
title The robot which can climb stairs and other tech news - BBC News

content_encoding UTF-8
stream_size 108460

url: <http://www.bbc.com/news/technology-37441971>

3. [redacted]

o) Query 5: California Wild Fires using Solr

PHP Solr Client Example		CSCI 572 Home Page												
<input type="text" value="localhost:8888/solr-php-client-master/?q=California+Wild+Fires"/> <input type="button" value="Submit"/> <input type="checkbox"/> Use Page Rank														
Results 1 - 10 of 17:														
1.														
<table border="1"> <tr> <td>id</td><td>/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/6161ce32-3903-454b-ba53-27f4d027cc97.html</td></tr> <tr> <td>stream_size</td><td>75072</td></tr> <tr> <td>title</td><td>NY Times Magazine - The New York Times</td></tr> <tr> <td>content_encoding</td><td>UTF-8</td></tr> <tr> <td>url:</td><td>http://www.nytimes.com/indexes/2009/10/25/magazine/index.html</td></tr> </table>			id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/6161ce32-3903-454b-ba53-27f4d027cc97.html	stream_size	75072	title	NY Times Magazine - The New York Times	content_encoding	UTF-8	url:	http://www.nytimes.com/indexes/2009/10/25/magazine/index.html		
id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/6161ce32-3903-454b-ba53-27f4d027cc97.html													
stream_size	75072													
title	NY Times Magazine - The New York Times													
content_encoding	UTF-8													
url:	http://www.nytimes.com/indexes/2009/10/25/magazine/index.html													
2.														
<table border="1"> <tr> <td>id</td><td>/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html</td></tr> <tr> <td>stream_size</td><td>211360</td></tr> <tr> <td>description</td><td>Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.</td></tr> <tr> <td>title</td><td>Today's Paper - The New York Times</td></tr> <tr> <td>content_encoding</td><td>UTF-8</td></tr> <tr> <td>url:</td><td>http://www.nytimes.com/pages/todayspaper/index.html</td></tr> </table>			id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html	stream_size	211360	description	Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.	title	Today's Paper - The New York Times	content_encoding	UTF-8	url:	http://www.nytimes.com/pages/todayspaper/index.html
id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html													
stream_size	211360													
description	Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.													
title	Today's Paper - The New York Times													
content_encoding	UTF-8													
url:	http://www.nytimes.com/pages/todayspaper/index.html													
3.														
<table border="1"> <tr> <td>id</td><td>/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/0d640342-a0fc-4d93-81d1-9b20f3face17.html</td></tr> <tr> <td>stream_size</td><td>120558</td></tr> <tr> <td>description</td><td>Find breaking news, multimedia and college football news on colleges and universities, the Rose Bowl, statistics, injury reports, the coaches and the players and B.C.S. rankings.</td></tr> <tr> <td>title</td><td>Sports News - College Football - The New York Times</td></tr> <tr> <td>content_encoding</td><td>UTF-8</td></tr> </table>			id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/0d640342-a0fc-4d93-81d1-9b20f3face17.html	stream_size	120558	description	Find breaking news, multimedia and college football news on colleges and universities, the Rose Bowl, statistics, injury reports, the coaches and the players and B.C.S. rankings.	title	Sports News - College Football - The New York Times	content_encoding	UTF-8		
id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/0d640342-a0fc-4d93-81d1-9b20f3face17.html													
stream_size	120558													
description	Find breaking news, multimedia and college football news on colleges and universities, the Rose Bowl, statistics, injury reports, the coaches and the players and B.C.S. rankings.													
title	Sports News - College Football - The New York Times													
content_encoding	UTF-8													

p) Query 5: California Wild Fires using PageRank

PHP Solr Client Example		CSCI 572 Home Page												
<input type="text" value="localhost:8888/solr-php-client-master/?q=California+Wild+Fires&pagerank=on"/> <input type="button" value="Submit"/> <input type="checkbox"/> Use Page Rank														
Results 1 - 10 of 17:														
1.														
<table border="1"> <tr> <td>id</td><td>/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html</td></tr> <tr> <td>stream_size</td><td>211360</td></tr> <tr> <td>description</td><td>Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.</td></tr> <tr> <td>title</td><td>Today's Paper - The New York Times</td></tr> <tr> <td>content_encoding</td><td>UTF-8</td></tr> <tr> <td>url:</td><td>http://www.nytimes.com/pages/todayspaper/index.html</td></tr> </table>			id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html	stream_size	211360	description	Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.	title	Today's Paper - The New York Times	content_encoding	UTF-8	url:	http://www.nytimes.com/pages/todayspaper/index.html
id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html													
stream_size	211360													
description	Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.													
title	Today's Paper - The New York Times													
content_encoding	UTF-8													
url:	http://www.nytimes.com/pages/todayspaper/index.html													
2.														
<table border="1"> <tr> <td>id</td><td>/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/720cc608b-8301-4b9c-9487-e0215d078527.html</td></tr> <tr> <td>description</td><td>The New York Times: Find breaking news, multimedia, reviews & opinion on Washington, business, sports, movies, travel, books, jobs, education, real estate, cars & more at nytimes.com.</td></tr> <tr> <td>title</td><td>The New York Times - Breaking News, World News & Multimedia</td></tr> <tr> <td>content_encoding</td><td>UTF-8</td></tr> <tr> <td>stream_size</td><td>191144</td></tr> <tr> <td>url:</td><td>http://www.nytimes.com/</td></tr> </table>			id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/720cc608b-8301-4b9c-9487-e0215d078527.html	description	The New York Times: Find breaking news, multimedia, reviews & opinion on Washington, business, sports, movies, travel, books, jobs, education, real estate, cars & more at nytimes.com.	title	The New York Times - Breaking News, World News & Multimedia	content_encoding	UTF-8	stream_size	191144	url:	http://www.nytimes.com/
id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/720cc608b-8301-4b9c-9487-e0215d078527.html													
description	The New York Times: Find breaking news, multimedia, reviews & opinion on Washington, business, sports, movies, travel, books, jobs, education, real estate, cars & more at nytimes.com.													
title	The New York Times - Breaking News, World News & Multimedia													
content_encoding	UTF-8													
stream_size	191144													
url:	http://www.nytimes.com/													
3.														
<table border="1"> <tr> <td>id</td><td>/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a4efd881-9f5b-4178-b6c1-1bd7099964da.html</td></tr> <tr> <td>stream_size</td><td>295164</td></tr> <tr> <td colspan="3">Your destination to buy, sell or research a new or used car: Research and compare cars, read car reviews, explore classic and collectible cars, search auto classifieds, find a</td></tr> </table>			id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a4efd881-9f5b-4178-b6c1-1bd7099964da.html	stream_size	295164	Your destination to buy, sell or research a new or used car: Research and compare cars, read car reviews, explore classic and collectible cars, search auto classifieds, find a							
id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a4efd881-9f5b-4178-b6c1-1bd7099964da.html													
stream_size	295164													
Your destination to buy, sell or research a new or used car: Research and compare cars, read car reviews, explore classic and collectible cars, search auto classifieds, find a														

q) Query 7: Harry Potter using Solr

PHP Solr Client Example CSCI 572 Home Page

localhost:8888/solr-php-client-master/?q=Harry+Potter

Search: Harry Potter Submit Use Page Rank

Results 1 - 10 of 47:

1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/284d3f79-fca8-4f02-96a0-c9fdc7ebec5f.html
content_encoding	UTF-8
author	Mekado Murphy
title	La guía de animales fantásticos del mundo de Harry Potter – Español
stream_size	52015
url:	http://www.nytimes.com/es/2016/09/23/la-guia-de-animales-fantasticos-del-mundo-de-harry-potter/

2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a015289e-0924-4b2a-afea-c79fecb8ddbc.html
description	The Harry Potter franchise is worth about £4bn to the London's economy, according to the London School of Economics.
title	Harry Potter's magical effect 'worth £4bn to London' - BBC News
content_encoding	UTF-8
stream_size	105934
url:	http://www.bbc.com/news/uk-england-london-37501694?intlink_from_url&link_location=live-reporting-map

r) Query 7: Harry Potter using PageRank

PHP Solr Client Example CSCI 572 Home Page

localhost:8888/solr-php-client-master/?q=Harry+Potter&pagerank=on

Search: Harry Potter Submit Use Page Rank

Results 1 - 10 of 47:

1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/ea9ae93b-c7a8-40dd-8886-b22e50b012c1.html
title	Noticias – Español
content_encoding	UTF-8
stream_size	130040
url:	http://www.nytimes.com/es/category/noticias/

2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/1e6b0cd0-cd0a-403a-827e-189e009366b4.html
stream_size	192597
description	Reviews, analysis and commentary about the latest devices, apps and technology trends.
title	Technology News - Personal Tech - The New York Times
content_encoding	UTF-8
url:	http://www.nytimes.com/pages/technology/personaltech/index.html

3.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/82c3fb54-7397-42e7-8fac-ade3546a8b5c.html
description	It's not just Harry Potter's house which had come on the market - here are some ordinary homes with extraordinary screen presence.
title	Not just Potter's house - ordinary homes for showbiz fans - BBC News
content_encoding	UTF-8
stream_size	152048

s) Query 8: Brazil using Solr

PHP Solr Client Example CSCI 572 Home Page

localhost:8888/solr-php-client-master/?q=Brazil

Search: Brazil Submit Use Page Rank

Results 1 - 10 of 662:

1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/5eaf7694-9685-4ade-8853-9d02427c30ef.html
content_encoding	UTF-8
author	Joe Nocera
description	A struggling economy, government corruption and the Zika virus have generated unwanted publicity before the Games.
title	As the Olympics Near, Brazil and Rio Let the Bad Times Roll - The New York Times
stream_size	101727
url:	http://www.nytimes.com/2016/06/25/sports/olympics/as-the-olympics-near-brazil-and-rio-let-the-bad-times-roll.html?ref=collection%2Fcolumn%2FSports%20Business
2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/2bb3b7fd-63ef-405c-bf7e-7dc671c28987.html
content_encoding	UTF-8

t) Query 8: Brazil using PageRank

PHP Solr Client Example CSCI 572 Home Page

localhost:8888/solr-php-client-master/?q=Brazil&pagerank=on

Search: Brazil Submit Use Page Rank

Results 1 - 10 of 662:

1.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/bd4da657-62a4-400a-9c74-69d87c6819a9.html
stream_size	211360
description	Use the Today's Paper page to see all the headlines from the Final City Edition of The New York Times organized in the same sections as they appeared in print. You can also view a front page image and read the quote of the day. Use the calendar feature to see a previous day's headlines or use search to look for articles published in a specific past issue.
title	Today's Paper - The New York Times
content_encoding	UTF-8
url:	http://www.nytimes.com/pages/todayspaper/index.html
2.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/7301a5b8-6fbc-4bda-bbbd-98324149606a.html
description	The latest BBC Business News: breaking personal finance, company, financial and economic news, plus insight and analysis into UK and global markets.
title	Business of Sport - BBC News
content_encoding	UTF-8
stream_size	186286
url:	http://www.bbc.com/news/business/business_of_sport
3.

id	/Users/saigeetha/Documents/School Documents/Fall 2016/Assignments/CSCI_572_HW3/crawl/a5f2fa3e-2ce6-4bd1-8abe-21250b451899.html
description	Get the latest African news from BBC News in Africa: breaking news, features, analysis and special reports plus audio and video from across the African continent.