

**The University of Texas at Dallas  
CS 6322  
Information Retrieval  
Spring 2023  
Class Project Report**

Project TITLE: Search engine for SWIMMING

TEAM: 1

Students:

1. Pallavi Dasarapu, [pallavi.dasarapu@utdallas.edu](mailto:pallavi.dasarapu@utdallas.edu)
2. Sangeetha Pradeep, [sangeetha.pradeep@utdallas.edu](mailto:sangeetha.pradeep@utdallas.edu)
  - 3. Abishek Kumar, [abishek.kumar@utdallas.edu](mailto:abishek.kumar@utdallas.edu)
  - 4. Sai Lasya Akkineni, [sailasya.akkineni@utdallas.edu](mailto:sailasya.akkineni@utdallas.edu)
  - 5. Devarshaa Sikharam, [devarshaa.sikharam@utdallas.edu](mailto:devarshaa.sikharam@utdallas.edu)

1. The Problem:

In this project, we have designed and implemented a search engine for **Swimming** called '**Dive**'. Here we present a well-designed architecture of Dive that effectively retrieves relevant information about swimming.

Topic Owners:

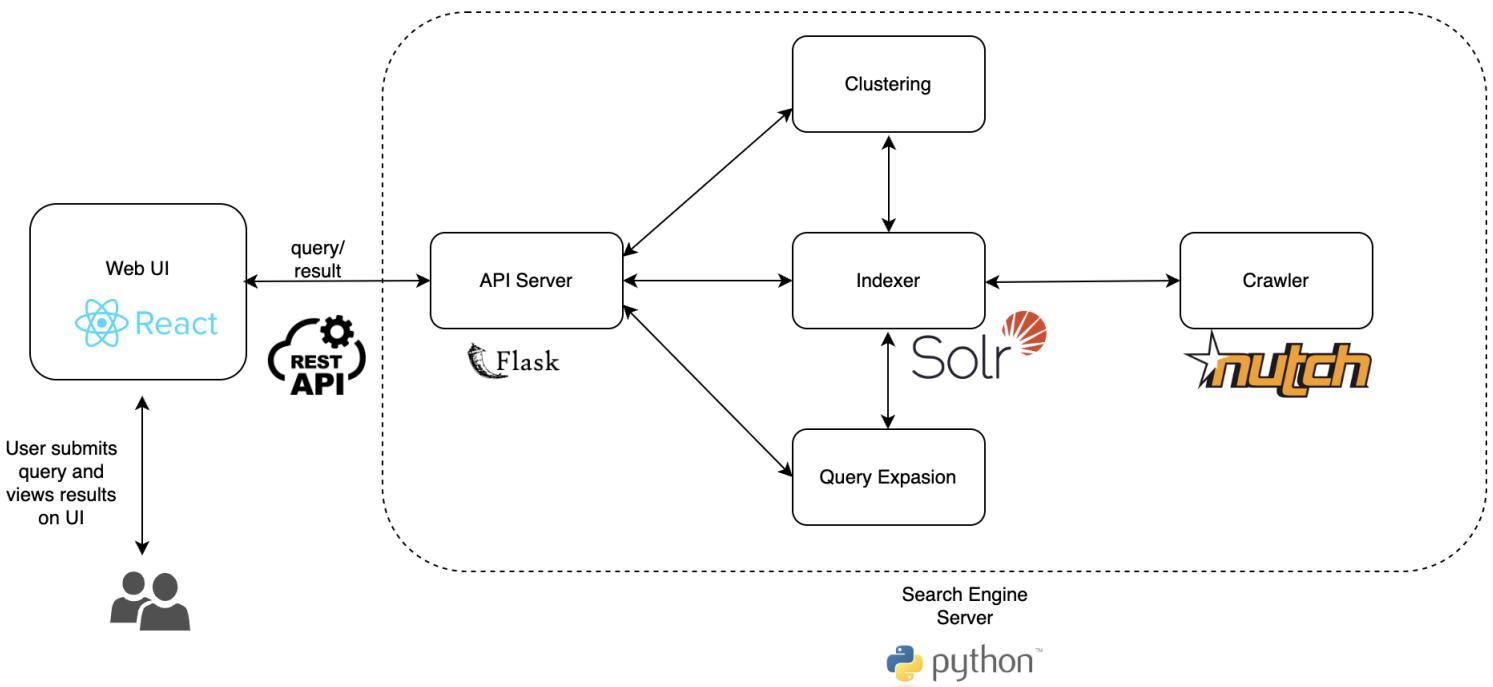
**Crawling:** Pallavi Dasarapu (PXD210008)

**Indexing and Relevance:** Sangeetha Pradeep

**User Interface and Comparison with Google and Bing:** Abishek Kumar (axk200020)

**Clustering:** Sai Lasya Akkineni (SXA210025)

**Query Expansion and Relevance Feedback:** Devarshaa Sikharam



Architectural Diagram of Dive - The Search Engine for Swimming

#### Learnings:

- **Technical Skills:** From this project, we learned how information retrieval works and acquired knowledge on using web crawler (Apache Nutch), indexer (Apache solr), programming various algorithms for ranked retrieval, clustering and query expansion and building a user-friendly interface for the search engine.
- **Scalability:** Being able to scale our search engine with the data growth was also one of the challenges. We learned different techniques and tools to handle this issue.

#### Challenges:

- **Scalability:** Being able to scale our search engine with the data growth was challenging even though our search engine only dealt with a smaller fraction of web pages in this project.
- **Data Collection:** One of the biggest challenges in building this search engine was collecting and indexing large volumes of data. This was time-consuming and resource-intensive.
- **Data Quality:** Ensuring the quality of the web pages crawled was also crucial. Data may be incomplete, outdated, or contain errors, affecting the relevance and reliability of search results.

#### How we dealt:

- Use web crawling tools like Apache Nutch to crawl and index relevant data combined with some level of data cleaning to ensure that the data is of high quality.
- We wanted to use services like AWS to perform storage and compute intensive tasks. We instead ran such tasks on dedicated UTD servers with higher compute power

Overall, building a search engine was challenging but a rewarding experience that helped us gain insights into the complexities of information retrieval.

*Crawling*  
*Pallavi Dasarapu (PXD210008)*

## 2. Crawling:

We utilized various Python libraries such as Scrapy, Beautiful Soup, and Selenium, among others, for web crawling purposes. After evaluating the pros and cons of different crawlers, we determined that Apache Nutch was the best tool for our project, and subsequently employed it to crawl approximately 110 seed URLs, acquired through different search engines like Google, Bing, and Yahoo, among others.

Following are a few of the seed urls used:

- 1.
2. <https://en.wikipedia.org/wiki/Swimming>
3. <https://www.britannica.com/sports/swimming-sport>
4. <https://www.swimmingworldmagazine.com/news/an-honest-sport-15-truths-about-swimming>
5. <https://dbpedia.org/page/Swimming>
6. <https://www.sportsnhobbies.org/what-are-the-different-types-of-swimming-sports.htm>
7. <https://www.rookieroad.com/swimming>
8. <https://olympics.com/en/sports/swimming>
9. <https://www.cdc.gov/healthywater/swimming>
10. <https://www.kiefer.com>
11. <https://www.shutterstock.com/search/swimming-sport>
12. <https://www.swimming.org>
13. <https://www.paralympic.org/swimming>
14. <https://www.britishswimming.org>
15. <https://healthysportindex.com/sports>
16. <https://www.sportsengine.com/swimming>
17. <https://moveunitedsport.org/sport/swimming>
18. <https://ibsasport.org/growing-blind-sports/helping-other-sports-grow/swimming>
19. <https://www.yardbarker.com/swimming>
20. <https://apnews.com/hub/swimming>
21. <https://gofrogs.com/sports/swimming-and-diving>
22. <https://www.telegraph.co.uk/swimming>
23. <https://www.independent.co.uk/topic/swimming>
24. <https://www.sportskeeda.com/swimming>
25. <https://auburntigers.com/sports/swimming-and-diving>
26. <https://www.bbc.com/sport/swimming>
27. <https://www.yahoo.com/tagged/swimming>
28. <https://www.imagineswimming.com>
29. <https://www.yelp.com>
30. <https://swimjim.com>
31. <http://aquaskills.com/en>
32. <https://swimeasyny.com>
33. <https://physiqueswimming.com>

*Crawling*

Pallavi Dasarapu (PXD210008)

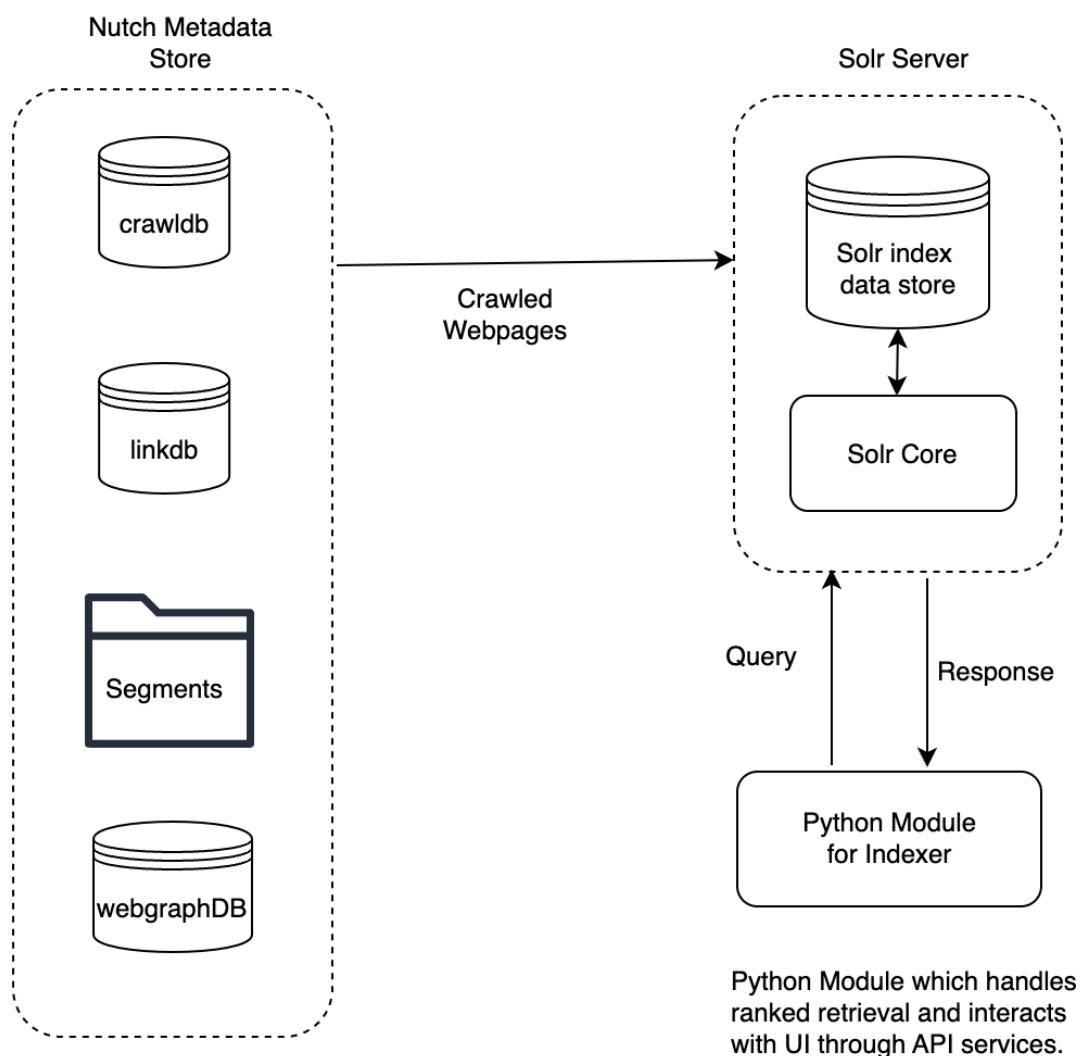
34. <https://www.goldfishswimschool.com>
35. <https://www.manhattanyouth.org/aquatics>
36. <https://my.lifetime.life/clubs/ny/sky-manhattan/programs/swim.html>
37. <https://takemetothewater.com>
38. <https://www.sunsationalswimschool.com>
39. <https://www.nysports4kids.com>
40. <https://aquaticsacademy.com>
41. <https://splashfit.org>
42. [https://en.wikipedia.org/wiki/Swimming\\_at\\_the\\_Summer\\_Olympics](https://en.wikipedia.org/wiki/Swimming_at_the_Summer_Olympics)
43. [https://en.wikipedia.org/wiki/List\\_of\\_swimmers](https://en.wikipedia.org/wiki/List_of_swimmers)
44. <https://www.triathlete.com>
45. <https://bigblueswimschool.com/blog/benefits-of-swimming-for-athletes>
46. <https://www.openwaterswimming.com>
47. [https://en.wikipedia.org/wiki/Open\\_water\\_swimming](https://en.wikipedia.org/wiki/Open_water_swimming)
48. <https://www.usms.org/fitness-and-training/articles-and-videos/articles/open-water-swimming-tips>
49. <https://www.openwaterpedia.com/wiki/>
50. <https://us.aquasphereswim.com/en/swim-goggles/open-water>
51. <https://oceannmanswim.com>
52. <https://www.pacificswim.co>
53. <https://www.redbull.com/ie-en/best-open-water-swim-races-in-the-world>
54. <https://www.dover.uk.com/channel-swimming/records>
55. [https://en.wikipedia.org/wiki/List\\_of\\_successful\\_English\\_Channel\\_swimmers](https://en.wikipedia.org/wiki/List_of_successful_English_Channel_swimmers)
56. <https://www.channelswimmingassociation.com>
57. <https://longswims.com/>

With the aid of Apache Nutch, we successfully crawled a total of 140,000 web pages. Nutch's crawl command helped us to efficiently gather, parse and index web pages from our seed URLs. Moreover, Nutch's deduplication feature ensured that we did not have any duplicate web pages in our crawled data. After each crawl iteration, the data was stored in dedicated folders. Additionally, I facilitated my teammates by granting necessary permissions and provided them with insights on the Nutch output to assist them in locating the relevant data and copying all the collected data, so that they could extract the required hyperlink information and proceed further.

### 3. Indexing and Relevance:

High Level Picture of Index:

To assemble the index, Apache Solr was used as the search indexer. The architecture of the system was designed to leverage the strengths of both Nutch and Solr to efficiently crawl, index, and compute page ranks and HITS scores for the web pages related to swimming.



Nutch data is used in conjunction with Solr to create the search engine index. The Nutch data are organised into the following three metadata stores:

- *The crawl database (crawldb)* - to keep track of every URL known to Nutch, including whether it was fetched and the timestamp of the fetch
- *The link database (linkdb)* - contains the list of known links to each URL, including the source URL and anchor text.
- *The segments directory* - contains a set of URLs that are fetched as a unit, with each subdirectory holding information about the status of fetching, raw content, parsed text, outlinks, and metadata.

Steps carried out to index the crawled web pages:

```
$ cd /to-nutch-root-folder (where binary and datastore resides)  
$ bin/nutch invertlinks crawl/linkdb -dir crawl/segments  
$ bin/nutch index crawl/crawldb -linkdb crawl/linkdb/ -dir crawl/segments/ -filter  
-normalize -deleteGone
```

Webgraph and statistics:

We used the `**bin/nutch webgraph**` command to generate the web graph for a given set of segments in Nutch.

```
$ bin/nutch webgraph -segmentDir crawl/segments/ -webgraphdb crawl/webgraphdb
```

The `**-webgraphdb**` option specifies the directory where the web graph will be stored in a format suitable for use with Hadoop or other distributed processing frameworks. It creates three databases, one for inlinks, one for outlinks, and a node database that holds the number of in and outlinks to a url and the current score for the url.

The resulting webgraph can be used for a variety of purposes, such as analyzing the structure of the web, identifying communities within the graph, or computing various metrics like PageRank or HITS. We have used the same for PageRank.

The WebGraph is an update-able database. Outlinks are stored by their fetch time or by the current system time if no fetch time is available. Only the most recent version of outlinks for a given url is stored. As more crawls are executed and the WebGraph updated, newer Outlinks will replace older Outlinks. This allows the WebGraph to adapt to changes in the link structure of the web.

We used a Python script to get some interesting statistics of the web graph thus generated and here are some statistics from analysis of the webgraph structure.

Total number of nodes	131095
Total number of links	2340991
Largest number of in-going links	10578
Largest number of out-going links	96

How information in the graph was used in Indexing?

Using the webgraph structure obtained, we scored the webpages using a Nutch plugin called '**LinkRank**'. LinkRank is an analysis program which is very similar to the PageRank algorithm. It is an iterative link analysis program that converges to stable global scores for each url. Once the LinkRank program has been run and link analysis is completed, the scores must be updated into the crawl database using the ScoreUpdater program. This updated data in the crawl database is then used for indexing. A detailed description of LinkRank and commands to run the same are discussed in the following section.

#### Weighing Schemes

Solr supports several weighting schemes for ranking search results: TF-IDF (Term Frequency-Inverse Document Frequency), BM25 (Best Matching 25), DFR (Divergence From Randomness), Language Models.

These weighting schemes can be configured in Solr's schema.xml file, and can be adjusted to suit the specific needs of a particular search application.

For our use case, we decided to go with the default weighting scheme supported by solr which is TF-IDF.

**TF-IDF (Term Frequency-Inverse Document Frequency)** - This weighting scheme assigns weights to terms based on their frequency in the document and the inverse frequency in the

entire corpus of documents. Terms that appear frequently in a document but rarely in the corpus are given higher weights.

It calculates a score for a word (term) in a document based on two factors:

1. *TF (Term Frequency)*: measures the frequency of a term in a document. It increases proportionally to the number of times a word appears in the document.
2. *IDF (Inverse Document Frequency)*: measures the importance of a term by calculating the inverse of its frequency across the entire corpus (i.e., all documents). It decreases the weight of words that occur very frequently in the corpus and increases the weight of words that occur rarely.

$$TF-IDF = TF \times IDF$$

Relevance Models:

### **Vector Space Model (VSM)**

Solr uses the Vector Space Model (VSM) for relevance ranking. VSM is a classic model for text retrieval that represents both documents and queries as vectors in a high-dimensional space. The relevance score between a query and a document is computed using a similarity function, such as cosine similarity, between their respective vectors in the space. Solr also provides various options to customize the VSM-based ranking, such as using different similarity functions, boosting certain fields, and applying query-time document boosting.

So when we query solr index, it retrieves results using VSM.

```
$ curl -X GET "http://localhost:8983/solr/nutch/select?wt=json&rows=30&q=swimming competitions"
```

The elements represented as vector space models are:

Query: swimming competitions

Document corpus

## Link Analysis Models

Graph-based relevance models are also called link analysis models. These models use the structure of the graph formed by hyperlinks between pages to determine the relevance of a page. We implemented two link analysis models which are PageRank and HITS (Hypertext Induced Topic Selection).

### PageRank

We performed the following steps in order to retrieve page ranked results.

1. We used the Nutch's linkRank plugin to calculate the PageRank for each page in the index using the "linkrank" command.
2. We then stored the PageRank values in a field in the Solr index, using the "updater" command. This will update the Solr index with the PageRank values for each page.
3. These PageRank values are used as a ranking factor in Solr's query ranking algorithm. This can be done by using the "pf" (phrase fields) and "qf" (query fields) parameters in the Solr query to boost the PageRank field. For example, to boost the PageRank field by a factor of 2, you would set the "pf" parameter to "PageRank^2"
4. Run queries against the Solr index, and the results will be ranked based on both text relevance and PageRank.

```
$ bin/nutch webgraph -segmentDir crawl/segments/ -webgraphdb crawl/webgraphdb  
$ bin/nutch linkrank -webgraphdb NewCrawl/webgraphdb/  
$ bin/nutch scoreupdater -crawldb crawl/crawldb -webgraphdb crawl/webgraphdb  
$ bin/nutch index crawl/crawldb -linkdb crawl/linkdb -dir crawl/segments/
```

### HITS (Hyperlink-Induced Topic Search)

HITS is a graph-based relevance model that considers both the authority and hub scores of web pages. The idea behind HITS is that web pages that are good hubs link to many good authorities, and web pages that are good authorities are linked to by many good hubs.

In HITS, each web page is given an **authority** score and a **hub** score. The authority score represents how likely a page is to contain high-quality information on a particular topic, while the hub score represents how well-connected a page is to other high-quality pages on that topic.

To implement HITS using Solr and Nutch, the web graph can be generated using Nutch's linkRank plugin. We then used a Python program to calculate the Hubs and the Authority scores for each URLs.

Here are some statistics:

1. Highest Scoring Hub: [https://en.wikipedia.org/wiki/Category:Open\\_water\\_swimming](https://en.wikipedia.org/wiki/Category:Open_water_swimming)  
Score: 0.01259
2. Highest Scoring Authority: <https://en.wikipedia.org/wiki/Swimming>  
Score: 0.02114925050427834

### Topic-based PageRanks

These are two topics used for testing topic based ranking.

1. Recreational Swimming
2. Competitive Swimming

Results for Recreational Swimming:

Title	URL	PageRankScore
Information about Pools, Hot Tubs, and Splash Pads by State   Healthy Swimming   Healthy Water   CDC	<a href="https://www.cdc.gov/healthywater/swimming/states-treated.html">https://www.cdc.gov/healthywater/swimming/states-treated.html</a>	0.0178
Talk:Open water swimming - Wikipedia	<a href="https://en.wikipedia.org/wiki/Talk:Open_water_swimming">https://en.wikipedia.org/wiki/Talk:Open_water_swimming</a>	0.016
Swimming Pools / Bathing Beaches / Recreational Aquatic Spray Grounds	<a href="https://www.health.ny.gov/environment/outdoors/swimming/">https://www.health.ny.gov/environment/outdoors/swimming/</a>	0.0089

Results for Competitive Swimming:

Title	URL	PageRankScore
Category:Competitive swimming by year - Wikimedia Commons	<a href="https://commons.wikimedia.org/wiki/Category:Competitive_swimming_by_year">https://commons.wikimedia.org/wiki/Category:Competitive_swimming_by_year</a>	0.0478
What are the Different Types of Competitive Swimming?	<a href="https://www.thehealthboard.com/what-are-the-different-types-of-competitive-swimming.htm">https://www.thehealthboard.com/what-are-the-different-types-of-competitive-swimming.htm</a>	0.0312
Open Water Swimming Collegiate National Championships - Openwaterpedia	<a href="https://www.openwaterpedia.com/wiki/American_Swimming_Association_Open_Water_Swimming_Collegiate_National_Championships">https://www.openwaterpedia.com/wiki/American_Swimming_Association_Open_Water_Swimming_Collegiate_National_Championships</a>	0.0276

Collaboration with UI and testing

We performed manual testing by using a set of queries to test the relevance models implemented as part of this project. Here are the results of randomly picked 5 queries tested on the UI and showing a contrast in the results of pagerank and HITS relevance model results.

*Indexing and Relevance*  
Sangeetha Pradeep (sxp210004)

Query : Olympic swimmers

Page Rank Result

Our results

- [Breaststroke - Openwaterpedia](#)  
Breaststroke Openwaterpedia Anonymous Not logged in Log in Openwaterpedia Search Check out our new site at openwaterswimming.com Breaststroke From...
- [Marathon swimmer - Openwaterpedia](#)  
Marathon swimmer Openwaterpedia Anonymous Not logged in Log in Openwaterpedia Search Check out our new site at openwaterswimming.com Marathon...
- [Tokyo Olympic Games - 10K Marathon Swim - Live Coverage – World Open Water Swimming Association](#)  
Tokyo Olympic Games K Marathon Swim Live Coverage World Open Water Swimming Association Home News News Top Educational Articles Top...
- [Pilot - Openwaterpedia](#)  
Pilot Openwaterpedia Anonymous Not logged in Log in Openwaterpedia Search Check out our new site at openwaterswimming.com Pilot From...
- [Oussama Mellouli - Openwaterpedia](#)  
Oussama Mellouli Openwaterpedia Anonymous Not logged in Log in Openwaterpedia Search Check out our new site at openwaterswimming.com Oussama...

HITS

Our results

- [Category:Olympic swimmers by country - Wikipedia](#)  
Category Olympic swimmers by country Wikipedia Jump to content Main menu Main menu move to sidebar hide Navigation Main page...
- [Category:Olympic swimmers by country - Wikimedia Commons](#)  
Category Olympic swimmers by country Wikimedia Commons Open main menu Home Random Nearby Log in Settings Donate About Wikimedia Commons...
- [Why Are There Lifeguards at Olympic Swimming Events?](#)  
Why Are There Lifeguards at Olympic Swimming Events Sports Sporting Events Games Activities x Sports Sporting Events Games Activities Swimming...
- [How Much Does A Swimmer Make?](#)  
How Much Does A Swimmer Make Sports Sporting Events Games Activities x Sports Sporting Events Games Activities Swimming What is...

In this query HITS provided better results compared to PageRank. HITS returned a Wikipedia result as the top one.

*Indexing and Relevance*  
Sangeetha Pradeep (sxp210004)

Query: Swimming workouts

Page Ranking

Our results

[Top 10 swimming ideas and inspiration](#)

Top swimming ideas and inspiration Pinterest Today Watch Shop Explore When autocomplete results are available use up and down arrows...

[Swimming Workouts - SwimSwam](#)

Swimming Workouts SwimSwam SwimSwam Subscribe to Newsletter Advertise Email Us Submit a Story Submit a Job Submit a College Recruit...

[Favorite Swimming Workouts](#)

Favorite Swimming Workouts Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr Basic Swimming...

[Swimming Workouts - Page 2 of 65 - SwimSwam](#)

Swimming Workouts Page of SwimSwam SwimSwam Subscribe to Newsletter Advertise Email Us Submit a Story Submit a Job Submit a...

HITS

Our results

[Swim.com | Compete, share and track your swimming workouts](#)

Swim com Compete share and track your swimming workouts How It Works Features Pools Clubs Blog Create your free account...

[Swimming Workouts and Nutrition Tips for Overweight People](#)

Swimming Workouts and Nutrition Tips for Overweight People Skip to Content Search Magnifying Glass Search for Close Search Home Learn...

[WOWSA Open Water Swimming Shop – World Open Water Swimming Association](#)

WOWSA Open Water Swimming Shop World Open Water Swimming Association Home News News Top Educational Articles Top Open Water Swimming...

[WOWSA Programs and Workouts – World Open Water Swimming Association](#)

WOWSA Programs and Workouts World Open Water Swimming Association Home News News Top Educational Articles Top Open Water Swimming Books...

[Best Swimming Workouts for Weight Loss | U.S. Masters Swimming](#)

Best Swimming Workouts for Weight Loss U S Masters Swimming Search Training Events Club Finder Workout Library About Us Join...

In this query, Pagerank returned a result from Pinterest while HITS returned a result from <https://www.swim.com/> a swim workout and training platform. HITS seems to out perform.

*Indexing and Relevance*  
Sangeetha Pradeep (sxp210004)

Query: Benefits of swimming  
Page Ranking

Our results

[\*\*The Health Benefits of Swimming\*\*](#)

The Health Benefits of Swimming Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...

[\*\*7 Health Benefits of Swimming\*\*](#)

Health Benefits of Swimming Menu Visit Verywell Fit's homepage Health Benefits of Swimming Fitness Beginners Walking Cardio Strength Running...

[\*\*10 Benefits of Swim Lessons for Kids | Big Blue Swim School\*\*](#)

Benefits of Swim Lessons for Kids Big Blue Swim School Big Blue Blog Kids Swimming Lessons Benefits Of Swim Lessons...

[\*\*Back to School: Mental Benefits of Swimming for Kids\*\*](#)

Back to School Mental Benefits of Swimming for Kids Big Blue Blog Kids Swimming Lessons A Back To School Boost...

[\*\*Swimming, Health, and Safety.\*\*](#)

Swimming Health and Safety Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr Basic...

HITS

Our results

[\*\*Water aerobics - Wikipedia\*\*](#)

Water aerobics Wikipedia Jump to content Main menu Main menu move to sidebar hide Navigation Main page Contents Current events...

[\*\*7 Health Benefits of Swimming\*\*](#)

Health Benefits of Swimming Menu Visit Verywell Fit's homepage Health Benefits of Swimming Fitness Beginners Walking Cardio Strength Running...

[\*\*Health Benefits of Swimming | Healthy Swimming | Healthy Water | CDC\*\*](#)

Health Benefits of Swimming Healthy Swimming Healthy Water CDC Skip directly to site content Skip directly to page options Skip...

[\*\*Working at Sunsational Swim School: 66 Reviews | Indeed.com\*\*](#)

Working at Sunsational Swim School Reviews Indeed.com Jobs Company reviews Find salaries Upload your resume Sign in Sign in...

[\*\*Jim Montgomery Masters Swim | U.S. Masters Swimming\*\*](#)

Jim Montgomery Masters Swim U.S. Masters Swimming Search Training Events Club Finder Workout Library About Us Join My USMS...

In this query, HITS returned a wikipedia page of water aerobics and has ranked the rest of the results better than PageRank.

*Indexing and Relevance*  
Sangeetha Pradeep (sxp210004)

Query: Swimming safety tips

Page Ranking

Our results

[Open Water - Canadian Red Cross](#)

Open Water Canadian Red Cross Skip to content Contact Shop Jobs My Red Cross Volunteer Search This Site Search Fran...

OPEN WATER SWIMMING SAFETY CONSIDERATIONS for INDIVIDUALS As public pools have closed due to Covid Swimming Natation Canada SNC has...

[20 Swimming ideas | swimming, swimming quotes, swimmers life](#)

Swimming ideas swimming swimming quotes swimmers life Pinterest Today Watch Shop Explore When autocomplete results are available use up and...

[Water Safety Blog - Sunsational Swim School](#)

Water Safety Blog Sunsational Swim School Toggle navigation Toggle navigation Sunsational Swim School LOGIN Sunsational Swim School ABOUT Blog About...

[Swimming Tips for Beginners and Experienced Swimmers](#)

Swimming Tips for Beginners and Experienced Swimmers Skip to Content Search Magnifying Glass Search for Close Search Home Learn To...

HITS

Our results

[Safe snorkeling - 10 tips for your safety and comfort](#)

Safe snorkeling tips for your safety and comfort Skip to content Snorkel Around The World Snorkeling Places Snorkel Gear Guide...

[Guide to Keep Kids Safe On and Off the Field](#)

Guide to Keep Kids Safe On and Off the Field Skip to main content Sign In Support My Teams Mobile...

[Simple Tips for Player Recognition | SportsEngine](#)

Simple Tips for Player Recognition SportsEngine Skip to main content Sign In Support My Teams Mobile App Training Camp What...

[Guest Posts and Link Requests](#)

Guest Posts and Link Requests Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...

[10 Golden Rules of Water Safety & Drowning Prevention - Swim Jim](#)

Golden Rules of Water Safety Drowning Prevention Swim Jim New York Texas Toggle navigation SwimJim Swim Levels SwimBaby Programs Learn...

In this query, HITS has ranked the results better than PageRank when looking at the first 10 urls. But PageRank seemed to rank a webpage close enough to Google results. (Redcross)

*Indexing and Relevance*  
Sangeetha Pradeep (sxp210004)

Query: Swimming competitions  
Page Ranking

Our results

[British Swimming Results | British Swimming](#)  
British Swimming Results British Swimming Events Results More Search term News Swimming News Para Swimming News Diving News Artistic Swimming...

[British Diving Results | British Swimming](#)  
British Diving Results British Swimming Events Results More Search term News Swimming News Para Swimming News Diving News Artistic Swimming...

[Para-Swimming Selection | Para-Swimming | British Swimming](#)  
Para Swimming Selection Para Swimming British Swimming Events Results More Search term News Swimming News Para Swimming News Diving News...

[British Para-Swimming Results | British Swimming](#)  
British Para Swimming Results British Swimming Events Results More Search term News Swimming News Para Swimming News Diving News Artistic...

[Me and the Water | Swim Ireland](#)  
Me and the Water Swim Ireland Membership Membership Overview Membership Options Clubmark Club Resources Get Swimming Get Swimming Overview Pop...

## HITS

Our results

[Open water swimming - Wikipedia](#)  
Open water swimming Wikipedia Jump to content Main menu Main menu move to sidebar hide Navigation Main page Contents Current...

[Mariam Sheikhalizadeh - Wikipedia](#)  
Mariam Sheikhalizadeh Wikipedia Open main menu Home Random Nearby Log in Settings Donate About Wikipedia Disclaimers Search Mariam Sheikhalizadeh Article...

[World Aquatics - Wikipedia](#)  
World Aquatics Wikipedia Open main menu Home Random Nearby Log in Settings Donate About Wikipedia Disclaimers Search World Aquatics Article...

[Medley swimming - Wikipedia](#)  
Medley swimming Wikipedia Open main menu Home Random Nearby Log in Settings Donate About Wikipedia Disclaimers Search Medley swimming Article...

[Japan at the 1932 Summer Olympics – Wikipedia](#)  
Japan at the Summer Olympics Wikipedia Open main menu Home Random Nearby Log in Settings Donate About Wikipedia Disclaimers Search...

*Indexing and Relevance*  
Sangeetha Pradeep (sxp210004)

In this query, PageRank retrieved the British swimming webpage while HITS retrieved wikipedia pages of Open water Swimming. HITS seemed to retrieve a more popular page. But both the models retrieved pages which are slightly off the pages that were retrieved by Google or Bing.

Collaboration with clustering

- Clustering required the entire data set from solr to perform offline computations. In order to facilitate this, I provided a json dump of the data in the solr.
- The solr was hosted so that the indexed results were accessible across devices from any location.
- The results of the indexing were used by clustering and in most of the cases it was noted that clustering provided better results than Page Ranking and HITS.

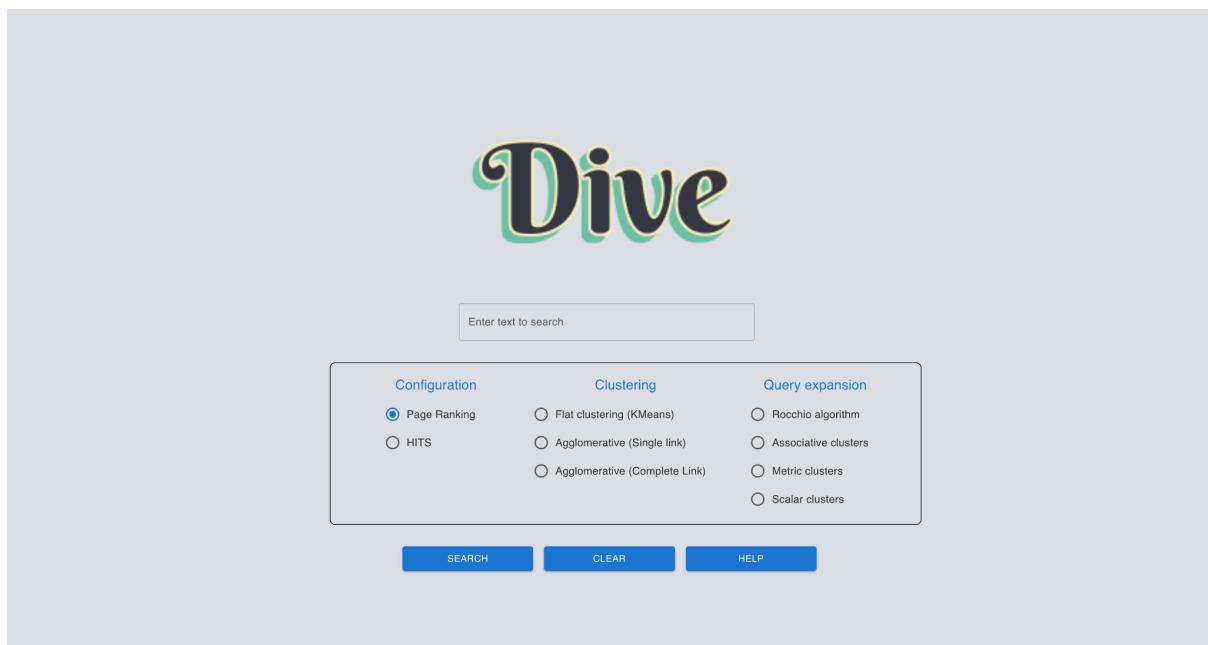
## 4. User interface and comparisons with Google and Bing:

For designing the interface, we used **Reactjs** framework for the frontend web development and **python** for the backend server. The Flask framework was used as the server framework for python. There are two main sections in the webpage.

1. Search section
2. Results section

The results section is hidden until someone searches for a query.

### Search Section:



In the Search section, there is a search bar and three buttons, **search**, **clear** and **help**. The search button initiates the search request and returns the results. The clear button clears the search result that has been done previously. The help button displays the set of instructions for using the website. The search bar is where you input the query text.

There is also a configuration box where the user can set preferences for his search. These are categorised into 3 sections.

1. Relevance model with two options:
  - a. **Page Ranking**
  - b. **HITS**.
2. Clustering models that the engine should use. There are three clustering models supported:
  - a. **Flat clustering (KMeans)**
  - b. **Agglomerative Single link**

- c. Agglomerative Complete Link.
3. Query expansion techniques that can be used as a part of the search.
    - a. Rocchio algorithm
    - b. Associative clusters
    - c. Metric clusters
    - d. Scalar clusters

## Results Section

The Results Section consists of three sub sections:

1. Our Results: The results produced by our search engine
2. Google Results: The results produced by google search engine
3. Bing Results: The results produced by bing search engine

The below images depict a sample result.

The image shows a side-by-side comparison of search results for the query "Swimming workouts for beginners".

**Our results:** This section displays four cards:

- Top 10 swimming strokes ideas and inspiration**: A Pinterest board link.
- Cardio Workouts for Absolute Beginners**: A link to Verywell Fit's homepage.
- Muscular Endurance: How to Measure and Improve It**: A link to Verywell Fit's homepage.
- Fasting and Fasted Training: a guide for female triathletes - Nutrition - TRI247**: A link to Triathlon News Elite.

**Google results:** This section shows the Google search interface with the query "Swimming workouts for beginners" entered. It lists several results:

- Byrdie**: A link to Byrdie.com featuring "7 Trainer-Approved Swim Workouts for Beginners".
- United States Masters Swimming**: A link to USMS.org featuring "Best Swimming Workouts for Beginners".
- Women's Health**: A link to Women's Health magazine featuring "16 Swim Workouts For Beginning, Intermediate, Advanced ...".

Below the search results, there is a "People also ask" section with the question "How many laps should I swim for a good workout?".

*User interface and comparisons with Google and Bing*  
Abishek Kumar (axk200020)

How have you worked with the student that has generated the index?  
How have you accessed the relevance models to provide the results in your user interface?

I worked with Sangeetha who generated the index. We worked together to make the indexing a python module and exported a function which acts as an entrypoint for query search. This entrypoint function accepts arguments such as the query and the relevance model to determine appropriate output. The respective relevance models were called based on the preference by the user. These data were transported using HTTP/TCP layer.

Elaborate on the number of queries you have used for testing the search engine. How many were used in collaboration with the student that built the relevance models and How many did you generate on your own?

Numerous queries were used for testing this search engine. Below are the effective ones that were useful for testing.

**Collaborated with student that built the relevance model:**

- Swimming pools near me
- Benefits of swimming for weight loss
- Swimming workouts for beginners
- Swimming stroke techniques

*User interface and comparisons with Google and Bing*  
Abishek Kumar (axk200020)

- Swimming safety tips
- How to improve swimming technique
- Swimming gear for beginners
- Swimming competitions in Dallas
- Swimming drills for freestyle
- Swimming for fitness: tips and tricks
- Swimming techniques

**Generated on my own:**

- Benefits of swimming
- Competitive swimming events
- Swimming pools near me
- Swimming for weight loss
- Swimming safety tips
- Swimming for beginners
- Swimming workouts
- Olympic swimmers
- Swimwear trends

These queries were selected to test various aspects of the search engine. There were 2 relevance options with 3 clustering models and 4 query expansion techniques. The above queries were chosen to counter cover every possibility.

How did you collaborate with the student that produced clusters to use the clustering information for relevance and presentation on the interface?

I collaborated with Lasya to follow similar steps that followed in integrating the generated index. in creating a clustering module. We worked on exporting the entrypoint for this module with a function signature that supported input parameters as arguments. This setup was similar to how it was explained in indexing.

How do you think your search engine compares to Google and Bing; explain your judgments.

**Relevance of search results:** This is a measure of how well the search engine matches the user's query with relevant search results. Google and Bing are known for their strong relevance algorithms that consider a wide range of factors, such as user behavior, link analysis, and content quality, to produce highly relevant results. Since our search engine is limited to the use of relevant models, It lies below Google and Bing.

*User interface and comparisons with Google and Bing*  
Abishek Kumar (axk200020)

**Speed of search results:** This measures how quickly the search engine returns the search results. Google and Bing are known for their fast response times due to their large-scale infrastructure and use of sophisticated caching mechanisms. Our search engine currently is running locally with a local hosted database. This specific setup performs better in terms of response times.

**User experience:** This measures the ease of use and user-friendliness of the search engine. Google and Bing have invested heavily in user experience, providing users with an intuitive search interface and a range of search features, such as image search, video search, and voice search. Our search engine currently only supports word search but is user friendly with flexible and clickable configurations.

**Privacy and security:** This measures the extent to which the search engine protects user privacy and secures user data. Google and Bing have privacy policies in place and use encryption to protect user data. Our current setup does not have TLS or any sort of encryption to protect the website.

If your search engine had a clustering component, how did you use the results of clustering in presenting the results of your search engine in the user interface.

Our Search engine presents different options for selecting the clustering model for the search query. This search option and query is sent to the backend service where results are generated by the clustering component. They are computed by solr server and groups it into precomputed clusters.

Discuss how you have decided to select the queries for the demonstration of your search engine

Since the demonstration is meant to showcase the accuracy and effectiveness of the search engine, it is important to select queries that are representative of the types of searches that users might perform on the search engine. The queries should cover a range of topics and be relevant to the intended audience of the search engine.

Accuracy

To demonstrate accuracy, the search engine should return relevant and accurate results for the selected queries. These are the queries that found to produce relevant and accurate results:

*User interface and comparisons with Google and Bing*  
Abishek Kumar (axk200020)

## Effectiveness

To demonstrate effectiveness, the search engine should return results quickly and efficiently, with minimal delay or downtime. These are the queries that found to produce effective results with less load:

Provide three examples of the queries and the results produced by your search engine, as well as the results of Google and Bing

These are the queries that produced some common results between google and bing and our search engine. Below are the attached screenshots.

Query1:

**"How to improve swimming technique"**

Our result

The screenshot shows a search results page with two main results. The first result is titled "The Stairs Are Too Big | GoSwim TV" and includes a snippet: "The Stairs Are Too Big GoSwim TV GoSwim TV Home Articles News Search Home Articles News Search The Stairs Are...". The second result is titled "Freestyle Swimming – 10 Tips to Improve Your Technique" and includes a snippet: "Freestyle Swimming Tips to Improve Your Technique Skip to Content Search Magnifying Glass Search for Close Search Home Learn To...". The background of the page has a light beige color with dark grey vertical bars on the sides.

User interface and comparisons with Google and Bing  
Abishek Kumar (axk200020)

Google

How To Improve Swimming Speed - 11 Simple Tips

Enjoy Swimming  
https://www.enjoy-swimming.com › Blog

**Freestyle Swimming – 10 Tips to Improve Your Technique**

Sep 19, 2022 – 1. Use a Neutral Head Position · 2. Press Your Buoy · 3. Do Not Lift Your Head to Breathe · 4. **Swim** on Your Sides · 5. Exhale in the Water · 6. Use a ...  
The Flutter Kick · Head-Lead Prone Balance Drill · The Freestyle Stroke: Overview

Runner's World  
https://www.runnersworld.com › Training › Triathlon

**Seven Steps To Better Swimming Technique**

Nov 23, 2009 – 1. **Swim** tall · 2. Drop an anchor · 3. Heavy rotation · 4. Keep your head down · 5.

Bing

https://www.enjoy-swimming.com/freestyle-swimming-technique.html  
**Freestyle Swimming – 10 Tips to Improve Your Technique**

more items

People also ask

<b>How to become a better swimmer in freestyle swimming?</b> In freestyle swimming, the proper technique is crucial for success. A good swimming technique allows you to either swim at a moderate pace in a relaxed way or to swim at a fast... <a href="http://www.enjoy-swimming.com/freestyle-s...">www.enjoy-swimming.com/freestyle-s...</a> <a href="#">Freestyle Swimming – 10 Tips to I...</a>	<b>What are the first steps in swimming?</b> One of the first steps is <b>learning how to do freestyle</b> . "The front crawl, or freestyle, is one of the first strokes you learn as a competitive swimmer," Marc Christian, Division I swim... <a href="http://www.dickssportinggoods.com/protips/...">www.dickssportinggoods.com/protips/...</a> <a href="#">Breaking Down the Freestyle Strok...</a>
---	---

V  
tl  
lr  
th  
fr  
fr  
h  
w  
F

Feedback

**Videos of Freestyle Swimming 10 steps Enjoyswimming**  
[bing.com/videos](http://bing.com/videos)

Query2:

**"Swimming for weight loss"**

User interface and comparisons with Google and Bing  
Abishek Kumar (axk200020)

## Our result

Running Couch to K K...

[\*\*12 Benefits of Swimming: Weight Loss, Health, and More\*\*](#)

Benefits of Swimming Weight Loss Health and More Health Conditions Featured Breast Cancer IBD Migraine Multiple Sclerosis MS Rheumatoid Arthritis...

[\*\*All Things Swimming & Water Safety Blog | Sunsational Swim School\*\*](#)

All Things Swimming Water Safety Blog Sunsational Swim School Toggle navigation Toggle navigation Sunsational Swim School LOGIN Sunsational Swim School...

## Google

hour three times a week, you'd burn a pound off in two weeks.

 Live Science  
<https://www.livescience.com> › Health › Exercise

**Is swimming good for weight loss?**

Nov 3, 2022 — Those of you who love a dip in the water will be glad to hear that depending on the duration and intensity of your workout, **swimming** can indeed ...

 U.S. News & World Report  
<https://health.usnews.com> › Wellness › Fitness

**Swimming for Weight Loss**

## Bing

See The University of Texas at Dallas results for **Swimming for we...>**

**Is swimming good for weight loss? | Live Science**  
<https://www.livescience.com/is-swimming-good-for-weight-loss> ▾  
Web Nov 2, 2022 · Overall, **swimming** is good for **weight loss** if performed in line with the Physical Activity Guidelines for Americans and alongside a healthy balanced diet. ....  
Author: Gemma Harris

**Swimming Workout For Weight Loss | The Most Popular Swim Shop**  
<https://www.swimoutlet.com/water-aerobics> ▾ 301K+ Facebook followers  
Ad Get Back in Shape with the Right Equipment. Shop Now - Always in Stock! Free Shipping On

Query3:

**“Swimming safety tips”**

User interface and comparisons with Google and Bing

Abishek Kumar (axk200020)

Our result

The screenshot shows a search result for "Water Safety American Red Cross". The title is "Water Safety | American Red Cross". Below the title, there is a navigation bar with links: Water Safety, American Red Cross, Menu, Hi, Account Summary, Donation History, Personal Information, Sign Out, Sign In, Español, and Your... . The main content area contains a heading "Food Safety – Water Quality and Health Council".

Bing

## Swimming Safety | Swim Safe | American Red Cross

<https://www.redcross.org/get-help/how-to-prepare...> ▾

(Web) Take smart **precautions** and **use layers of protection**. Stay within an arm's reach of any weak or inexperienced swimmer who is in the water. Have young children or inexperienced swimmers wear U.S. Coast Guard-approved life jackets around water, but do not rely...

### Beach and Ocean Safety

Swimming Safety Tips. Life Jacket Safety  
Saves Lives. Find a Red Cross Provider. ...

### Sign Up

The Water Safety for Parents and  
Caregivers Online Course focuses on ...

### Water Safety

It only takes a moment. A child or weak  
swimmer can drown in the time it takes ...

Other content from redcross.org

Google

The screenshot shows a search result for "Water Safety". The snippet from "Gateway Region YMCA" includes the text: "Closely supervise children. · Stay hydrated. · Never **swim** alone. Even ...". Below the snippet is a link to "Gateway Region YMCA" with the URL "https://gwrymca.org › blog › water-safety-tips" and a "More" button. The snippet for "Water Safety Tips" includes the date "Jun 14, 2019" and the text: "Water Safety Tips · 1. Never **Swim** Alone · 2. Supervise Children When They're in the Water · 3. Don't Play Breath-Holding Games · 4. Always Wear a ...". Below this is a snippet for "Images for Swimming safety tips" with four thumbnail images: "red cross", "open water", "goldfish swim", and "swimming infographic".

## 5. CLUSTERING

All the documents were crawled by Pallavi (student responsible for crawling) and indexed by Sangeetha (the student responsible for indexing). I got the indexed data generated by Apache Solr from Sangeetha. The query “\*” is used to retrieve all the crawled documents in a JSON format, and the results were dumped to a JSON file. The size of this file was around 2 GB.

### Sample of indexed data:

```
{  
  "responseHeader":{  
    "status":0,  
    "QTime":8,  
    "params":{  
      "q":"*-*",  
      "rows":1166,  
      "wt":"json"}},  
  "response":{ "numFound":1166, "start":0, "numFoundExact":true, "docs": [  
    {  
      "tstamp": "2023-04-18T04:57:08.145Z",  
      "anchor": "[Winter Swimming]",  
      "digest": "312eb7e341c55ac802c298c84e3ecc31",  
      "boost": 0.40240812,  
      "id": "http://ecopsychology.swami-center.org/winter-swimming.shtml",  
      "title": "Winter Swimming",  
      "url": "http://ecopsychology.swami-center.org/winter-swimming.shtml",  
      "version": 1763567567268478976,  
      "content": "Winter Swimming\n \nContemporary knowledge about God, Evolution, and the meaning of  
human life. Methodology of spiritual development.\n \n \nEcopsychology /Winter Swimming\nWinter  
Swimming\nWinter swimming , i.e. swimming in ice-cold water, is a wonderful method not only of tempering  
the body, but also of increasing the energy might of the organism.\nStatistics holds that among those  
practicing winter swimming the sickness rate decreases for cold-type diseases 60 (!) times, and for  
other diseases – 30 times. As a therapeutic method winter swimming can heal many diseases, including  
radiculitis, both high and low blood-pressure, pulmonary tuberculosis, pancreatic diabetes, chronic  
gastrointestinal diseases, inflammations in genitals, menstrual cycle abnormalities, dermatoses, and so  
on [ 32 , 65 ].\nOne should not begin to swim in cold water necessarily in autumn, as some people think:  
one may start right during the hardest frost, if one likes. But it is recommended to master first the  
psychic self-regulation techniques that we described above. In this case, winter swimming will directly  
contribute to strengthening the ability to preserve the subtlest psychic states. Coming out into the  
frost naked and submerging into ice-cold water become not only an act of volition, but also a controlled  
experiment on oneself in order to maintain the subtlest states of the consciousness in the conditionally  
adverse environment. It is beneficial to disidentify oneself from bodily sensations and to merely  
    }  
  ]  
}
```

### DATA PRE-PROCESSING:

- As in the indexed data sample above, response key contains Json value which has list of docs.
- Each doc contains data inside content key.
- Iterated through each doc, extracted the “content” key’s value from the JSON, which has the data in plain text format, removing the DOM and HTML tags.
- The URL key has the URL of the corresponding website.
- Content of all the docs is stored in a list and all corresponding urls in another list.
- scikit-learn’s TfidfVectorizer is used for vectorization of the data.
- Using TruncatedSVD to reduce the dimensionality of TF-IDF document vectors (latent semantic analysis. SVD results are then normalized.

## **CLUSTERING ALGORITHMS:**

### **FLAT CLUSTERING**

Flat clustering does single grouping or partitioning of data. In the project, flat clustering is implemented using KMeans clustering algorithm.

#### **KMeans Clustering:**

Given a set of observations ( $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ), where each observation is a  $d$ -dimensional real vector,  $k$ -means clustering aims to partition the  $n$  observations into  $k$  ( $\leq n$ ) sets  $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$  to minimize the within-cluster sum of squares (WCSS) (i.e., variance). WCSS is the sum of the squares of distances of all the observations belonging to a cluster from the centroid of the cluster. The objective is to find:

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

where  $\boldsymbol{\mu}_i$  is the mean of points in  $S_i$ .

#### **KMeans Clustering Design and Implementation:**

- The preprocessed data is after dimensionality reduction is used for KMeans clustering. Scikit-Learn's KMeans library is used to apply the KMeans clustering algorithm with a predefined value of  $k$ .
- The cluster obtained for a document after applying the KMeans algorithm as well as the URL corresponding to that document were converted to Pandas series.
- The results were stored in a text file (clustering\_f.txt) in (URL, ClusterID) format.

#### **Selection of predefined number of clusters for KMeans (value of k):**

The elbow method and silhouette score are used to select the predefined value of  $k$ .

Performed KMeans clustering for different values of  $k$ , ranging from 4 to 20 on the dataset, and the observed clusters were compared for selecting the value of  $k$ . The Elbow Method is more of a decision rule, while the Silhouette is a metric used for validation while clustering.  $K$  value is chosen for which WSS first starts to diminish and Silhouette score reaches to global maximum. After observing the results testing different values of  $k$ , a value of  $k = 15$  gave the most relevant results. As a result, all the fetched documents are stored in one of the 15 clusters with cluster id ranging from 0 to 14.

*Sai Lasya Akkineni(SXA210025) - Clustering*

### **Results of KMeans Clustering:**

Sample of clusters generated by KMeans from the text file:

```
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/1997/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/1998/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/2000/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/2008/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/2009/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/2010/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/2011/,7
https://longswims.com/events/maraton-acuatica-internacional-ciudad-de-rosario/2013/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1961/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1962/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1963/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1964/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1965/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1966/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1967/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1974/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1987/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1988/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1990/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1991/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1992/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1993/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1994/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1995/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1996/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/1997/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2000/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2001/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2002/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2003/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2004/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2005/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2006/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2007/,7
https://longswims.com/events/maraton-acuatico-rio-coronda/2008/,7
```

As seen in the above picture, all events related to swimming are grouped into cluster 7.

### **HIERARCHIAL AGGLOMERATIVE CLUSTERING:**

Agglomerative clustering with single-link and complete-link are implemented for hierarchical clustering of documents.

### **Agglomerative Clustering:**

It is also known as the bottom-up approach or hierarchical agglomerative clustering (HAC). This algorithm does not require us to prespecify the number of clusters. The bottom-up algorithm treats each data as a singleton cluster at the outset and then successively agglomerates pairs of clusters until all clusters have been merged into a single cluster that contains all data. The results of hierarchical clustering are usually represented by a dendrogram.

For computing distance, there are various distance metrics to choose from: Euclidean, Manhattan, Mahalanobis, etc . For computing the linkage criteria, to compute the distance between different clusters, there are a lot of metrics to choose from: maximum or complete link, minimum or single link, unweighted average link, weighted average link, centroid link, ward etc.

For the project, I have used Euclidean distance metric and performed two types of hierarchical clustering single-link and complete-link.

### **Agglomerative Clustering (Single Link and Complete Link) Design & Implementation:**

- The preprocessed data after dimensionality reduction is used to perform agglomerative clustering.
- Python's fastcluster library is used to compute the results of hierachial clustering efficiently.
- The linkage method of fastcluster library is used to perform hierarchical, agglomerative clustering on vector data.
  - `fastcluster.linkage(X, method='single', metric='euclidean')`
- X is the input data; method can be passed as single or complete and metric is the distance used for calculating similarity.
- fastcluster runs the hierachial clustering and determines the suitable number of clusters for the data.
- Dendrogram is plotted to visualize the cluster results.
- The clusters obtained for a document after applying agglomerative clustering algorithm as well as the URLs corresponding to that document were converted to pandas series.
- The single link hierachial clustering results were stored in a text file (clustering\_hs.txt) and complete link hierachial clustering results in (clustering\_hc.txt) in (URL, ClusterID) format.

Sai Lasya Akkineni(SXA210025) - Clustering

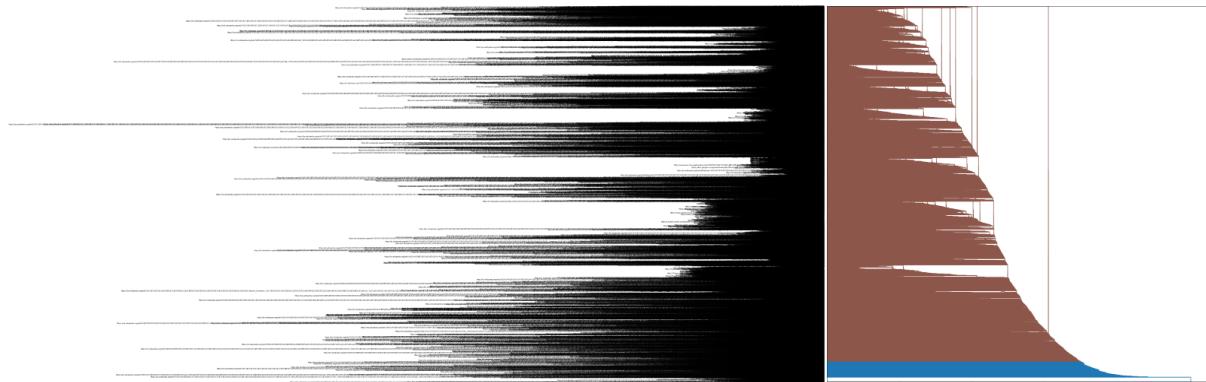
**Single-Link hierarchical clustering results:**

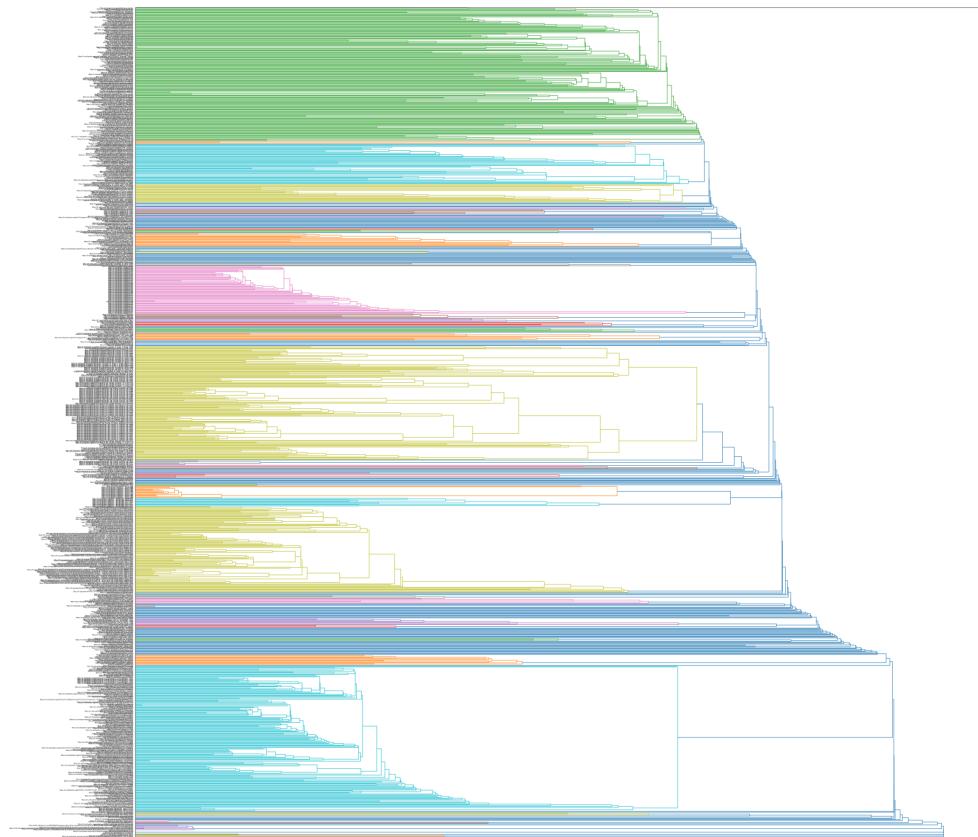
Sample of clusters generated by single link hierachial clustering from the text file:

```
http://aquaskills.com/en/stunts-tv-and-film/schedules,3.0
http://aquaskills.com/en/aquapool-parties,3.0
http://aquaskills.com/en/component/tags/tag/swim-lessons-nyc,3.0
http://aquaskills.com/en/component/tags/tag/adult-swim-classes,3.0
http://aquaskills.com/en/component/tags/tag/best-swim-school,3.0
http://aquaskills.com/en/about-us/aquaskills-locations,3.0
http://aquaskills.com/en/swim-lessons/mini-group-swim-lessons,3.0
http://aquaskills.com/en/videos,3.0
http://aquaskills.com/en/photos,3.0
http://aquaskills.com/en/swim-lessons/mommy-or-daddy-me,3.0
http://aquaskills.com/en/registration,3.0
http://aquaskills.com/en/announcements,3.0
http://aquaskills.com/en/announcements/46-aquaskills-announcements,3.0
```

As observed in the picture, information related to swimming classes, registration, photos etc are grouped into same cluster

Dendrograms for large amounts of data are very cluttered. Below are some of the samples from subset of the dataset.





**Number of clusters observed on entire dataset:**

fastcluster's agglomerative clustering selects the best number of clusters based on input data provided and different distance and linkage metrics used. It was observed that 10 clusters were created from our dataset.

**Complete-Link hierarchical clustering results:**

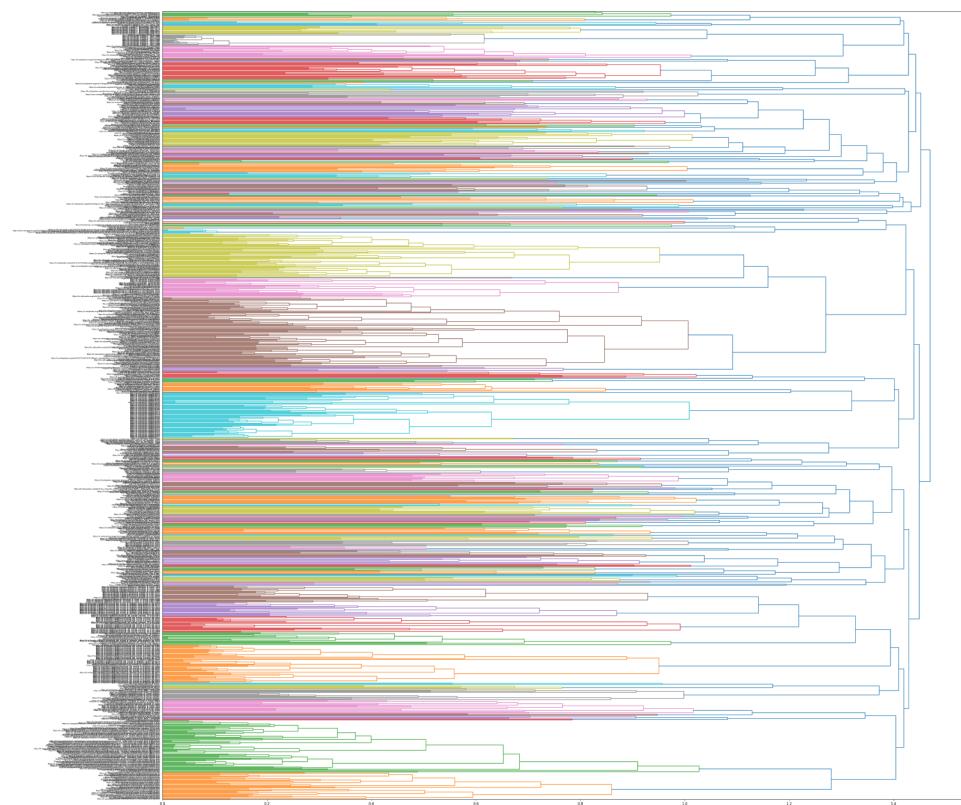
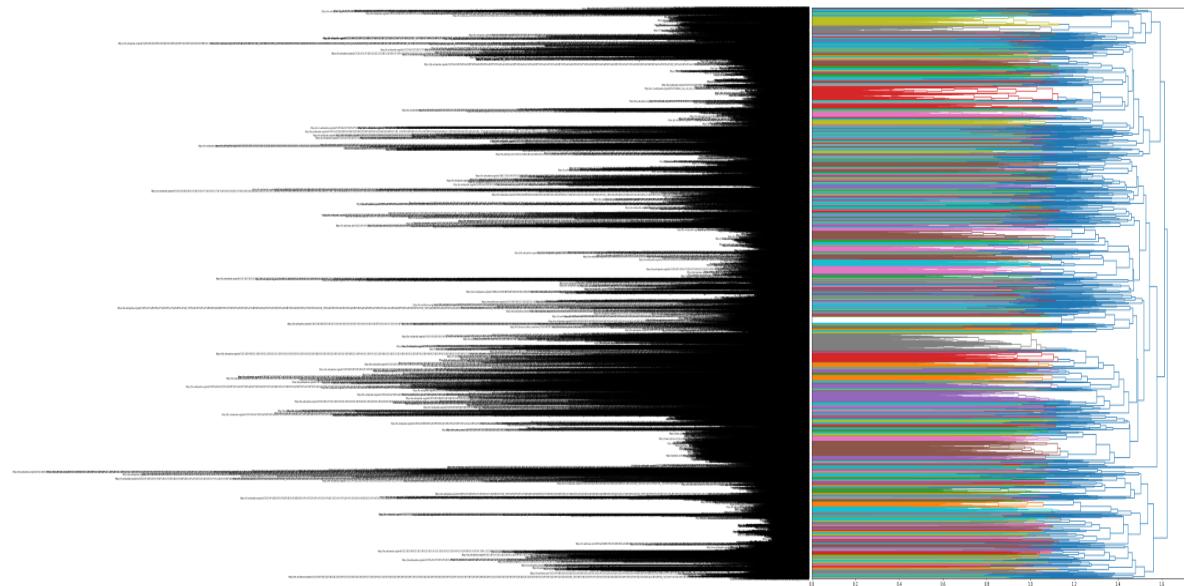
Sample of clusters generated by complete link hierarchical clustering from the text file:

```
https://www.healthline.com/health/exercise-fitness/how-to-swim,10.0
https://www.healthline.com/health/fitness-exercise/swim-workouts-target-your-belly,10.0
https://www.healthy.arkansas.gov/programs-services/topics/swimming-pools-and-related-facilities,10.0
"http://www.msdh.state.ms.us/msdhsite/_static/30,0,95,669.html",10.0
https://www.health.pa.gov/topics/programs/Bathing%20Places/Pages/Bathing%20Places.aspx,10.0
https://www.health.ny.gov/environmental/outdoors/swimming/,10.0
https://www.health.state.mn.us/communities/environment/recreation/pools/index.html,10.0
https://www.cdc.gov/habs/cdc-at-work.html,10.0
https://www.cdc.gov/fluoridation/guidelines/index.htm,10.0
https://www.cdc.gov/fluoridation/faqs/wellwater.htm,10.0
https://www.cdc.gov/fluoridation/index.html,10.0
https://deq.jic.no.gov/ac/bw/,10.0
```

As observed in the above picture, news about health and workout related to swimming are grouped into one cluster.

Dendrograms for large amounts of data are very cluttered. Below are some of the samples from subset of the dataset.

Sai Lasya Akkineni(SXA210025) - Clustering



#### **Number of clusters observed:**

fastcluster's agglomerative clustering selects the best number of clusters based on input data provided and different distance and linkage metrics used. 10 clusters were created from our dataset.

#### **INTEGRATION WITH UI**

#### **Flat and Hierarchical (Single-link and Complete-link) clustering results incorporation and integration with User Interface:**

- On the UI, there are three clustering modes: one for flat clustering (Flat clustering (KMeans)) and two for hierarchical clustering (single link and complete link) - (Agglomerative (Single link), Agglomerative (Complete Link)).
- If any of the clustering modes is selected, a GET request is sent to the backend with the corresponding clustering mode as the request type parameter.
- On the backend, if the request type contains clustering 40 relevant results are fetched from Solr index based on page ranking and on top of that clustering is applied.
- As part of clustering, based on clustering type corresponding precomputed clusters are fetched from the text file and loaded into a map. The Map contains key as the url and clusterId as the value.
- Using the precomputed clusters, the results are prioritized in such a way that cluster to which top-ranked result belongs to, all the other remaining results belonging to the same cluster in the order of decreasing page rank are ranked next.
- This process repeats until we process all the 40 relevant results.
- The updated results ranked based on the clustering are sent as response to the front-end. Collaborated with Abhishek on this part.

#### **OPTIMIZATION WITH CLUSTERING**

##### **Query selection and impact on results and relevancy:**

The criterion for selecting the queries is obtained from the type of clusters formed for all the types of clustering methods used. I manually created the queries depending on the results obtained from the relevance models, page rank, and HITS. As observed in the cluster samples above, one of the selection criteria is events related to swimming. Another selection criteria is swimming classes and news about health related to swimming.

I used 50 queries for each type of clustering: flat and hierarchical (single-link and complete-link) to test the impact of the clustering on the results. After clustering is applied, similar results among the relevant results are grouped together and shown in the user interface. The non-relevant or dissimilar results get deprioritized.

##### **Number of queries experimented with to improve results of clustering:**

Around 50 queries are used for each type of clustering to improve the results. Some of the query results are listed below. Screenshots that show the impact of clustering on improving the relevancy of the retrieved results are also provided below.

Sai Lasya Akkineni(SXA210025) - Clustering

## EXAMPLES OF QUERIES WITH CLUSTERING:

Around 50 queries are used for testing flat clustering and both types of hierarchical clustering. Attached are five of the queries I used. The first image shows results after page ranking, the second image shows results after flat clustering, the third image shows results after single link hierarchical clustering, and the last image about complete link hierarchical clustering.

Query 1: Benefits of swimming

### PageRanking

The screenshot displays a search interface with the following components:

- Search Bar:** "Enter text to search" with the query "Benefits of swimming".
- Configuration Panel:** Contains three sections:
  - Configuration:** Radio buttons for "Page Ranking" (selected), "HITS", "Clustering", "Query expansion".
  - Clustering:** Radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", "Agglomerative (Complete Link)".
  - Query expansion:** Radio buttons for "Rocchio algorithm", "Associative clusters", "Metric clusters", "Scalar clusters".
- Buttons:** "SEARCH", "CLEAR", "HELP".
- Results Section:** Titled "Results".
  - Our results:** A list of search results:
    - [The Health Benefits of Swimming](#)
    - [7 Health Benefits of Swimming](#)
    - [10 Benefits of Swim Lessons for Kids | Big Blue Swim School](#)
    - [Back to School: Mental Benefits of Swimming for Kids](#)
    - [Swimming, Health, and Safety](#)
  - Google results:** A screenshot of a Google search results page for "Benefits of swimming". It shows the search bar, filters (Images, Videos, etc.), and a snippet of the first result from "Everyday Health" about swimming being a full-body workout.

# Sai Lasya Akkineni(SXA210025) - Clustering

## Flat clustering

The screenshot displays a user interface for a flat clustering application. At the top, there is a search bar with the placeholder "Enter text to search" and the query "Benefits of swimming". Below the search bar are three main configuration sections: Configuration, Clustering, and Query expansion.

- Configuration:** Contains two radio buttons: "Page Ranking" (selected) and "HITS".
- Clustering:** Contains three radio buttons: "Flat clustering (KMeans)" (selected), "Agglomerative (Single link)", and "Agglomerative (Complete Link)".
- Query expansion:** Contains five radio buttons: "Rocchio algorithm", "Associative clusters", "Metric clusters", and "Scalar clusters".

At the bottom of the configuration area are three buttons: "SEARCH", "CLEAR", and "HELP".

The main interface is titled "Results" and is divided into two sections: "Our results" and "Google results".

- Our results:** A list of four search results:
  - The Health Benefits of Swimming**: Includes links to "Skip to Content", "Search", "Magnifying Glass", "Search for Close", "Search Home", "Learn To Swim", and "pr...".
  - 7 Health Benefits of Swimming**: Includes links to "Health Benefits of Swimming", "Menu", "Visit Verywell Fit's homepage", "Health Benefits of Swimming", "Fitness", "Beginners", "Walking", "Cardio", "Strength", and "Running...".
  - Back to School: Mental Benefits of Swimming for Kids**: Includes links to "Back to School", "Mental Benefits of Swimming for Kids", "Big Blue Blog", "Kids Swimming Lessons", and "A Back To School Boost...".
  - Swimming, Health, and Safety**: Includes links to "Swimming", "Health", "Safety", "Skip to Content", "Search", "Magnifying Glass", "Search for Close", "Search Home", "Learn To Swim", and "pr Basic...".
- Google results:** A screenshot of a Google search results page for the query "Benefits of swimming".
  - Search bar: "Benefits of swimming".
  - Filter buttons: "Images", "Videos", "+ For adults", "+ Everyday", "+ 10", "+ Essay", "+ 5", "+ More".
  - Text: "About 572,000,000 results (0.47 seconds)".
  - Result 1: "Everyday Health" - <https://www.everydayhealth.com/fitness/health-ben...>.  
Summary: "Jul 11, 2022 – Swimming Can Help Keep You Lean ... Swimming is a full-body workout. Swimming engages everything from the arms, shoulders, and legs to the core, ...".
  - People also ask (dropdown):
    - "What are the 5 benefits of swimming?"
    - "Does swimming burn belly fat?"
    - "How many times a week should I swim to tone up?"
    - "Will swimming 3 times a week tone me up?"
  - Result 2: "better health.vic.gov.au" - <https://www.betterhealth.vic.gov.au/healthyliving/s...>.  
Summary: "Other benefits of swimming - being a relaxing and peaceful form of exercise - alleviating stress - improving coordination, balance and posture - improving ...".
  - Text: "Health benefits of swimming - Other benefits of swimming".

## Sai Lasya Akkineni(SXA210025) - Clustering Agglomerative (Single Link)

The screenshot displays a user interface for a clustering application. At the top, there is a search bar containing the text "Benefits of swimming". Below the search bar are three main sections: "Configuration", "Clustering", and "Query expansion".

- Configuration:** Contains radio buttons for "Page Ranking" (selected), "HITS", "Flat clustering (KMeans)", "Agglomerative (Single link)" (selected), "Agglomerative (Complete Link)", "Rocchio algorithm", "Associative clusters", "Metric clusters", and "Scalar clusters".
- Clustering:** Buttons for "SEARCH", "CLEAR", and "HELP".
- Query expansion:** Buttons for "Images", "Videos", "For adults", "Everyday", "+ 10", "Essay", "+ 5", and "Ment".

The main area is titled "Results" and is divided into two sections: "Our results" and "Google results".

- Our results:** A grid of five cards:
  - The Health Benefits of Swimming**: The Health Benefits of Swimming Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...
  - 10 Benefits of Swim Lessons for Kids | Big Blue Swim School**: Benefits of Swim Lessons for Kids Big Blue Swim School Big Blue Blog Kids Swimming Lessons Benefits Of Swim Lessons...
  - Back to School: Mental Benefits of Swimming for Kids**: Back to School Mental Benefits of Swimming for Kids Big Blue Blog Kids Swimming Lessons A Back To School Boost...
  - Swimming\_Health\_and\_Safety**: Swimming Health and Safety Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr Basic...
  - Nadia Harrald - Big Blue Blog | Kids Swimming Lessons**: Nadia Harrald Big Blue Blog Kids Swimming Lessons Nadia Harrald posts goalgetter Be a Goal Getter Steps to Better Habit...
- Google results:** A search results page for "Benefits of swimming" on Google. It shows the following information:
  - About 572,000,000 results (0.47 seconds)
  - Top result: **Everyday Health** (<https://www.everydayhealth.com/fitness/health-ben...>)  
9 Health Benefits of Swimming  
Jul 11, 2022 – Swimming Can Help Keep You Lean ... Swimming is a full-body workout.  
\*Swimming engages everything from the arms, shoulders, and legs to the core, ...
  - People also ask:
    - What are the 5 benefits of swimming?
    - Does swimming burn belly fat?
    - How many times a week should I swim to tone up?
    - Will swimming 3 times a week tone me up?
  - Other results:
    - better health.vic.gov.au (<https://www.betterhealth.vic.gov.au/healthyiving/s...>)  
Swimming - health benefits  
Other benefits of swimming - being a relaxing and peaceful form of exercise - alleviating stress - improving coordination, balance and posture - improving ...  
Health benefits of swimming - Other benefits of swimming
    - Healthline (<https://www.healthline.com/Wellness Topics/Fitness>)

## Sai Lasya Akkineni(SXA210025) - Clustering Agglomerative (Complete Link)

The screenshot displays a web-based clustering application interface. At the top, there is a search bar containing the text "Benefits of swimming". Below the search bar are three main configuration sections: "Configuration", "Clustering", and "Query expansion".

- Configuration:** Contains radio buttons for "Page Ranking" (selected), "HITS", and "Agglomerative (Complete Link)".
- Clustering:** Contains radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", and "Agglomerative (Complete Link)" (selected).
- Query expansion:** Contains radio buttons for "Rocchio algorithm", "Associative clusters", "Metric clusters", and "Scalar clusters".

Below these sections are three buttons: "SEARCH", "CLEAR", and "HELP".

The main area is titled "Results" and is divided into two sections: "Our results" and "Google results".

- Our results:** A list of five search results:
  - The Health Benefits of Swimming**: Includes links to "Skip to Content", "Search", "Magnifying Glass", "Search for Close", "Search Home", "Learn To Swim pr...", and "...".
  - Swimming, Health, and Safety**: Includes links to "Swimming Health and Safety", "Skip to Content", "Search", "Magnifying Glass", "Search for Close", "Search Home", "Learn To Swim pr Basic...", and "...".
  - The Psychological Benefits of Swimming**: Includes links to "The Psychological Benefits of Swimming", "Skip to Content", "Search", "Magnifying Glass", "Search for Close", "Search Home", "Learn To Swim pr...", and "...".
  - The Importance Of Swimming Skills For Safety, Health and Enjoyment**: Includes links to "The Importance Of Swimming Skills For Safety Health and Enjoyment", "Skip to Content", "Search", "Magnifying Glass", "Search for Close", "Search Home", "Learn To Swim pr...", and "...".
  - 7 Health Benefits of Swimming**: Includes links to "Health Benefits of Swimming", "Menu Visit Verywell Fit's homepage", "Health Benefits of Swimming", "Fitness", "Beginners", "Walking", "Cardio", "Strength", "Running...", and "...".
- Google results:** A screenshot of a Google search results page for "Benefits of swimming".
  - Header: "Google Benefits of swimming" with search filters: "Images", "Videos", "For adults", "Everyday", "10", "Essay", "5", and "Ment".
  - Text: "About 572,000,000 results (0.47 seconds)".
  - Result 1: "Everyday Health" (<https://www.everydayhealth.com/fitness/health-benefits-of-swimming/>)
    - Summary: "9 Health Benefits of Swimming Jul 11, 2022 - Swimming Can Help Keep You Lean ... Swimming is a full-body workout. \*Swimming engages everything from the arms, shoulders, and legs to the core. ..."
  - Related questions (People also ask):
    - "What are the 5 benefits of swimming?"
    - "Does swimming burn belly fat?"
    - "How many times a week should I swim to tone up?"
    - "Will swimming 3 times a week tone me up?"
  - Other results:
    - "better health.vic.gov.au" (<https://www.betterhealth.vic.gov.au/healthy-living/s...>)
    - "Swimming - health benefits" (Summary: "One of the best forms of exercise - a relaxing and peaceful form of exercise - alleviating stress - improving coordination, balance and posture - improving ...")
    - "Health benefits of swimming | Other benefits of swimming" (Summary: "Healthline")

*Sai Lasya Akkineni(SXA210025) - Clustering*  
Query 2: Swimming for beginners

PageRanking

Enter text to search  
**Swimming for beginners**

**Configuration**  
 Page Ranking  
 HITS

**Clustering**  
 Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**  
 Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

**Results**

**Our results**

- The Ultimate Swimming Guide For Beginners - Human Race**  
The Ultimate Swimming Guide For Beginners Human Race Close Events All Events Results Support Shop Preparation Training Advice Training Events...
- Triathlon swim training advice, plans & research | Tri247.com**  
Triathlon swim training advice plans research Tri com Skip to content Menu News Triathlon News Elite Age Group Paratriathlon Events...
- TRI247 - Triathlon news & events plus training, gear & nutrition tips**  
TRI Triathlon news events plus training gear nutrition tips Skip to content Menu News Triathlon News Elite Age Group Paratriathlon...
- Cold Water Swimming - TRI247**  
Cold Water Swimming TRI Skip to content Menu News Triathlon News Elite Age Group Paratriathlon Events Events Calendar All Events...
- Top 10 swimming strokes ideas and inspiration**  
Top swimming strokes ideas and inspiration Pinterest Today Watch Shop Explore When autocomplete results are available use up and down...

**Google results**

Google Swimming for beginners

About 58,300,000 results (0.41 seconds)

**Beginner SWIM TIPS**

... Here we go on my side left arm is out I'm scrolling Underwater Now we're gonna show you a ... More

YouTube https://www.youtube.com/FarrKrebs Beginner Swim Tips For Adults - YouTube

Everyday Health https://www.everydayhealth.com/fitness/how-to... About featured snippets Feedback

# Sai Lasya Akkineni(SXA210025) - Clustering

## Flat clustering

Enter text to search  
Swimming for beginners

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

SEARCH    CLEAR    HELP

### Results

**Our results**

[The Ultimate Swimming Guide For Beginners - Human Race](#)  
The Ultimate Swimming Guide For Beginners Human Race Close Events All Events Results Support Shop Preparation Training Advice Training Events...

[Cold Water Swimming - TRI247](#)  
Cold Water Swimming TRI Skip to content Menu News Triathlon News Elite Age Group Paratriathlon Events Events Calendar All Events...

[Top 10 swimming strokes ideas and inspiration](#)  
Top swimming strokes ideas and inspiration Pinterest Today Watch Shop Explore When autocomplete results are available use up and down...

[Cold Water Swimming Locations In The UK - TRI247](#)  
Cold Water Swimming Locations In The UK TRI Skip to content Menu News Triathlon News Elite Age Group Paratriathlon Events...

[4 Basic Swimming Strokes Every Beginner Should Learn](#)  
Basic Swimming Strokes Every Beginner Should Learn Skip to Content Search Magnifying Glass Search for Close Search Home Learn To...

**Google results**

Google    Swimming for beginners    About 58,300,000 results (0.41 seconds)

[Google Videos Images Shopping Adults Near me Lap Freestyle](#)



\*... Here we go on my side left arm is out I'm scrolling. Underwater. Now we're gonna show you a ... More

YouTube    https://www.youtube.com /Fares Ksebat    Beginner Swim Tips For Adults - YouTube

About featured snippets · Feedback

Everyday Health    https://www.everydayhealth.com /fitness /how-to-pe...

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Single Link)

Enter text to search  
Swimming for beginners

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

**Our results**

[The Ultimate Swimming Guide For Beginners - Human Race](#)  
The Ultimate Swimming Guide For Beginners Human Race Close Events All Events Results  
Support Shop Preparation Training Advice Training Events...

[4 Basic Swimming Strokes Every Beginner Should Learn](#)  
Basic Swimming Strokes Every Beginner Should Learn Skip to Content Search Magnifying Glass  
Search for Close Search Home Learn To...

[Swimming Tips for Beginners and Experienced Swimmers](#)  
Swimming Tips for Beginners and Experienced Swimmers Skip to Content Search Magnifying Glass  
Search for Close Search Home Learn To...

[Outdoor Swimmer's Weekly Newsletter - Outdoor Swimmer Magazine](#)  
Outdoor Swimmer's Weekly Newsletter Outdoor Swimmer Magazine Outdoor Swimmer Extra You can now read the best content from Outdoor...

[Comments on: Swimming Tips for Beginners and Experienced Swimmers](#)  
Comments on Swimming Tips for Beginners and Experienced Swimmers Comments on Swimming

**Google results**

Google   Swimming for beginners   X   Videos   Images   Shopping   Adults   Near me   Lap   Freestyle

About 58,300,000 results (0.41 seconds)



YouTube   https://www.youtube.com - Fares Kiebati   Beginner Swim Tips For Adults - YouTube

Everyday Health   https://www.everydayhealth.com - fitness - how-to-go...   About featured snippets   Feedback

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Complete Link)

Enter text to search  
Swimming for beginners

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

**Our results**

- [The Ultimate Swimming Guide For Beginners - Human Race](#)  
The Ultimate Swimming Guide For Beginners Human Race Close Events All Events Results Support Shop Preparation Training Advice Training Events...
- [Swimming Tips for Beginners and Experienced Swimmers](#)  
Swimming Tips for Beginners and Experienced Swimmers Skip to Content Search Magnifying Glass Search for Close Search Home Learn To...
- [Triathlon swim training advice, plans & research | Tri247.com](#)  
Triathlon swim training advice plans research Tri.com Skip to content Menu News Triathlon News Elite Age Group Paratriathlon Events...
- [TRI247 - Triathlon news & events plus training, gear & nutrition tips](#)  
TRI Triathlon news events plus training gear nutrition tips Skip to content Menu News Triathlon News Elite Age Group Paratriathlon...
- [Cold Water Swimming - TRI247](#)  
Cold Water Swimming TRI Skip to content Menu News Triathlon News Elite Age Group Paratriathlon Events Events Calendar All Events...

**Google results**

Google   Swimming for beginners   X     

About 58,300,000 results (0.41 seconds)

  
... Here we go on my side left arm is out I'm scrolling. Underwater. Now we're gonna show you a ... More

YouTube   https://www.youtube.com /Fares Kiebaï

Beginner Swim Tips For Adults - YouTube

About featured snippets · Feedback

Everyday Health   https://www.everydayhealth.com /fitness /how-to-get...  
A Beginner's Guide to Get Started With Swimming  
Jul 5, 2022 – Safety Tips for Beginners - Don't swim alone. No matter your experience or fitness

## Sai Lasya Akkineni(SXA210025) - Clustering

### Query 3: How Can I Swim Without Getting Tired PageRanking

Enter text to search  
**How Can I Swim Without Getting Tired**

Configuration	Clustering	Query expansion
<input checked="" type="radio"/> Page Ranking	<input type="radio"/> Flat clustering (KMeans)	<input type="radio"/> Rocchio algorithm
<input type="radio"/> HITS	<input type="radio"/> Agglomerative (Single link)	<input type="radio"/> Associative clusters
	<input type="radio"/> Agglomerative (Complete Link)	<input type="radio"/> Metric clusters
		<input type="radio"/> Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

**Our results**

- Comments on: Learn How To Swim Easily**  
Comments on Learn How To Swim Easily Comments on Learn How To Swim Easily Learning to Swim and Having Fun...
- Comments on: Front Crawl / Freestyle Swimming Technique: Breathing**  
Comments on Front Crawl Freestyle Swimming Technique Breathing Comments on Front Crawl Freestyle Swimming Technique Breathing Learning to Swim and...
- Comments on: Learn How to Tread Water – Techniques and Exercises**  
Comments on Learn How to Tread Water Techniques and Exercises Comments on Learn How to Tread Water Techniques and Exercises...
- The 3 Stages of Proper Swim Development | TrainingPeaks**  
The Stages of Proper Swim Development TrainingPeaks The Stages of Proper Swim Development TrainingPeaks Training Articles Subscribe Training Articles Coach...
- Butterfly Stroke Drills - One Armed Butterfly**  
Butterfly Stroke Drills One Armed Butterfly Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim...

**Google results**

Google   How Can I Swim Without Getting Tired   About 30,700,000 results (0.43 seconds)

YouTube   https://www.youtube.com · Global Triathlon Network · How To Swim Without Getting Tired - YouTube



... And less efficient. So some mobility work could be in order you can do this on dry land. And also ... More

About featured snippets · Feedback

People also ask :

Why do I get tired so fast while swimming?

## Sai Lasya Akkineni(SXA210025) - Clustering

### Flat Clustering

Enter text to search  
How Can I Swim Without Getting Tired

**Configuration**  
 Page Ranking  
 HITS

**Clustering**  
 Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**  
 Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH** **CLEAR** **HELP**

### Results

**Our results**

- [Comments on: Learn How To Swim Easily](#)  
Comments on Learn How To Swim Easily Comments on Learn How To Swim Easily Learning to Swim and Having Fun...
- [The 3 Stages of Proper Swim Development | TrainingPeaks](#)  
The Stages of Proper Swim Development TrainingPeaks The Stages of Proper Swim Development TrainingPeaks Training Articles Subscribe Training Articles Coach...
- [Mr. Brojen Das \( King Of Channel \)](#)  
Mr Brojen Das King Of Channel Conquered The English Channel Friday October MORNING NEWS REVOLUTION ANNIVERSARY SUPPLEMENT XXXI Conquered The...
- [Jessi Harewicz - Around Bowen Island | Marathon Swimmers Federation](#)  
Jessi Harewicz Around Bowen Island Marathon Swimmers Federation News Forum Database Swims Rules Members Jessi Harewicz Around Bowen Island Bowen...
- [LSX 700 Hot Tub Model from Michael Phelps Legend Series](#)  
LSX Hot Tub Model from Michael Phelps Legend Series Hot Tubs Shop By Price See All Shop By Size Adults...

**Google results**

Google How Can I Swim Without Getting Tired

About 30,700,000 results (0.43 seconds)



\*... And less efficient. So some mobility work could be in order you can do this on dry land. And also... More

YouTube https://www.youtube.com · Global Triathlon Network · How To Swim Without Getting Tired - YouTube

About featured snippets · Feedback

People also ask

Why do I get tired so fast while swimming?

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Single Link)

Enter text to search  
How Can I Swim Without Getting Tired

**Configuration**  
 Page Ranking  
 HITS

**Clustering**  
 Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**  
 Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH** **CLEAR** **HELP**

### Results

**Our results**

**Comments on: Learn How To Swim Easily**  
Comments on Learn How To Swim Easily Comments on Learn How To Swim Easily Learning to Swim and Having Fun...

**Comments on: Front Crawl / Freestyle Swimming Technique: Breathing**  
Comments on Front Crawl Freestyle Swimming Technique Breathing Comments on Front Crawl Freestyle Swimming Technique Breathing Learning to Swim and...

**Comments on: Learn How to Tread Water – Techniques and Exercises**  
Comments on Learn How to Tread Water Techniques and Exercises Comments on Learn How to Tread Water Techniques and Exercises...

**Butterfly Stroke Drills - One Armed Butterfly**  
Butterfly Stroke Drills One Armed Butterfly Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim...

**Learn How to Tread Water - Techniques and Exercises**  
Learn How to Tread Water Techniques and Exercises Skip to Content Search Magnifying Glass Search for Close Search Home Learn...

**Google results**

Google | How Can I Swim Without Getting Tired

About 30,700,000 results (0.43 seconds)



... And less efficient. So some mobility work could be in order you can do this on dry land. And also ... More

YouTube | <https://www.youtube.com> · Global Triathlon Network · How To Swim Without Getting Tired - YouTube

About featured snippets · Feedback

People also ask :  
Why do I get tired so fast while swimming?

## Sai Lasya Akkineni(SXA210025) - Clustering Agglomerative (Complete Link)

Enter text in search  
How Can I Swim Without Getting Tired

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

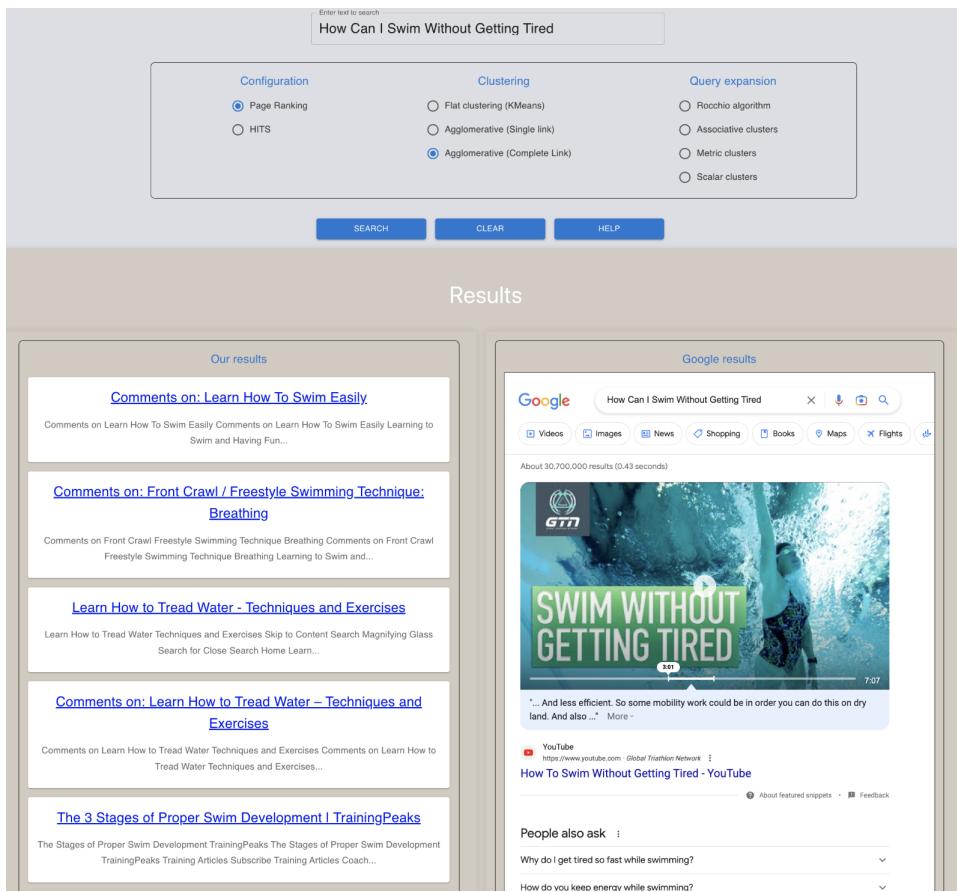
**Our results**

- Comments on: Learn How To Swim Easily**  
Comments on Learn How To Swim Easily Comments on Learn How To Swim Easily Learning to Swim and Having Fun...
- Comments on: Front Crawl / Freestyle Swimming Technique; Breathing**  
Comments on Front Crawl Freestyle Swimming Technique Breathing Comments on Front Crawl Freestyle Swimming Technique Breathing Learning to Swim and...
- Learn How to Tread Water - Techniques and Exercises**  
Learn How to Tread Water Techniques and Exercises Skip to Content Search Magnifying Glass Search for Close Search Home Learn...
- Comments on: Learn How to Tread Water – Techniques and Exercises**  
Comments on Learn How to Tread Water Techniques and Exercises Comments on Learn How to Tread Water Techniques and Exercises...
- The 3 Stages of Proper Swim Development | TrainingPeaks**  
The Stages of Proper Swim Development TrainingPeaks The Stages of Proper Swim Development TrainingPeaks Training Articles Subscribe Training Articles Coach...

**Google results**

Google   How Can I Swim Without Getting Tired   About 30,700,000 results (0.43 seconds)

**YouTube**   https://www.youtube.com · Global Triathlon Network   How To Swim Without Getting Tired - YouTube



# Sai Lasya Akkineni(SXA210025) - Clustering

Query 4: calories burnt in swimming

PageRanking

Enter text to search  
calories burnt in swimming

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

**Our results**

- [How To Make The Most Out of Paddle Boarding With An Apple Watch – AquaSportsPlanet](#)  
How To Make The Most Out of Paddle Boarding With An Apple Watch AquaSportsPlanet Skip to Content Search Search for...
- [Cooking Oil Smoke Points: High, Low, and Why It Matters](#)  
Cooking Oil Smoke Points High Low and Why It Matters Menu Visit Verywell Fit's homepage Know Your Cooking Oil...
- [How Swimming in Cold Water Helps Burn Calories | livestrong](#)  
How Swimming in Cold Water Helps Burn Calories livestrong Eat Better Get Fit Manage Weight Live Well More Challenges Topics...
- [What's the best swim stroke for burning calories? | Live Science](#)  
What's the best swim stroke for burning calories Live Science Skip to main content Open menu Close menu Live...
- [Calorie Calculator | ACTIVE](#)  
Calorie Calculator ACTIVE Sign In Sign Up My Events Running Sports Running Couch to K K K Half Marathon Marathon...

**Google results**

Google search results for "calories burnt in swimming". The query is expanded to include "1000 meters", "Vs running", "Freestyle swimming", "Breaststroke", and "30". The results page shows a snippet from Shape magazine about calorie burn rates for different swimming strokes, accompanied by a photo of a swimmer.

## Sai Lasya Akkineni(SXA210025) - Clustering

### Flat clustering

Enter text to search  
calories burnt in swimming

Configuration      Clustering      Query expansion

Page Ranking       Flat clustering (KMeans)       Rocchio algorithm  
 HITS       Agglomerative (Single link)       Associative clusters  
                           Agglomerative (Complete Link)       Metric clusters  
                           Scalar clusters

SEARCH    CLEAR    HELP

### Results

**Our results**

- [How To Make The Most Out of Paddle Boarding With An Apple Watch – AquaSportsPlanet](#)  
How To Make The Most Out of Paddle Boarding With An Apple Watch AquaSportsPlanet Skip to Content Search for...
- [Cooking Oil Smoke Points: High, Low, and Why It Matters](#)  
Cooking Oil Smoke Points High Low and Why It Matters Menu Visit Verywell Fit's homepage Know Your Cooking Oil...
- [Calorie Calculator ACTIVE](#)  
Calorie Calculator ACTIVE Sign In Sign Up My Events Running Sports Running Couch to K K K Half Marathon Marathon...
- [This Is Why Jet Skiing Is Great Exercise: A Doctor's View – AquaSportsPlanet](#)  
This Is Why Jet Skiing Is Great Exercise A Doctor's View AquaSportsPlanet Skip to Content Search for Close...
- [Best fitness trackers 2023 | Live Science](#)

**Google results**

Google    calories burnt in swimming

Images + 1000 meters + Vs running + Freestyle swimming + Breaststroke + 30

About 9,450,000 results (0.47 seconds)

Showing results for **calories burnt in swimming**  
Search instead for calories burnt in swimming

While swimming, your body burns 3.5 METs (223 calories per hour) treading water at a moderate effort; 8.3 METs (528 calories per hour) for a medium-speed, vigorous effort crawl; and 13.8 METs (878 calories per hour) for butterfly stroke. (These estimates are for a 140-pound adult.) Nov 23, 2022



Shape https://www.shape.com / Workouts by Type > Cardio

How Many Calories Does Swimming Burn? - Shape

About featured snippets · Feedback

People also ask

How many calories do you burn in 30 minutes of swimming?  
Is swimming a good way to lose weight?  
How long do you have to swim to burn 500 calories?

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Single Link)

Enter text to search  
calories burnt in swimming

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

**Our results**

- [How To Make The Most Out of Paddle Boarding With An Apple Watch – AquaSportsPlanet](#)  
How To Make The Most Out of Paddle Boarding With An Apple Watch AquaSportsPlanet Skip to Content Search for...
- [How Swimming in Cold Water Helps Burn Calories | livestrong](#)  
How Swimming in Cold Water Helps Burn Calories livestrong Eat Better Get Fit Manage Weight Live Well More Challenges Topics...
- [What's the best swim stroke for burning calories? | Live Science](#)  
What's the best swim stroke for burning calories Live Science Skip to main content Open menu Close menu Live...
- [This Is Why Jet Skiing Is Great Exercise: A Doctor's View – AquaSportsPlanet](#)  
This Is Why Jet Skiing Is Great Exercise A Doctor's View AquaSportsPlanet Skip to Content Search for Close...
- [12 Benefits of Swimming: Weight Loss, Health, and More](#)

**Google results**

Google   calories burnt in swimming   **X**

Images   + 1000 meters   + Vs running   + Freestyle swimming   + Breaststroke   + 30

About 9,450,000 results (0.47 seconds)

Showing results for **calories burnt in swimming**  
Search instead for calories burnt in swimming

While swimming, your body burns 3.5 METs (223 calories per hour) treading water at a moderate effort; 8.3 METs (528 calories per hour) for a medium-speed, vigorous effort crawl; and 13.8 METs (876 calories per hour) for butterfly stroke. (These estimates are for a 140-pound adult.) Nov 23, 2022



Shape   https://www.shape.com › Workouts by Type › Cardio   ⓘ   How Many Calories Does Swimming Burn? - Shape

About featured snippets · ⓘ Feedback

People also ask :

- How many calories do you burn in 30 minutes of swimming?
- Is swimming a good way to lose weight?
- How long do you have to swim to burn 500 calories?

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Complete Link)

Enter text to search  
calories burnt in swimming

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH**   **CLEAR**   **HELP**

### Results

**Our results**

[How To Make The Most Out of Paddle Boarding With An Apple Watch – AquaSportsPlanet](#)  
How To Make The Most Out of Paddle Boarding With An Apple Watch AquaSportsPlanet Skip to Content Search Search for...

[12 Benefits of Swimming: Weight Loss, Health, and More](#)  
Benefits of Swimming Weight Loss Health and More Health Conditions Featured Breast Cancer IBD Migraine Multiple Sclerosis MS Rheumatoid Arthritis...

[US Masters Open Water Swimming Drink Mix | INFINIT Nutrition](#)  
US Masters Open Water Swimming Drink Mix INFINIT Nutrition JavaScript seems to be disabled in your browser For the best...

[Cooking Oil Smoke Points: High, Low, and Why It Matters](#)  
Cooking Oil Smoke Points High Low and Why It Matters Menu Visit Verywell Fit s homepage Know Your Cooking Oil...

[Calorie Calculator ACTIVE](#)  
Calorie Calculator ACTIVE Sign In Sign Up My Events Running Sports Running Couch to K K K Half Marathon Marathon....

**Google results**

Google | calories burnt in swimming | Images | + 1000 meters | + Vs running | + Freestyle swimming | + Breaststroke | + 30 About 9,450,000 results (0.47 seconds)  
Showing results for **calories burnt in swimming**  
Search instead for calories burnt in swimming

While swimming, your body burns 3.5 METs (223 calories per hour) treading water at a moderate effort; 8.3 METs (528 calories per hour) for a medium-speed, vigorous effort crawl; and 13.8 METs (878 calories per hour) for butterfly stroke. (These estimates are for a 140-pound adult.) Nov 23, 2022



Shape https://www.shape.com • Workouts by Type • Cardio

How Many Calories Does Swimming Burn? - Shape

About featured snippets • Feedback

**People also ask :**

How many calories do you burn in 30 minutes of swimming?  
Is swimming a good way to lose weight?  
How long do you have to swim to burn 500 calories?  
Does swimming burn belly fat?

Feedback

Sai Lasya Akkineni(SXA210025) - Clustering

Query 5: common mistakes in swimming

## PageRanking

The screenshot shows a search interface with the following components:

- Search Bar:** "Enter text to search" with the query "common mistakes in swimming".
- Configuration Panel:** Contains three sections:
  - Configuration:** Radio buttons for "Page Ranking" (selected), "HITS", and "Clustering".
  - Clustering:** Radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", and "Agglomerative (Complete Link)".
  - Query expansion:** Radio buttons for "Rocchio algorithm", "Associative clusters", "Metric clusters", and "Scalar clusters".
- Buttons:** "SEARCH", "CLEAR", and "HELP".
- Results Section:** Titled "Results".
  - Our results:** A list of search results:
    - [Swimming Technique - Tweaks and Mistakes](#)
    - [3 Common Kicking Mistakes Swimmers Make | Swim Jim](#)
    - [Freestyle Stroke Swimming Mistakes: Overreaching](#)
    - [Front Crawl / Freestyle Swimming Technique: The Flutter Kick](#)
    - [Front Crawl - A Few Simple Tips](#)
  - Google results:** A screenshot of a Google search results page for "common mistakes in swimming". It shows the search bar, filters (Images, Videos, News, etc.), and a snippet from a Team USA article about common freestyle mistakes.

## Sai Lasya Akkineni(SXA210025) - Clustering

### Flat clustering

The screenshot shows a web-based clustering interface. At the top, there's a search bar with the query "common mistakes in swimming". Below it is a configuration panel with three tabs: "Configuration", "Clustering", and "Query expansion". Under "Configuration", "Page Ranking" is selected. Under "Clustering", "Flat clustering (KMeans)" is selected. Under "Query expansion", none of the options are selected. At the bottom of this panel are "SEARCH", "CLEAR", and "HELP" buttons.

The main area is titled "Results". It is divided into two sections: "Our results" on the left and "Google results" on the right.

**Our results:** This section displays five search results:

- Swimming Technique - Tweaks and Mistakes**  
Swimming Technique Tweaks and Mistakes Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...
- Freestyle Stroke Swimming Mistakes: Overreaching**  
Freestyle Stroke Swimming Mistakes Overreaching Skip to Content Search Magnifying Glass Search for Close Search Home Learn To swim pr...
- Front Crawl / Freestyle Swimming Technique: The Flutter Kick**  
Front Crawl Freestyle Swimming Technique The Flutter Kick Skip to Content Search Magnifying Glass Search for Close Search Home Learn...
- Front Crawl - A Few Simple Tips**  
Front Crawl A Few Simple Tips Skip to Content Search Magnifying Glass Search for Close Search Home Learn To swim...
- Front Crawl Swimming Mistakes - Wide Arm Recovery**  
Front Crawl Swimming Mistakes Wide Arm Recovery Skip to Content Search Magnifying Glass Search for Close Search Home Learn To...

**Google results:** This section shows the results from a Google search for the same query. It includes the search bar, filters (Images, Videos, News, etc.), and a snippet of the first result from Team USA.

**People also ask:** A list of related questions with dropdown menus:

- What are common mistakes in swimming?
- What are the 5 common mistakes while performing the freestyle?
- What are 3 5 common mistakes swimmers make while performing backstroke?
- What is commonly the biggest stumbling block in learning to swim?

**Feedback:** A link to provide feedback.

**MySwimPro** result:  
The 5 Most Common Freestyle Mistakes - MySwimPro  
Jan 26, 2021 – The 5 Most Common Freestyle Mistakes - 1. Looking Forward - 2. Kicking too Much - 3. Breathing Incorrectly - 4. Crossing Over on Hand Entry - 5.

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Single Link)

Enter text to search: common mistakes in swimming

**Configuration**

Page Ranking  
 HITS

**Clustering**

Flat clustering (KMeans)  
 Agglomerative (Single link)  
 Agglomerative (Complete Link)

**Query expansion**

Rocchio algorithm  
 Associative clusters  
 Metric clusters  
 Scalar clusters

**SEARCH** **CLEAR** **HELP**

### Results

**Our results**

**Swimming Technique - Tweaks and Mistakes**  
Swimming Technique Tweaks and Mistakes Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...

**3 Common Kicking Mistakes Swimmers Make | Swim Jim**  
Common Kicking Mistakes Swimmers Make Swim Jim New York Texas Toggle navigation SwimJim  
Swim Levels SwimBaby Programs Learn to Swim...

**Freestyle Stroke Swimming Mistakes: Overreaching**  
Freestyle Stroke Swimming Mistakes Overreaching Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...

**Front Crawl / Freestyle Swimming Technique: The Flutter Kick**  
Front Crawl Freestyle Swimming Technique The Flutter Kick Skip to Content Search Magnifying Glass Search for Close Search Home Learn...

**Front Crawl - A Few Simple Tips**  
Front Crawl A Few Simple Tips Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim...

**Google results**

Google common mistakes in swimming

Images Videos News + In freestyle + For Beginners + In breaststroke

About 14,700,000 results (0.45 seconds)

Team USA https://www.teamusa.org/MultisportLab/January/10 Common Freestyle Swim Mistakes and How to Correct ...  
Jan 11, 2019 – 1. Crossing the midline of the body on entry - 2. Elbow dropping on entry - 3. Windmill the arms - 4. Pushing the water to the side - 5. No early - 1. Crossing The Midline Of... 2. Elbow Dropping On Entry 3. Windmill The Arms

People also ask

What are common mistakes in swimming?  
What are the 5 common mistakes while performing the freestyle?  
What are 3 5 common mistakes swimmers make while performing backstroke?  
What is commonly the biggest stumbling block in learning to swim?

MySwimPro https://blog.myswimpro.com · 2021/01/26 · the-5-mo...  
The 5 Most Common Freestyle Mistakes - MySwimPro  
Jan 26, 2021 – The 5 Most Common Freestyle Mistakes - 1. Looking Forward 2. Kicking too Much 3. Breathing Incorrectly 4. Crossing Over on Hand Entry 5.

## Sai Lasya Akkineni(SXA210025) - Clustering

### Agglomerative (Complete Link)

The screenshot shows a search interface with two main sections: 'Our results' and 'Google results'.

**Configuration:**

- Page Ranking** (selected)
- HITS

**Clustering:**

- Flat clustering (KMeans)
- Agglomerative (Single link)
- Agglomerative (Complete Link)** (selected)

**Query expansion:**

- Rocchio algorithm
- Associative clusters
- Metric clusters
- Scalar clusters

**Buttons:** SEARCH, CLEAR, HELP

**Results Section:**

#### Our results

- Swimming Technique - Tweaks and Mistakes**  
Swimming Technique Tweaks and Mistakes Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...
- Freestyle Stroke Swimming Mistakes: Overreaching**  
Freestyle Stroke Swimming Mistakes Overreaching Skip to Content Search Magnifying Glass Search for Close Search Home Learn To Swim pr...
- Front Crawl / Freestyle Swimming Technique: The Flutter Kick**  
Front Crawl Freestyle Swimming Technique The Flutter Kick Skip to Content Search Magnifying Glass Search for Close Search Home Learn...
- Front Crawl - A Few Simple Tips**  
Front Crawl A Few Simple Tips Skip to Content Search Magnifying Glass Search for Close Search Home Learn...
- Front Crawl Swimming Mistakes - Wide Arm Recovery**  
Front Crawl Swimming Mistakes Wide Arm Recovery Skip to Content Search Magnifying Glass Search for Close Search Home Learn To...

#### Google results

Google search results for "common mistakes in swimming".

- Team USA** <https://www.teamusa.org/MultisportLab-January> ...  
10 Common Freestyle Swim Mistakes and How to Correct ...  
Jan 11, 2019 — 1. Crossing the midline of the body on entry · 2. Elbow dropping on entry · 3. Windmill the arms · 4. Pushing the water to the side · 5. No early ...  
1. Crossing The Midline Of... · 2. Elbow Dropping On Entry · 3. Windmill The Arms
- People also ask** ...
  - What are common mistakes in swimming?
  - What are the 5 common mistakes while performing the freestyle?
  - What are 3 5 common mistakes swimmers make while performing backstroke?
  - What is commonly the biggest stumbling block in learning to swim?
- MySwimPro** [https://blog.myswimpro.com/2021/01/26/the-5most...](https://blog.myswimpro.com/2021/01/26/the-5-most-common-freestyle-mistakes/) ...  
The 5 Most Common Freestyle Mistakes - MySwimPro  
Jan 26, 2021 — The 5 Most Common Freestyle Mistakes · 1. Looking Forward · 2. Kicking too Much · 3. Breathing Incorrectly · 4. Crossing Over on Hand Entry · 5.
- Videos** ...

Above are the results for five queries with page ranking and different types of clustering. Based on the precomputed clusters, different clustering methods display different urls in the top based on the similarity of relevant urls retrieved through page ranking.

## **6.Query expansion and relevance feedback:**

Query Expansion is any technique used to improve the precision of search results. As part of 'DIVE', I have implemented the following local query expansion techniques:

- Rocchio Algorithm
- Automatic Query Expansion using
  - Association Clusters
  - Metric Clusters
  - Scalar Clusters

### **Rocchio Algorithm**

Formula Used:

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

Number of results from the original query considered = 30

Number of local relevant documents => = 20

Number of local non-relevant documents=> = 10

Alpha = 1, Beta = 0.8 and Gamma = 0.2 since positive weight > negative weight

**Queries Considered:**

The 20 queries used to test the algorithm were selected based on query length, specific to generic queries to assess the relevance of the documents being retrieved. The selected queries and their respective expanded queries are as follows:

Query	Expanded Query
Swimming	Swimming hall fame intern water
Medals	Medals olymp gold silver bronz
Swimming accidents	Swimming accidents water pool safeti
Best swimming schools	Best swimming schools lesson colleg sport
Olympics winners	Olympics winners swim world km femal
Swimsuits	Swimsuits swim swimwear water add
Breathing techniques	Breathing techniques swim learn stroke
List of competitions	List of competitions judo rank ijf world
Swimming rules	Swimming rules swimmer stroke meet
Caeleb Dressel	Caeleb Dressel swim world news
Learn butterfly stroke	Learn butterfly stroke swim drill
Japanese swimmers	Japanese swimmers swim year hour
High dive	High dive tank us scuba view
Triathlon	Triathlon de item le wetsuit
List of swimming competitions	List of swimming competitions world open usm
Deep diving	Deep diving diver underwat scuba water
Types of pools	Types of pools swim water swimmer
Freestyle stroke	Freestyle stroke swim techniqu learn
Improve technique	Improve technique swim learn stroke lesson
Debby Meyer	Debby Meyer meter olymp age imag getti

*Devarsha Sikharan(DXS210063) - Query Expansion and Relevance Feedback*

**Examples of relevant and irrelevant web pages:**

**Triathlon:**

**Relevant:**

[https://en.m.wikipedia.org/wiki/International\\_Triathlon\\_Union](https://en.m.wikipedia.org/wiki/International_Triathlon_Union)

[https://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_Triathlon](https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Triathlon)

<https://it.aquasphereswim.com/it/mute-nuoto/allenamento/pursuit-v3---muta-triathlon-donna-SU838.html>

**Irrelevant:**

<https://uk.aquasphereswim.com/en/usermanuals>

**Caeleb Dressel:**

**Relevant:**

<https://www.swimmingworldmagazine.com/news/world-championships-caeleb-dressel-out-for-remainder-of-the-meet/>

<https://www.swimmingworldmagazine.com/meet/2019-atlanta-classic>

<https://www.sportskeeda.com/swimming/news-olympian-caeleb-dressel-returns-training-reports>

**Irrelevant:**

<https://apnews.com/hub/michael-phelps>

<https://apnews.com/hub/katie-ledecky>

**Judgement:**

As there are the queries I formulated, I have made the judgement based on the relevance between my information need and the returned results. If the result is close to my need, I chose it as relevant. If the result is related to the topic but not to my need or irrelevant to the topic, I deemed it irrelevant.

### Automatic Query Expansion:

I have implemented automatic query expansion using association clusters, metric clusters and scalar clusters. The results of eac for 3 example queries are given here and in the 'automaticQueryExpansion' document.

### Query Selection:

I have used the 20 queries from Rocchio Algorithm so that I can compare the results among the 4 approaches implemented. I have also tried another 30 queries of which some are irrelevant to swimming, about people relevant to swimming, short and long queries, clear and vague queries etc.

### Note:

Since the Local Document Set (20), Local Vocabulary, Local stems and correlation values of top 10 stems are huge, I have attached them in another document titled 'automaticQueryExpansion'.

### Associations Clustering:

#### Example 1:

**Original Query:** Best Swimming Schools

**Expanded Query:** Best Swimming Schools lesson univers

### Search Engine Results:

The screenshot shows a search interface with the following components:

- Search Bar:** "Best Swimming Schools".
- Configuration Panel:**
  - Page Ranking** (selected)
  - HITS**
- Clustering Options:**
  - Flat clustering (KMeans)**
  - Agglomerative (Single link)**
  - Agglomerative (Complete Link)**
- Query expansion Options:**
  - Rocchio algorithm**
  - Associative clusters** (selected)
  - Metric clusters**
  - Scalar clusters**
- Buttons:** SEARCH, CLEAR, HELP.
- Results Section:** "Results" heading, followed by the expanded query: "Best Swimming Schools men women divis view top colleg naia lesson".
- Left Panel (Our results):**
  - The NAIA Changes its National Championship Invitational Rules Again**
  - Best Swimming Colleges | SportsEngine**
  - College basketball - Wikipedia**
  - Basketball Scholarships | How to Get a Basketball Scholarship**
- Right Panel (Google results):**
  - Google** search bar with query "Best Swimming Schools".
  - Results for Plano, TX 75075**
  - Filtering:** News, Images, Shopping, Maps, DI, Women's, Videos, Division 3, Men's.
  - About 189,000,000 results (0.56 seconds)**
  - Top Result:** NCSA (https://www.recreaps.org/test-colleges-best-men...) with ranking 1.
  - Best Men's Swimming Colleges** listed with rankings 1-5.
  - People also ask:**
    - Who is ranked 1 in women's college swimming?
    - What is the best college women's swim team?
    - What is Auburn women's swimming ranked?

**Example 2:**

**Original Query:** Olympic Medals

**Expanded Query:** olympic medals gold china silver men bronz

**Search Engine Results:**

The screenshot shows a search interface with the following components:

- Search Bar:** Displays the query "Olympic Medals".
- Configuration Panel:** Contains three sections:
  - Configuration:** Radio buttons for "Page Ranking" (selected), "HITS", and "Clustering".
  - Clustering:** Radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", and "Agglomerative (Complete Link)".
  - Query expansion:** Radio buttons for "Rocchio algorithm", "Associative clusters" (selected), "Metric clusters", and "Scalar clusters".
- Buttons:** "SEARCH", "CLEAR", and "HELP".
- Results Section:** Headed "Results" with the expanded query "Olympic Medals game gold first time".
- Our results:** A list of search results:
  - [United States at the Summer Olympics - Wikipedia](#)
  - [The Olympic Games Quiz | Britannica](#)
  - [Top 10 Greatest Olympians Of All Time | Sportshubnet](#)
  - [Top 10 Greatest Swimmers Of All Time | Sportshubnet](#)
- Google results:** A screenshot of a Google search results page for "Olympic Medals". It shows the search bar, filters (Images, News, Videos, Shopping, + By country, + 2022, + 2021, + India, + Count all-time, All filters, Tools), and a snippet about Olympic medals. Below the snippet are images of Gold, Silver, and Bronze medals.

**Example 3:**

**Original Query:** types of strokes

**Expanded Query:** types of strokes pool race olymp cours train

**Search Engine Results:**

The screenshot shows a search interface with the following components:

- Configuration:** Includes options for "Page Ranking" (selected), "HITS", "Clustering" (with "Flat clustering (KMeans)" selected), "Agglomerative (Single link)", "Agglomerative (Complete Link)", "Query expansion" (with "Rocchio algorithm" selected), "Associative clusters" (selected), "Metric clusters", and "Scalar clusters".
- Buttons:** SEARCH, CLEAR, HELP.
- Results:** The word "Results" is centered below the configuration panel.
- Expanded query:** types of strokes pool cours competit race event olymp swimmer use
- Our results:** A section containing several links:
  - [What is Swimming?](#)
  - [What are the Different Types of Competitive Swimming?](#)
  - [What Are The Rules Of Swimming?](#)
  - [What Are Some Basic Swimming Terms And Lingo To Know?](#)
- Google results:** A screenshot of a Google search results page for the query "types of strokes". The results include:
  - A sponsored result from TexasHealth.org about stroke types.
  - A result from VirtualStrokeCare.com.
  - A snippet: "What are the types of stroke? There are two types of stroke: ..."
  - Image thumbnails for "Common Types of Stroke".

## Metric Clustering:

### Example 1:

**Original Query:** drowning while swimming

**Expanded Query:** drowning while swimming unspecifi subset

### Search Engine Results:

The screenshot shows a search interface with the following components:

- Configuration:** Options include "Page Ranking" (selected), "HITS", "Clustering" (with "Metric clusters" selected), and "Query expansion" (with "Metric clusters" selected).
- Search Bar:** The query "drowning while swimming" is entered.
- Buttons:** SEARCH, CLEAR, HELP.
- Results Section:** The heading "Results" is displayed above a list of search results. The expanded query is shown as "Expanded query: drowning while swimming retriev death water j alcohol journal peopl origin".
- Our results:** A list of local search results:
  - Drowning Facts | Drowning Prevention | CDC
  - Drowning Wikipedia
  - Prevention | Drowning Prevention | CDC
  - Safety | Free Full-Text | The Influence of Alcohol and Drugs on Drowning among Victims of Senior Ye
  - Review of the role of alcohol in drowning associated with recreational aquatic activity | Injury Pre
- Google results:** A screenshot of the Google search results page for the query "drowning while swimming". It shows the search bar, navigation links, and a list of search results.

Devarshaa Sikharam(DXS210063) - Query Expansion and Relevance Feedback

## Example 2:

**Original Query:** Japanese Swimmers

**Expanded Query:** japanese swimmers suieijutsu tastugo canadian

## Search Engine Results:

The screenshot shows a search interface with the following components:

- Search Bar:** "Enter text to search" with the query "Japanese Swimmers".
- Configuration Panel:** Includes:
  - Configuration:** Radio buttons for "Page Ranking" (selected) and "HITS".
  - Clustering:** Radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", and "Agglomerative (Complete Link)".
  - Query expansion:** Radio buttons for "Rocchio algorithm", "Associative clusters", "Metric clusters" (selected), and "Scalar clusters".
- Buttons:** "SEARCH", "CLEAR", and "HELP".
- Results Section:** A large area labeled "Results" containing the expanded query: "Japanese Swimmers cincinnati tastugo volunt plight pace".
- Our results:** A list of search results:
  - [Tatsugo Kawaishi - Wikipedia](#)
  - [Julie Otsuka - Wikipedia](#)
  - [Yoshi Oyakawa - Wikipedia](#)
  - [Masters I Swimming World](#)
  - [Support craft - Openwaterpedia](#)
- Google results:** A screenshot of a Google search results page for "Japanese Swimmers". It shows a grid of swimmer profiles with names like Rikako Ikee, Kosuke Kitajima, Dalya Seto, etc. Below the grid is a section titled "Images for Japanese Swimmers" with a preview of several images.

### Example 3:

**Original Query:** Scuba Diving

**Expanded Query:** scuba diving overwhelmingli sherwood ranch

### Search Engine Results:

The screenshot shows a search interface with the following components:

- Query Bar:** "Scuba Diving"
- Configuration Panel:** Includes "Configuration" (radio buttons for "Page Ranking" and "HITS"), "Clustering" (radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", "Agglomerative (Complete Link)", "Rocchio algorithm", "Associative clusters", "Metric clusters", and "Scalar clusters"), and "Query expansion".
- Buttons:** "SEARCH", "CLEAR", and "HELP".
- Results Section:** Headed "Results" with the expanded query "Scuba Diving muck inshore inland unman paint".
- Our results:** A list of search results:
  - Diving Techniques | General Diving Techniques for Khao Lak Dives
  - Can You Paint Your Scuba Tank? – AquaSportsPlanet
  - Template:Underwater diving - Wikipedia
  - Dynamic apnea – Wikipedia
  - Swimming at the 1900 Summer Olympics – Men's underwater swimming - Wikipedia
- Google results:** A screenshot of a Google search results page for "Scuba Diving". It shows:
  - Google search bar: "Scuba Diving"
  - Search results:
    - Results for "Plane, TX 75075" (dropdown for "Use precise location")
    - Wikipedia entry for "Scuba diving": "Scuba diving is a mode of underwater diving whereby divers use breathing equipment that is completely independent of a surface air supply, and therefore has..."
    - Image of a scuba diver.
    - Section "Max depth": "How deep can you scuba dive"
    - Section "Equipment": "Scuba diving equipment"
    - Section "Scuba Diving Games": "BZ", "BND", "DEEP DIVING SIMULATOR", "DEEP DIVING ADVENTURES"

## Scalar Clustering:

### Example 1:

**Original Query:** swimming drills

**Expanded Query:** swimming drills learn

### Search Engine Results:

The screenshot shows a search interface with the following components:

- Configuration:** Options include "Page Ranking" (selected), "HITS", "Clustering" (radio buttons for "Flat clustering (KMeans)", "Agglomerative (Single link)", and "Agglomerative (Complete Link)"), and "Query expansion" (radio buttons for "Rocchio algorithm", "Associative clusters", "Metric clusters", and "Scalar clusters").
- Buttons:** SEARCH, CLEAR, and HELP.
- Results:** The word "Results" is centered below the configuration area.
- Expanded query:** "swimming drills learn water post"
- Our results:** A list of search results:
  - Breaststroke Archives - Enjoy Swimming
  - Backstroke Archives - Enjoy Swimming
  - Backstroke Drills Archives - Enjoy Swimming
  - Swimming Strokes SEO Title
  - Learn Swimming Breaststroke - Drills and Exercises
  - Backstroke Swimming Drills - Hand-Lead Side balance
- Google results:** A screenshot of a Google search results page for "swimming drills". It shows the search bar, filters (Shopping, Videos, Images, Pdf, Presetyle, For beginners PDF, Advanced, Freestyle), and a snippet about 10 essential pool swim drills for triathletes. Below the snippet, there's a result from MySwimPro about 5 Freestyle Drills For Beginner Swimmers. The "People also ask" section includes questions like "What are drills in swimming?", "How do you swim freestyle drills for beginners?", and "How can I train better to swim?".

## Example 2:

**Original Query:** what are freestyle strokes

**Expanded Query:** what are freestyle strokes swim swimmer

## Search Engine Results:

The image shows two side-by-side search results pages. On the left is a local search interface with a configuration panel for Page Ranking, Clustering (Flat clustering, Agglomerative, Agglomerative (Complete Link)), and Query expansion (Rocchio algorithm, Associative clusters, Metric clusters, Scalar clusters). The query 'what are freestyle strokes' is entered in the search bar. On the right is a Google search results page for the expanded query 'what are freestyle strokes swim swimmer'. The results include a snippet from 'Enjoy Swimming' explaining the freestyle stroke, two images of swimmers performing the stroke, and a link to 'Freestyle Stroke: Overview and Swimming Technique'. A sidebar titled 'People also ask' lists related questions like 'What strokes are allowed in freestyle?' and 'What is the difference between freestyle and stroke?'. At the bottom, there's a 'Feedback' link.

Devarshaa Sikharam(DXS210063) - Query Expansion and Relevance Feedback

### Example 3:

**Original Query:** swimsuits

**Expanded Query:** swimsuits women us one swim

**Search Engine Results:**

The screenshot shows a search interface with the following components:

- Configuration:** Includes "Page Ranking" (selected), "HITS", "Clustering" (with options: Flat clustering (KMeans), Agglomerative (Single link), Agglomerative (Complete Link)), and "Query expansion" (with options: Rocchio algorithm, Associative clusters, Metric clusters, Scalar clusters).
- Buttons:** SEARCH, CLEAR, HELP.
- Results Section:** Headed "Results" with the expanded query "swimsuits piec one women swim water top suit new short".
- Our results:** A list of search results:
  - [Swimsuit - Wikipedia](#)
  - [Men's Drag Suits for Swimming | Drag Swimsuits](#)
  - [Bathing Suits | Women's Athletic Bikinis and One Piece Swimwear](#)
  - [Tankini Swimsuits | SwimOutlet.com](#)
  - [Women's Tech Suits | SwimOutlet.com](#)
  - [Water Aerobics Swimsuits & Swimwear - Aqua Aerobics Swimwear & Swimsuits - Kiefer Aquatics](#)
- Google results:** A screenshot of a Google search results page for "swimsuits". It shows sponsored ads for Speedo and Summersalt swimwear, along with other search filters like Shopping, Images, Videos, etc.

**Collaboration for Result Reflection:**

I have collaborated with Abishek to reflect results from query expansion in the UI. We needed to alter the results object structure to accommodate the return of the expanded query along with the retrieved documents as all other parts returned only the result set. We also worked on a common run.py file for the query expansion module which takes in user request type, query and directs the execution based on the type of request.

**Queries Selected and Their Expansions:**

I have tried over 50 queries for each method and picked the queries mentioned in the examples above as they are diverse in terms of topic, length and relevance. I found that the queries expanded using Rocchio Algorithm, association clusters and scalar clusters were better than those from metric clusters.

*Sai Lasya Akkineni(SXA210025)*  
*Devarshaa Sikharam(DXS210063)*  
*Abishek Kumar (axk200020)*  
*Pallavi Dasarapu (PXD210008)*  
*Sangeetha Pradeep (sxp210004)*

## **7. Discussion:**

To build the search engine, Pallavi approximately used 110 seed urls and crawled around 143k webpages. Apache Nutch is used for crawling. The crawled data is then indexed by Sangeetha using Apache Solr. HITS and Page ranking are also implemented. Indexed data is then used by Lasya to perform flat clustering and agglomerative clustering (single-link vs complete link). Scikit-learn library is used for flat clustering(KMeans) and fastcluster for agglomerative clustering. Devarshaa worked on query expansion using Rocchio algorithm, associative, metric and scalar clusters. We all collaborated with Abishek who implemented User Interface using React and integrated with backend using Flask API. UI has different options for retrieval of web pages based on the query.

## **8. Conclusion:**

Experimenting with different queries using different techniques improved the performance in different ways. Page ranking and HITS improved the relevance of results with query. Clustering grouped the similar relevant results together improving the relevance of top results shown to user. Query expansion helped in creating an enhanced query from the input query given by the user improving the accuracy of retrieving relevant results.