

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

**Project TITLE: Basketball**

**TEAM 3**

X1(Akshaiy Ramlinkam), [axj210002@utdallas.edu](mailto:axj210002@utdallas.edu)

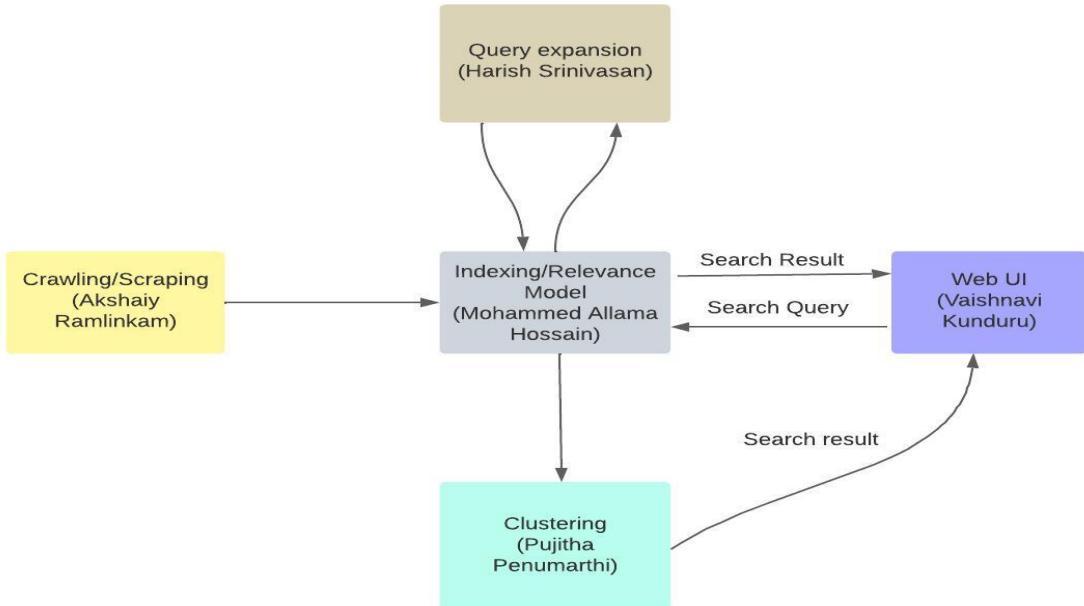
X2(Allama Hossain), [mah200001@utdallas.edu](mailto:mah200001@utdallas.edu)

X3(Vaishnavi Kunduru), [kxv200027@utdallas.edu](mailto:kxv200027@utdallas.edu)

X4(Pujitha Penumarthy), [vxp200010@utdallas.edu](mailto:vxp200010@utdallas.edu)

X5(Harish Srinivasan), [hxs200044@utdallas.edu](mailto:hxs200044@utdallas.edu)

## THE PROBLEM:



The goal of our project was to build a search engine pertaining to the topic - **Basketball**. We executed this project in 5 stages as mentioned below:

1. **Crawling** - 100693 relevant web pages were crawled by **Akshaiy Ramlinkam**, and the output data was produced in the required format to the teammates who were responsible for indexing and building relevance models, and clustering.
2. **Indexing and Relevance Models** - **Mohammed Allama Hossain** handled this part. He used the crawled data, indexed it, and built relevance models such as HITS, PageRank, and Vector Space relevance model. He then sent the PageRank results to the teammates responsible for clustering, query expansion, and the user interface..
3. **User Interface** - **Vaishnavi Kunduru** handled this part of the project. She integrated the frontend with the python scripts sent by teammates who performed indexing and relevance models, clustering, and query expansion. She also used the Google and Bing API to show the expected results in the frontend.
4. **Clustering** - This part was coded by **Pujitha Penumarthi**. She performed flat clustering and hierarchical clustering (single and complete) using the data provided by Akshaiy. She stored the results of clustering in csv file. Then she used that to combine the relevance model results from Allama and optimized those results using the clusters that she had already built. Then she sent these results to Vaishnavi, who was responsible for the user interface.
5. **Query Expansion** - **Harish Srinivasan** was in charge of this part. He received the data from Allama and performed the following query expansion techniques: Associative clustering, Scalar clustering and Metric clustering. Then the expanded query based on associative clustering was sent to Vaishnavi, who would then trigger a search by inputting this query into Allama's relevance models. The output results are then displayed in the user interface.

## **CRAWLING:**

To perform web crawling, I used the open source tool **Scrapy** which is python-based. Using the inbuilt selectors (CSS/xpath) provided within scrapy, I was able to crawl and scrape required data from the web pages. I understood that my teammate who was working on indexing and relevance models (Allama) required a CSV file with columns such as **Parent link (the link that led to the current page)**, **Child link (current page)**, **Title and Content**. I was able to provide the crawling output in that format. To crawl, I had to first find suitable domains to gather the right information from. I also made sure that the crawler doesn't derail into random topics.

Overall, I was able to get data from **100693 web pages using 70 start URLs**.

I used the following 2 methods for crawling:

1. **CrawlSpider:** I selected the start URLs to be all the NBA teams' wiki pages, and applied crawling by allowing only URLs that contain NBA/basketball/FIBA etc. and denying URLs that might lead to irrelevant sites such as football/hockey/baseball etc. These constraints were set using the **Rule** and **LinkExtractor** class present in Scrapy.

I used CrawlSpider for the following websites:

- a. [en.wikipedia.org/wiki](http://en.wikipedia.org/wiki)
- b. [www.espn.com/nba/story](http://www.espn.com/nba/story)
- c. [www.basketballnews.com/stories](http://www.basketballnews.com/stories)
- d. [www.ncaa.com/news/basketball-men/article](http://www.ncaa.com/news/basketball-men/article)

2. **scrapy.Spider:** For sites which contain basketball news articles without any links leading to other news articles, I took advantage of the “**Next**” button. Here, as I can only crawl using the “**Next**” button in the page, I used the simple CSS selector to traverse through all the pages containing article links till I’ve visited all the articles from the site. I made sure that the web pages were not duplicated by using the **RFPDupeFilter** class in Scrapy.

I used scrapy.Spider for the following websites:

- a. [www.theringer.com/nba/archives](http://www.theringer.com/nba/archives)
- b. [www.global.nba.com/news](http://www.global.nba.com/news)
- c. [www.basketball.realgm.com](http://www.basketball.realgm.com)
- d. [www.basketballforever.com/latest-news](http://www.basketballforever.com/latest-news)
- e. [www.ballislife.com/news](http://www.ballislife.com/news)
- f. [www.sportingnews.com/us/nba/news](http://www.sportingnews.com/us/nba/news)

**Start URLs that I used:**

Having **wiki\_url2** = “<https://en.wikipedia.org/wiki/>”

```
wiki_url2+"Portland_Trail_Blazers", wiki_url2+"New_York_Knicks",
wiki_url2+"Boston_Celtics",

wiki_url2+"Philadelphia_76ers", wiki_url2+"Toronto_Raptors",
wiki_url2+"Chicago_Bulls",

wiki_url2+"Cleveland_Cavaliers", wiki_url2+"Detroit_Pistons",
wiki_url2+"Indiana_Pacers",

wiki_url2+"Milwaukee_Bucks", wiki_url2+"Atlanta_Hawks",
wiki_url2+"Charlotte_Hornets",

wiki_url2+"Miami_Heat", wiki_url2+"Orlando_Magic",
wiki_url2+"Washington_Wizards",

wiki_url2+"Denver_Nuggets", wiki_url2+"Minnesota_Timberwolves",
wiki_url2+"Oklahoma_City_Thunder",

wiki_url2+"Utah_Jazz", wiki_url2+"Golden_State_Warriors",
wiki_url2+"Los_Angeles_Clippers",

wiki_url2+"Los_Angeles_Lakers", wiki_url2+"Phoenix_Suns",
wiki_url2+"Sacramento_Kings",

wiki_url2+"Dallas_Mavericks", wiki_url2+"Houston_Rockets",
wiki_url2+"Memphis_Grizzlies",

wiki_url2+"New_Orleans_Pelicans", wiki_url2+"San_Antonio_Spurs",
wiki_url2+"Basketball",

wiki_url2+"National_Basketball_Association", wiki_url2+"EuroBasket",
wiki_url2+"College_basketball",

wiki_url2+"List_of_National_Basketball_Association_rivalries",
wiki_url2+"NBA_draft", wiki_url2+"Basketball_positions"
```

**Rules for the above crawl (I have denied the crawler from entering into links that satisfy the format mentioned in “pattern” and only crawl into links satisfying the condition mentioned in the “allow” parameter):**

```
pattern =
re.compile(r".*(category|talk|baseball|football|template|rugby|hockey).*"
, re.IGNORECASE)

le1 = LinkExtractor(deny = pattern,
allow=r"https://en.wikipedia.org/wiki/.*(?:NBA|basketball|coach|NCAA|FIBA
|league|division|draft|WNBA).*",
restrict_xpaths="//main[@class='mw-body']//a")
```

- [https://en.wikipedia.org/wiki/List\\_of\\_National\\_Basketball\\_Association\\_Players\\_\(A-Z\)](https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_Players_(A-Z))
- <https://www.theringer.com/nba/archives>
- <http://global.nba.com/news/>
- <https://basketballforever.com/latest-news>
- <https://ballislife.com/news/>
- [https://www.espn.com/nba/story/\\_/id/36191140/heat-rally-stun-bulls-face-bucks-first-round](https://www.espn.com/nba/story/_/id/36191140/heat-rally-stun-bulls-face-bucks-first-round)
- [https://www.espn.com/nba/story/\\_/id/36127955/with-lebron-front-row-bronny-james-leads-us-nike-ho
op-summit](https://www.espn.com/nba/story/_/id/36127955/with-lebron-front-row-bronny-james-leads-us-nike-ho
op-summit)

**The rule used for the crawling of the 2 URLs above:**

```
le = LinkExtractor(allow=r"https://www.espn.com/nba/story/.*")

rule1 = Rule(le, callback='parse_item', follow=True)
```

- <https://www.sportingnews.com/us/nba/news>

- <https://www.basketballnews.com/>

**The rule used for the crawling of the URL above:**

```
le = LinkExtractor(allow=r"https://www.basketballnews.com/stories/.*)")

rule1 = Rule(le, callback='parse_item', follow=True)
```

- [https://www.ncaa.com/news/basketball-men/article/2022-12-05/college-basketballs-net-rankings-expained?utm\\_campaign=mbk-rr-links](https://www.ncaa.com/news/basketball-men/article/2022-12-05/college-basketballs-net-rankings-expained?utm_campaign=mbk-rr-links)

**The rule used for the crawling of the URL above:**

```
le = LinkExtractor(allow=r"https://www.ncaa.com/news/basketball-men/article/.*)"

rule1 = Rule(le, callback='parse_item', follow=True)
```

**NOTE:**

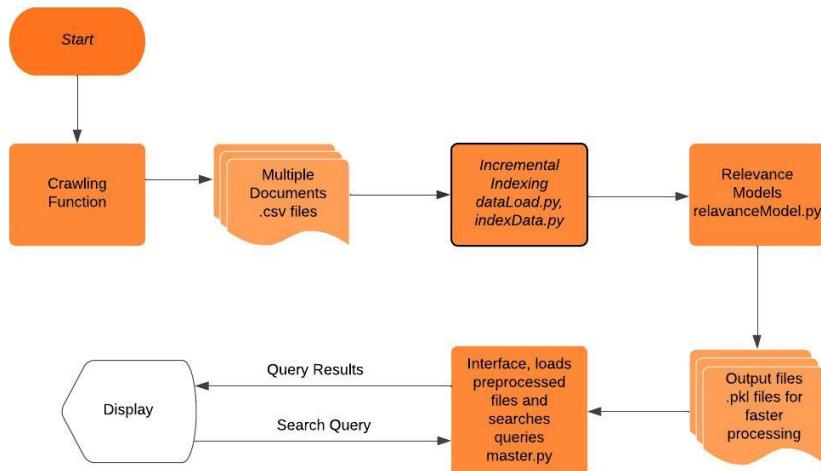
- I also had to send my crawled data to Pujitha, who was responsible for the Clustering part. We had a discussion on the format in which she needed the crawled data. As her main needs were based on the title and content of the web pages, she was also okay with the output file being a CSV file with the following columns: **Parent Link, Child Link, Title, Content.**
- As I used Scrapy to perform crawling, the **robots.txt** file will by default not allow the crawling of sensitive/protected web pages.
- Using the **100693** unique web pages that were crawled, the results were too skewed towards the wikipedia pages. So we used only **62376** URLs by intentionally removing most of the wikipedia pages to get results containing other news article sites.

## 2. Incremental Indexing, Link Analysis of Web Graph and Relevance Models

### Indexing:

In this project I created a function named dataTokenizer to index the content and title of the incoming crawled data. We used a global documentID as the key to index these data. The indexed data was stored in a pickle file which was appended on every function run thereby ensuring incremental indexing of the incoming crawled data. The indexing program created the dictionary and the posting files to be used later in the vector space model. Indexing of the crawled data can be done by simply specifying the path of the data in the dataLoad.py file and then by running the following command

```
> python3 dataLoad.py
```



### Web graph creation:

The web graph was generated using the NetworkX library in Python. To prepare the graph, we had to assemble the data in a dataframe and specify the edge direction from the column named ‘Source Link’ to the column named ‘Link’. This library allowed us to find out statistics of the web graph created. The library also provided PageRank and HITS functionality.

Following are the required statistics:

- Number of nodes - 62376

- Number of edges - 59400
- Largest number of ingoing and outgoing links for the graph - 520
  - [https://en.wikipedia.org/wiki/List\\_of\\_National\\_Basketball\\_Association\\_players\\_\(B\)](https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(B))

### **Graph to Index connection and Relevance Model creation:**

Since the data was indexed on the basis of a global DocumentID, we used the same DocumentID to connect the information from the graph to the index. This means that PageRank and HITS results are ordered on the basis of DocumentID.

The vector space model is used to search the results and the PageRank and the HITS algorithm augments the search results produced by vector space. The vector space model will display results having the highest tf-idf score and this result will be further enhanced by combining the results of the PageRank and HITS algorithm. In a way, we are ordering the results of vector space again on the basis of PageRank and HITS. The weighing scheme used in the vector space model is as follows:

$tf = 1 + \log(tf)$  for the query and document vectors

$idf = \log(N/df)$  where N is the total number of documents and df is the document frequency.

$tf\text{-}idf = tf * idf$  TF-IDF score.

The two relevance model schemes were created using the library NetworkX. I ran both the models to a maximum iteration of 1000 while keeping the error tolerance to check convergence at  $10^{-6}$ . Since we did not have enough time to differentiate and mark important links, we had to resort to keeping the weight of all links as a default 1.

### **Topic Based Page Ranking:**

For the purpose of testing topic based page ranking, I selected the topic of ‘LeBron James’ and the following links were returned with highest PageRank scores:

<a href="https://www.espn.com/nba/story/_/id/36127955/what-lebron-front-row-bronny-james-leads-us-nike-hoop-summit">https://www.espn.com/nba/story/_/id/36127955/what-lebron-front-row-bronny-james-leads-us-nike-hoop-summit</a>	4.225079670947696e-05
<a href="https://www.espn.com/nba/story/_/id/25995187/10-things-like-including-mvp-race-nba">https://www.espn.com/nba/story/_/id/25995187/10-things-like-including-mvp-race-nba</a>	4.225079670947696e-05
<a href="https://www.espn.com/nba/story/_/id/26228166/lebron-vs-mj-lakers-season-great-nba-if">https://www.espn.com/nba/story/_/id/26228166/lebron-vs-mj-lakers-season-great-nba-if</a>	4.225079670947696e-05
<a href="https://www.espn.com/nba/story/_/id/18560667/zach-lowes-10-things-like-featuring-lebron-james-passing-nba">https://www.espn.com/nba/story/_/id/18560667/zach-lowes-10-things-like-featuring-lebron-james-passing-nba</a>	4.225079670947696e-05

<a href="https://www.espn.com/nba/story/_/id/25513076/theres-point-revolution-lebron-noticed">https://www.espn.com/nba/story/_/id/25513076/theres-point-revolution-lebron-noticed</a>	4.225079670947696e-05
---	-----------------------

The following links had the highest HITS authority scores:

<a href="https://en.wikipedia.org/wiki/Mike_Bytzura">https://en.wikipedia.org/wiki/Mike_Bytzura</a>	0.0019230769230769236
<a href="https://en.wikipedia.org/wiki/Tommy_Byrnes">https://en.wikipedia.org/wiki/Tommy_Byrnes</a>	0.0019230769230769236
<a href="https://en.wikipedia.org/wiki/Marty_Byrnes">https://en.wikipedia.org/wiki/Marty_Byrnes</a>	0.0019230769230769236
<a href="https://en.wikipedia.org/wiki/Andrew_Bynum">https://en.wikipedia.org/wiki/Andrew_Bynum</a>	0.0019230769230769236
<a href="https://en.wikipedia.org/wiki/Will_Bynum">https://en.wikipedia.org/wiki/Will_Bynum</a>	0.0019230769230769236

The following links had the highest HITS hubs scores:

<a href="https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(B)">https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(B)</a>	1.0000000000000002
<a href="https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(M)">https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(M)</a>	8.478322998107585e-15
<a href="https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(H)">https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(H)</a>	4.06765943705588e-15
<a href="https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(C)">https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(C)</a>	2.5830957497781744e-15
<a href="https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(K)">https://en.wikipedia.org/wiki/List_of_National_Basketball_Association_players_(K)</a>	8.63299279936631e-16

From the above results, we see that wikipedia pages pertaining to players have the highest HITS authority scores, whereas all index wikipedia pages pointing to the starting letter of basketball players have the highest hub scores. As learnt in theory, links with highest hub scores are pointing to pages having high relevance (in our example, it is pointing to the wiki pages of basketball players) whereas the wiki pages of players itself have the highest authority scores.

### **Collaboration with other students:**

In the very first project meeting, I discussed with Vaishnavi (user interface), Harish (query expansion) and Pujitha (clustering) about the different kinds of data structure we can use to return the results and we all agreed upon a list of dictionaries. So all my output functions returned the results in the form of a list of dictionaries which were exploited by other team members to execute their functionalities.

Once Vaishnavi (user interface) was ready with her part and had integrated my results into the user interface, we tested random queries related to basketball to check if the user interface results and command line results matched or not.

### **Query Testing:**

I tested 12 queries having direct hits, as in, queries directly picked up from the content or title of the crawled data and the rest 12 queries were random search strings such as ‘LeBron James’, ‘Chicago Bulls’, ‘Abdul Kareem’. The query gave exact results for all queries having direct hits, and other results were also related to the given search query. Even for the shorter queries, the results were mostly significant for all search queries. I judged the results based on how much the title and content matched the search query. In the below examples I will illustrate how in some cases a particular model gave better results as compared to the other models.

#### **Search Query - Abdul Kareem**

While searching for ‘Abdul Kareem’, we see that vector space gave results more centric to the player himself such as his stats, best moments etc as well as details about the Kareem Abdul Jabbar award, but in the case of HITS and PageRank, we saw that even though the results were not centric to the player himself, but it was relevant in the recent context as LeBron James recently passed the record of Abdul Kareem a few months back.

The screenshot shows a search interface with a query input field, a dropdown menu, and a search button. The results are categorized by model: Vector Space, PageRank, and HITS. Each category lists search results with titles and URLs.

**Results for: abdel kareem**

**Our Search Engine Results**

**Vector Space**

- [Kareem Abdul-Jabbar: College stats, best moments, quotes](#)  
https://www.ncaa.com/news/basketball-men/article/2020-05-04/kareem-abdul-jabbar-college-stats-best-moments-quotes?utm\_c...
- [Andy Katz shares his top 11 candidates for the Kareem Abdul-Jabbar award](#)  
https://www.ncaa.com/news/basketball-men/article/2019-06-05/andy-katz-shares-his-top-11-candidates-kareem-abdul-jabbar?utm\_c...
- [Basketball Hall of Fame names 20 centers to 2023 Kareem Abdul-Jabbar Award watch list](#)  
https://www.ncaa.com/news/basketball-men/article/2022-10-28/basketball-hall-fame-names-20-centers-2023-kareem-abdul-jabba...

**PageRank**

- [What LeBron passing Kareem means for NBA, GOAT debate](#)  
https://www.basketballnews.com/stories/what-lebron-passing-kareem-means-for-nba-goat-debate
- [LeBron James passes Kareem Abdul-Jabbar as NBA all-time scoring leader](#)  
https://www.basketballnews.com/stories/nba-breaking-news-lebron-james-passes-kareem-abduljabbar-as-nba-all-time-scoring-le...
- [What LeBron passing Kareem means to NBA, greatest debate](#)  
https://www.basketballnews.com/stories/what-lebron-passing-kareem-means-to-nba-greatest-debate

**HITS**

- [LeBron James breaks Kareem Abdul-Jabbar's NBA scoring record](#)  
https://www.espn.com/nba/story/\_/id/35611846/lakers-lebron-james-passes-kareem-abdul-jabbar-become-nba-all-scoring-leader
- [LeBron James passes Kareem Abdul-Jabbar for all-time FG playoff record](#)  
https://www.espn.com/nba/story/\_/id/23566498/cleveland-cavaliers-lebron-james-now-all-leader-postseason-field-goals
- [All-Time #NBArank: Kareem tops list of greatest centers ever](#)  
https://www.espn.com/nba/story/\_/page/nbarankCs/ranking-greatest-centers-nba-history

## Search Query - LeBron

In this particular context, all the models returned relevant results.

Type your query...

**Results for: lebron**

**Our Search Engine Results**

**Vector Space**

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)  
<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>
- [Nuggets' Michael Malone to coach Team LeBron in 2023 NBA All-Star Game](#)  
<https://www.basketballnews.com/stories/nuggets-michael-malone-to-coach-team-lebron-in-2023-nba-allstar-game>
- [Season of NBA moments: LeBron, Lillard, Mitchell and more](#)  
<https://www.basketballnews.com/stories/season-of-nba-moments-lebron-lillard-mitchell-and-more>

**PageRank**

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)  
<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>
- [LeBron James closes in on NBA record, Lakers falls to Pelicans](#)  
<https://www.basketballnews.com/stories/lebron-james-closes-in-on-nba-record-lakers-falls-to-pelicans>
- ['I'm not going anywhere': LeBron James plans to play 'a few more years'](#)  
<https://www.basketballnews.com/stories/im-not-going-anywhere-lebron-james-plans-to-play-a-few-more-years>

**HITS**

- [People have been hating on LeBron James all season](#)  
<https://www.basketballnews.com/stories/nba-rumors-lebron-james-los-angeles-lakers-hating-criticism-radio-tv>
- [LeBron, Giannis lead NBA jersey sales for second half of season](#)  
<https://www.basketballnews.com/stories/lebron-james-giannis-antetokounmpo-lead-jersey-sales-for-2nd-half-of-season>
- [LeBron James passes Wilt Chamberlain on all-time field goals made list](#)  
<https://www.basketballnews.com/stories/nba-news-lebron-james-all-time-field-goals-made-list-wilt-chamberlain-los-angeles-lakers>

## **Search Query - Chicago Bulls**

In this context, the vector space model returned bad results as the top three results referred to the college basketball team ‘Loyola Chicago’. However, the PageRank and HITS algorithm returned good results which were referring to the NBA team ‘Chicago Bulls’.

Type your query... --Choose one option-- Search

**Results for: *Chicago Bulls***

**Our Search Engine Results**

**Vector Space**

- [NCAA Video Vault: Cinderella Loyola Chicago's three straight game winners in 2018 March Madness](#)  
<https://www.ncaa.com/news/basketball-men/article/2022-01-11/cinderella-loyola-chicagos-three-straight-gam>
- [NCAA Video Vault: Cinderella Loyola Chicago's three straight game winners in 2018 March Madness](#)  
<https://www.ncaa.com/news/basketball-men/article/2022-01-11/cinderella-loyola-chicagos-three-straight-gam>
- [The world has fallen in love with Loyola Chicago's Sister Jean](#)  
<https://www.ncaa.com/news/basketball-men/article/2018-03-15/why-world-has-fallen-love-loyola-chicago-su>

**PageRank**

- [Derrick Jones Jr. re-signs with Chicago on two-year deal](#)  
<https://www.basketballnews.com/stories/derrick-jones-jr-resigns-with-chicago-on-a-twosear-deal>
- [Leonard, Powell lead Clippers past Bulls, 108-103](#)  
<https://www.basketballnews.com/stories/leonard-powell-lead-clippers-past-bulls-108103>
- [Zach LaVine scores 36 points, Bulls beat Grizzlies 128-107](#)  
<https://www.basketballnews.com/stories/zach-lavine-scores-36-points-bulls-beat-grizzlies-128107>

**HITS**

- [Zach LaVine scores 36 points, Bulls beat Grizzlies 128-107](#)  
<https://www.basketballnews.com/stories/zach-lavine-scores-36-points-bulls-beat-grizzlies-128107>
- [Dinwiddie has 25 in Brooklyn return, Nets beat Bulls 116-105](#)  
<https://www.basketballnews.com/stories/dinwiddie-has-25-in-brooklyn-return-nets-beat-bulls-116105>
- [Spurs, Blazers, Lakers have best odds to steal Zach LaVine from Bulls](#)  
<https://www.basketballnews.com/stories/spurs-blazers-lakers-have-best-odds-to-steal-zach-lavine-from-bulls>

## **Collaboration with Pujitha (clustering)**

I collaborated with Pujitha to understand how she wanted my results and I provided her the results through a list of dictionaries. I also understood how my results would be fed into the clustering algorithm to improve the search results. For the queries mentioned in my report as well as for other queries, we found that clustering provided better results in comparison to PageRank and HITS relevance models.

## 4. User Interface and Comparisons with Google and Bing

Technologies used: HTML, Flask, Jinja

For UI, I have used simple HTML to create the native-application for the search engine. In the UI design a query input field, a dropdown list is used to choose either Relevance model or Clustering or Query expansion, and a “Search” button is on top of the screen.

Once we enter a query, select a category, and click “Search”, the form is submitted as a POST request to the backend. I’ve used Flask for backend, since Relevance model, Clustering and Query expansion are developed using python, it would be easy to integrate. In response to the POST request, the results are generated based on query and model values. The generated results are then displayed on frontend using Jinja template along with Google search Results and Bing Search Results.

1. (Vector-Space, HITS, PageRank, Google, Bing) when we choose Relevance model from dropdown
2. (Flat Clustering, Single clustering, Complete clustering Google, Bing) when we choose Clustering from dropdown
3. (Expanded query and Query expansion results, Google, Bing) when we Choose query expansion from dropdown

For google search results, I have used Google Custom Search API which requires API key, search engine ID and a query. Once these parameters are passed on to the API through GET request, it returns a JSON response which has metadata describing requested search, metadata describing our search engine and search results. The search results are in the form of a list of dictionaries with key ‘items’. For bing search results, I have used Bing Web Search API which requires a subscription key and query as payload sent through GET request. With the request to API it would provide a JSON response. The JSON response has metadata related to the requested search and search results. The search results have a key named ‘value’ and is in the form of a list of dictionaries.

Since the results format is a list of dictionaries with google and bing, I discussed with Allama, who is responsible for Relevance model, Harish who is responsible for query expansion and Pujitha who is responsible for clustering about the output format. To maintain the consistency, we all agreed upon the output format as List of Dictionaries. Then We were able to simultaneously develop our individual parts.

The first step would be importing python modules written by Allama in Part 2. After importing, I have created a class object which has all methods required to generate results for relevance models. The search query is then passed on as input argument for the search method of the object. This will return a list of dictionaries with Title, and Link as keys for each list entry. Allama gave me 10 queries and I used 50 queries separately to test the communication between the User Interface and the API, and also to ensure that the API requests could handle many requests back to back.

I also discussed with Pujitha Penumarthy, who is responsible for Part 3, about how to integrate relevance model with clustering. After importing the python modules, I have created a class object with all clustering methods. With a given query, the pagerank results are generated from Part2. These results are then passed on to the clustering method. This method would generate clustering results for all the clustering outputs as a list of dictionaries with Title and Link as keys for each list entry.

I then discussed with Harish, who is responsible for Part 5 about how to integrate relevance model with query expansion. At first, I generated pagerank results with the given query from part-2 and the top relevant results were considered. I have created a class object with all query expansion methods, after importing the necessary python modules. The search query and relevant documents are passed on as input for this query expansion method. This generates an expanded query which is sent back to the relevance models in part 2 to provide appropriate results for the expanded query. The output will be a list of dictionaries with Title and Link as keys.

Once all modules are connected, we all generated search results for queries. We have listed common 50 queries and tested them. For demonstration we have chosen queries that we believe have been showing both accurate results and show improvement as we include clustering and query expansion.

The best queries that we are planning to execute are: ["abdul kareem", "kobe", "mvp", "lebron", "chicago", "chicago bulls", "giannis", "giannis bucks", "darvin ham", "laker", "nba", "fiba", "Karl Malone"].

## TESTING:

### Comparison with google and bing

For query - nba

### Relevance model Results:

Type your query... --Choose one option-- Search

Results for: nba

**Our Search Engine Results**

**Vector Space**

- College basketball's top 10 players in the NBA's 2020 Eastern Conference Finals  
[https://www.ncaa.com/news/basketball-men/article/2020-09-15/college-basketballs-top-10-players-nbas-2020-eastern-conference-finals?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2020-09-15/college-basketballs-top-10-players-nbas-2020-eastern-conference-finals?utm_campaign=inline-article)
- Ja Morant and R.J. Barrett are expected to be the No. 2 and No. 3 picks in the NBA Draft. Let's break down their college games  
<https://www.ncaa.com/news/basketball-men/article/2019-06-19/ja-morant-and-rj-barrett-are-expected-be-no-2-and-no-3-picks>
- Ja Morant: NBA draft projection, college highlights, statistics  
<https://www.ncaa.com/news/basketball-men/article/2019-06-20/ja-morant-nba-draft-projection-college-highlights-statistics>

**PageRank**

- LaMarcus Aldridge, 7-time NBA All-Star, announces retirement  
<https://www.basketballnews.com/stories/lamarcus-aldrige-7time-nba-allstar-announces-retirement>
- Grady Dick to leave Kansas for 2023 NBA Draft after one season  
<https://www.basketballnews.com/stories/grady-dick-to-leave-kansas-for-2023-nba-draft-after-one-season>
- LeBron James closes in on NBA record. Lakers falls to Pelicans  
<https://www.basketballnews.com/stories/lebron-james-closes-in-on-nba-record-lakers-falls-to-pelicans>

**HITS**

- NBA, players' union reach deal for a new labor agreement  
<https://www.basketballnews.com/stories/nba-players-union-reach-deal-for-a-new-labor-agreement>
- Malone, Rivers named NBA's Coaches of the Month for January  
<https://www.basketballnews.com/stories/malone-rivers-named-nbas-coaches-of-the-month-for-january>
- College basketball rankings: Andy Katz's Power 36 following the NBA draft  
[https://www.ncaa.com/news/basketball-men/article/2019-06-24/college-basketball-rankings-andy-katzs-power-36-following?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2019-06-24/college-basketball-rankings-andy-katzs-power-36-following?utm_campaign=inline-article)

### Clustering Results:

**Flat Clustering**

- LaMarcus Aldridge, 7-time NBA All-Star, announces retirement  
<https://www.basketballnews.com/stories/lamarcus-aldrige-7time-nba-allstar-announces-retirement>
- UCLA guard Jaylen Clark declares for 2023 NBA Draft  
<https://www.basketballnews.com/stories/ucla-guard-jaylen-clark-declares-for-2023-nba-draft>
- NBA roundtable: Which team desperately needs to make a trade?  
<https://www.basketballnews.com/stories/nba-roundtable-which-team-desperately-needs-to-make-a-trade>
- UConn guard Jordan Hawkins declares for 2023 NBA Draft  
<https://www.basketballnews.com/stories/uconn-guard-jordan-hawkins-declares-for-2023-nba-draft>
- It's Embiid vs. Doncic for scoring title in NBA's final week  
<https://www.basketballnews.com/stories/its-embiid-vs-doncic-for-scoring-title-in-nbas-final-week>

**Single Clustering**

- LaMarcus Aldridge, 7-time NBA All-Star, announces retirement  
<https://www.basketballnews.com/stories/lamarcus-aldrige-7time-nba-allstar-announces-retirement>
- Views from the bottom: 5 young NBA players to keep an eye on  
<https://www.basketballnews.com/stories/views-from-the-bottom-5-young-nba-players-to-keep-an-eye-on>
- Season of NBA moments: LeBron, Lillard, Mitchell and more  
<https://www.basketballnews.com/stories/season-of-nba-moments-lebron-lillard-mitchell-and-more>
- College basketball's top 10 players in the NBA's 2020 Eastern Conference Finals  
[https://www.ncaa.com/news/basketball-men/article/2020-09-15/college-basketballs-top-10-players-nbas-2020-eastern-conference-finals?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2020-09-15/college-basketballs-top-10-players-nbas-2020-eastern-conference-finals?utm_campaign=inline-article)
- NBA in Mexico: G League Capitanes bring the 've laying foundation  
<https://www.basketballnews.com/stories/nba-in-mexico-g-league-capitanes-hope-theyre-laying-foundation>

**Complete Clustering**

- LaMarcus Aldridge, 7-time NBA All-Star, announces retirement  
<https://www.basketballnews.com/stories/lamarcus-aldrige-7time-nba-allstar-announces-retirement>
- Views from the bottom: 5 young NBA players to keep an eye on  
<https://www.basketballnews.com/stories/views-from-the-bottom-5-young-nba-players-to-keep-an-eye-on>
- LaMarcus Aldridge, 7-time NBA All-Star, announces retirement  
<https://www.basketballnews.com/stories/lamarcus-aldrige-7time-nba-allstar-announces-retirement>
- Embiid wins NBA's scoring race, Harden takes assist title  
<https://www.basketballnews.com/stories/embiid-wins-nbas-scoring-race-harden-takes-assist-title>
- NBA Stats Notebook: Giving James Harden the flowers he deserves  
<https://www.basketballnews.com/stories/nba-stats-notebook-giving-james-harden-the-flowers-he-deserves-philladelphia-76ers-sixers-joel-embiid-tobias-harris>

## Query Expansion, Google & Bing Results:

**Google Search Results**

- [The official site of the NBA for the latest NBA Scores, Stats & News ...](#)  
<https://www.nba.com/>  
The official site of the National Basketball Association. Follow the action on NBA scores, schedules, stats, news, Team and Player news.
- [NBA \(@nba\) • Instagram photos and videos](#)  
<https://www.instagram.com/nba/?hl=en>  
79.1m Followers, 1118 Following, 59.4k Posts - See Instagram photos and videos from NBA (@nba)
- [2023 NBA Playoffs | Home | NBA.com](#)  
<https://www.nba.com/playoffs/2023>  
LeBron James comes up clutch in the 4th quarter and OT to power a Game 4 win and a 3-1 series lead on Memphis. in 8 hours - Inside the NBA: Anthony Davis' peaks ...

**Bing Search Results**

- [National Basketball Association - Official Site](#)  
<http://www.nba.com/>  
The official site of the National Basketball Association. Your home for scores, schedules, stats, League Pass, video recaps, news, fantasy, rankings and more for NBA players and teams.
- [NBA on ESPN - Scores, Stats and Highlights](#)  
<https://www.espn.com/nba/>  
Visit ESPN for NBA live scores, video highlights and latest news. Stream games on ESPN and play Fantasy Basketball.
- [NBA Basketball Scores - NBA Scoreboard - ESPN](#)  
<https://www.espn.com/nba/scoreboard>  
Real-time NBA Basketball scores on ESPN. Former Celtics coach Ime Udoka has agreed to a deal to become the next Houston Rockets head coach, sources told ESPN's Adrian Wojnarowski on Monday.

For query - fiba

## Relevance model Results:

**Our Search Engine Results**

**Vector Space**

- [FIBA EuroBasket 2022: Top teams, players and prospects to watch](#)  
<https://www.basketballnews.com/stories/op-teams-players-prospects-to-watch-fiba-eurobasket-2022-giannis-antetokounmpo-nikola-jokic-luka-doncic>
- [Jokic vs Giannis: Serbia defeats Greece in FIBA World Cup Qualifier](#)  
<https://www.basketballnews.com/stories/jokic-vs-giannis-serbia-defeats-greece-in-fiba-world-cup-qualifier>
- [2022 FIBA AmeriCup schedule released](#)  
<https://www.basketballnews.com/stories/the-2022-fiba-americup-schedule-has-been-released>

**PageRank**

- [Meet Quincy Diggins: FIBA's top-ranked American 3x3 player](#)  
<https://www.basketballnews.com/stories/fiba-international-basketball-meet-quincy-diggs-top-ranked-united-states-american-3x3-player-team-vienna-austria-akron-zips-keith-dambrot-lebron-james-redwell-oberwart-gunners>
- [Things to monitor as FIBA World Cup's next qualifying window arrives](#)  
<https://www.basketballnews.com/stories/things-to-monitor-as-the-next-fiba-world-cups-next-qualifying-window-arrives>
- [Joe Johnson proving he can still get buckets in FIBA Americup Qualifiers](#)  
<https://www.basketballnews.com/stories/nba-fiba-americup-joe-johnson-isaiah-thomas-detroit-pistons-free-agent-qualifiers>

**HITS**

- [FIBA All-Stars: Gasol, Giannis and the very best NBA players at the World Cup](#)  
[https://www.espn.com/nba/story/\\_/id/27620210/fiba-all-stars-gasol-giannis-very-best-nba-players-world-cup](https://www.espn.com/nba/story/_/id/27620210/fiba-all-stars-gasol-giannis-very-best-nba-players-world-cup)
- [Juan Palacios hopes to guide Colombia to FIBA World Cup](#)  
<https://www.basketballnews.com/stories/international-basketball-juan-palacios-colombia-mens-national-team-fiba-world-cup-rick-pitino-louisville-cardinals>
- [FIBA World Cup Preview: Breaking down all 32 teams](#)  
[https://www.espn.com/nba/story/\\_/id/27489783/fiba-world-cup-preview-breaking-all-32-teams](https://www.espn.com/nba/story/_/id/27489783/fiba-world-cup-preview-breaking-all-32-teams)

## Clustering Results:

### Flat Clustering

- Meet Quincy Diggins, FIBA's top-ranked American 3x3 player  
<https://www.basketballnews.com/stories/fiba-international-basketball-meet-quincy-diggins-top-ranked-united-states-american-3x3-player-team-vienna-austria-akron-zips-keith-dambrot-lebron-james-redwell-oberwart-gummers>
- Things to monitor as FIBA World Cup's next qualifying window arrives  
<https://www.basketballnews.com/stories/things-to-monitor-as-the-next-fiba-world-cups-next-qualifying-window-arrives>
- Joe Johnson proving he can still get buckets in FIBA Americup Qualifiers  
<https://www.basketballnews.com/stories/nba-fiba-americup-joe-johnson-isaiah-thomas-detroit-pistons-free-agent-qualifiers>
- Ben Simmons in exhibition games but not FIBA World Cup  
[https://www.espn.com/nba/story/\\_/id/27204354/ben-simmons-exhibition-games-not-fiba-world-cup](https://www.espn.com/nba/story/_/id/27204354/ben-simmons-exhibition-games-not-fiba-world-cup)
- FIBA EuroBasket 2022: Top teams, players and prospects to watch  
<https://www.basketballnews.com/stories/top-teams-players-prospects-to-watch-fiba-eurobasket-2022-giannis-anterokounmpo-nikola-jokic-luka-doncic>

### Single Clustering

- Meet Quincy Diggins, FIBA's top-ranked American 3x3 player  
<https://www.basketballnews.com/stories/fiba-international-basketball-meet-quincy-diggins-top-ranked-united-states-american-3x3-player-team-vienna-austria-akron-zips-keith-dambrot-lebron-james-redwell-oberwart-gummers>
- FIBA U17 Basketball World Cup results: USA takes home Gold over Spain  
<https://www.sportingnews.com/us/nba/news/fiba-u17-basketball-world-cup-results-dates-time-schedule-how-to-watch/hxek4tzbfbqmkamlvr7bu>
- A'Ja Wilson's FIBA World Cup dominance caps off a remarkable year  
<https://www.sportingnews.com/us/nba/news/a-jay-wilson-fiba-world-cup-dominance/h62amylarfe2pjyhc2verj>
- FIBA EuroBasket 2022 rosters: Here are the full rosters for all 24 teams  
<https://www.sportingnews.com/us/nba/news/fiba-eurobasket-full-rosters-2022-sac5h2bemhlc5ylnthgqa>
- FIBA Women's Basketball World Cup 2022: USA break World Cup record with 145 point outing  
<https://www.sportingnews.com/us/nba/news/fiba-womens-basketball-world-cup-2022-usa-break-world-cup-record-145-point-thrashing/jhsjt6gqfwdnnmflktivzr>

### Complete Clustering

- Meet Quincy Diggins, FIBA's top-ranked American 3x3 player  
<https://www.basketballnews.com/stories/fiba-international-basketball-meet-quincy-diggins-top-ranked-united-states-american-3x3-player-team-vienna-austria-akron-zips-keith-dambrot-lebron-james-redwell-oberwart-gummers>
- FIBA Women's Basketball World Cup 2022: Quarter-finals set as USA look unstoppable Australia tops 'Group of Death'  
<https://www.sportingnews.com/us/nba/news/fiba-womens-basketball-world-cup-2022-quarter-finals-set-as-usa-look-unstoppable-australia-tops-group-death/lghqfhkce4q2pd81smhhfi9>
- FIBA World Cup Roundup Day 6: Greece, Australia, Czech Republic Get Big Wins  
<http://global.nba.com/news/fiba-world-cup-roundup-day-6-greece-australia-czech-republic-get-big-wins/>
- Meet Quincy Diggins, FIBA's top-ranked American 3x3 player  
<https://www.basketballnews.com/stories/fiba-international-basketball-meet-quincy-diggins-top-ranked-united-states-american-3x3-player-team-vienna-austria-akron-zips-keith-dambrot-lebron-james-redwell-oberwart-gummers>
- Isiah Thomas scores 19 points in FIBA qualifying game, U.S. wins  
<https://www.basketballnews.com/stories/nba-news-isaiyah-thomas-fiba-americup-team-usa-debut-win>

## Query Expansion, Google & Bing Results:

### Our Search Engine Results

#### Query expansion

- fiba world team win
- Miami basketball is seeing double as men's and women's teams advance to the Elite Eight on the same day.  
[https://www.ncaa.com/news/basketball-men/article/2023-03-24/miami-basketball-seeing-double-mens-and-womens-teams-advance-elite-eight-same?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2023-03-24/miami-basketball-seeing-double-mens-and-womens-teams-advance-elite-eight-same?utm_campaign=inline-article)
  - Duke knocks Texas Tech from undefeated list, but both teams in position for tournament runs  
<https://www.ncaa.com/news/basketball-men/article/2018-12-20/duke-texas-tech-basketball-score-new-york>
  - Tennessee's Grant Williams wins March Madness Preseason Player of the Year  
<https://www.ncaa.com/news/basketball-men/article/2018-10-26/tennessees-grant-williams-wins-march-madness-preseason>

### Google Search Results

- FIBA basketball: International Basketball Federation (FIBA)  
<https://www.fiba.basketball/>  
 FIBA, International Basketball Federation, is the world governing body for basketball. Founded in 1932, we bring together 212 National Basketball ...
- FIBA (@FIBA) | Twitter  
<https://twitter.com/FIBA>  
 FIBA @FIBA. Official account of the International Basketball Federation Home Of Hoops. linktr.ee/fiba Born June 18 Joined April 2009.
- FIBA | Financial & International Business Association  
<https://fibainc.net/>  
 FIBA is a non-profit trade association dedicated to cultivating excellence in International Banking. CONTACT US: 80 SW 8th St #2170, Miami, FL 33130. FIBA ...

### Bing Search Results

- International Basketball Federation (FIBA) - FIBA basketball  
<https://www.fiba.basketball/>  
 The official website of FIBA, the International Basketball Federation, and the governing body of Basketball. FIBA organises the most famous and prestigious international basketball competitions including the FIBA Basketball World Cup, the FIBA World Championship for Women and the FIBA 3x3 World Tour. The webs 70,00 players profile ...
- FIBA Live Scores - FIBA basketball  
<https://www.fiba.basketball/live-scores>  
 The official website of FIBA, the International Basketball Federation, and the governing body of Basketball. FIBA organises the most famous and prestigious international basketball competitions including the FIBA Basketball World Cup, the FIBA World Championship for Women and the FIBA 3x3 World Tour. The webs 70,00 players profile ...
- FIBA - Wikipedia  
<https://en.wikipedia.org/wiki/FIBA>  
 FIBA headquarters in Mies, Switzerland. The International Basketball Federation (FIBA / fiba / FEE-ba, French: Fédération internationale de basket-ball) is an association of national organizations which governs the sport of basketball worldwide. Originally known as the Fédération internationale de basket-ball amateur (

For query - Karl Malone

Relevance model Results:

Type your query...

**Results for: Karl Malone**

**Our Search Engine Results**

**Vector Space**

- [Growing up a Buck- George Karl arrives in Milwaukee](https://www.basketballnews.com/stories/growing-up-a-buck-george-karl-arrives-in-milwaukee)  
https://www.basketballnews.com/stories/growing-up-a-buck-george-karl-arrives-in-milwaukee
- [Nuggets' Michael Malone to coach Team LeBron in 2023 NBA All-Star Game](https://www.basketballnews.com/stories/nuggets-michael-malone-to-coach-team-lebron-in-2023-nba-all-star-game)  
https://www.basketballnews.com/stories/nuggets-michael-malone-to-coach-team-lebron-in-2023-nba-allstar-game
- [Malone, Rivers named NBA's Coaches of the Month for January](https://www.basketballnews.com/stories/malone-rivers-named-nbas-coaches-of-the-month-for-january)  
https://www.basketballnews.com/stories/malone-rivers-named-nbas-coaches-of-the-month-for-january

**PageRank**

- [Mike Malone, Jamal Murray, Michael Porter Jr. won't be back soon](https://www.basketballnews.com/stories/malone-jamal-murray-michael-porter-jr-wont-be-back-soon)  
https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon
- [Mike Malone: Nikola Jokic 'should be an All-Defensive player'](https://www.basketballnews.com/stories/mike-malone-nikola-jokic-should-be-an-all-defensive-player)  
https://www.basketballnews.com/stories/mike-malone-nikola-jokic-should-be-an-alldefensive-player
- [Mike Malone tests positive for COVID, enters protocols](https://www.basketballnews.com/stories/mike-malone-tests-positive-for-covid-enters-protocols)  
https://www.basketballnews.com/stories/mike-malone-tests-positive-for-covid-enters-protocols

**HITS**

- [Malone: Coaches should have family in bubble](https://www.basketballnews.com/stories/nba-rumors-coaching-staff-families-bubble-rules-denver-nuggets)  
https://www.basketballnews.com/stories/nba-rumors-coaching-staff-families-bubble-rules-denver-nuggets
- [Basketball Hall of Fame names 20 power forwards to watch list for 2022 Karl Malone Award](https://www.ncaa.com/news/basketball-men/article/2021-10-21/basketball-hall-fame-names-20-power-forwards-watch-list-2022-karl-malone-award?utm_campaign=inline-article)  
https://www.ncaa.com/news/basketball-men/article/2021-10-21/basketball-hall-fame-names-20-power-forwards-watch-list-2022-karl-malone-award?utm\_campaign=inline-article
- [25 contenders for the Karl Malone Award as the nation's best power forward](https://www.ncaa.com/news/basketball-men/article/2020-11-03/25-contenders-karl-malone-award-nations-best-power-forward)  
https://www.ncaa.com/news/basketball-men/article/2020-11-03/25-contenders-karl-malone-award-nations-best-power-forward

Clustering Results:

**Flat Clustering**

- [Mike Malone, Jamal Murray, Michael Porter Jr. won't be back soon](https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon)  
https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon
- [Mike Malone: Nikola Jokic 'should be an All-Defensive player'](https://www.basketballnews.com/stories/mike-malone-nikola-jokic-should-be-an-all-defensive-player)  
https://www.basketballnews.com/stories/mike-malone-nikola-jokic-should-be-an-alldefensive-player
- [Mike Malone tests positive for COVID, enters protocols](https://www.basketballnews.com/stories/mike-malone-tests-positive-for-covid-enters-protocols)  
https://www.basketballnews.com/stories/mike-malone-tests-positive-for-covid-enters-protocols
- [Malone: Coaches should have family in bubble](https://www.basketballnews.com/stories/nba-rumors-coaching-staff-families-bubble-rules-denver-nuggets)  
https://www.basketballnews.com/stories/nba-rumors-coaching-staff-families-bubble-rules-denver-nuggets
- [Denver Nuggets coach Michael Malone says he had coronavirus in March](https://www.espn.com/nba/story/_/id/29316542/denver-nuggets-coach-michael-malone-says-had-coronavirus-march)  
https://www.espn.com/nba/story/\_/id/29316542/denver-nuggets-coach-michael-malone-says-had-coronavirus-march

**Single Clustering**

- [Mike Malone, Jamal Murray, Michael Porter Jr. won't be back soon](https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon)  
https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon
- [Damian Lillard & his \\$100+ million Adidas contract? offends ringless Barry Sanders, Chris Webber & Karl Malone | Footlocker Ad](https://www.basketballlife.com/damian-lillard-his-100-million-adidas-contract-offends-ringless-barry-sanders-chris-webber-karl-malone-footlocker-ad)  
https://www.basketballlife.com/damian-lillard-his-100-million-adidas-contract-offends-ringless-barry-sanders-chris-webber-karl-malone-footlocker-ad
- [Remembering When Karl Malone Got Dunked On By The Only Player He Would Pay To See - Vince Carter](https://www.basketballlife.com/remembering-when-karl-malone-got-dunked-on-by-the-only-player-he-would-pay-to-see-vince-carter/)  
https://www.basketballlife.com/remembering-when-karl-malone-got-dunked-on-by-the-only-player-he-would-pay-to-see-vince-carter/
- [Anthony Mason Could Guard Michael Jordan, Karl Malone & Hakeem The Dream](https://www.basketballlife.com/anthony-mason-could-guard-michael-jordan-karl-malone-hakeem-the-dream/)  
https://www.basketballlife.com/anthony-mason-could-guard-michael-jordan-karl-malone-hakeem-the-dream/
- [Kevin Durant talks about his friendship with mentor Karl Malone](https://www.basketballlife.com/kevin-durant-talks-about-his-friendship-with-mentor-karl-malone/)  
https://www.basketballlife.com/kevin-durant-talks-about-his-friendship-with-mentor-karl-malone/

**Complete Clustering**

- [Mike Malone, Jamal Murray, Michael Porter Jr. won't be back soon](https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon)  
https://www.basketballnews.com/stories/malone-michael-porter-jr-and-jamal-murrays-wont-be-back-soon
- [Nuggets Coach Michael Malone Wants His Family Allowed In The Bubble](https://basketballforever.com/2020/09/04/nuggets-coach-michael-malone-wants-his-family-allowed-in-the-bubble)  
https://basketballforever.com/2020/09/04/nuggets-coach-michael-malone-wants-his-family-allowed-in-the-bubble
- [Mike Malone: Nikola Jokic 'should be an All-Defensive player'](https://www.basketballnews.com/stories/mike-malone-nikola-jokic-should-be-an-all-defensive-player)  
https://www.basketballnews.com/stories/mike-malone-nikola-jokic-should-be-an-alldefensive-player
- [Moses Malone Jr. Sees James Harden For "Masterminding" A Robbery](https://www.basketballforever.com/2017/05/16/moses-malone-jr-sees-james-harden-masterminding-robery)  
https://www.basketballforever.com/2017/05/16/moses-malone-jr-sees-james-harden-masterminding-robery
- [Three-time NBA MVP Moses Malone dies at age 60](https://www.espn.com/nba/story/_/id/13645531/three-time-nba-mvp-moses-malone-dies-age-60)  
https://www.espn.com/nba/story/\_/id/13645531/three-time-nba-mvp-moses-malone-dies-age-60

## Query Expansion, Google & Bing Results:

### Query expansion

Karl Malone coach michael nba

- [Hall of Fame coach Lute Olson dies at 85](#)  
<https://www.ncaa.com/news/basketball-men/article/2020-08-28/hall-fame-coach-lute-olson-dies-85>
- [The men's college basketball coaches best at developing talent, ranked by Andy Katz](#)  
[https://www.ncaa.com/news/basketball-men/article/2022-09-28/mens-college-basketball-coaches-best-developing-talent-ranked-andy-katz?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2022-09-28/mens-college-basketball-coaches-best-developing-talent-ranked-andy-katz?utm_campaign=inline-article)
- [Alabama coach Nate Oats gets new 6-year, \\$30 million deal](#)  
[https://www.basketballnewsws.com/stories/alabama-coach-nate-oats-gets-new-6year-30-million-deal](https://www.basketballnewsws.com/stories/alabama-coach-nate-oats-gets-new-6-year-30-million-deal)

### Google Search Results

- [Karl Malone - Wikipedia](#)

[https://en.wikipedia.org/wiki/Karl\\_Malone](https://en.wikipedia.org/wiki/Karl_Malone)

Karl Anthony Malone (born July 24, 1963) is an American former professional basketball player in the National Basketball Association (NBA).

- [Karl Malone Stats, Height, Weight, Position, Draft Status and more ..](#)

<https://www.basketball-reference.com/players/m/malonka01.html>

Karl Anthony Malone : Twitter: TheDeliverer\_32 ; Position: Power Forward - Shoots: Right ; Born: July 24, 1963 in Summerfield, Louisiana us ; College: Louisiana ...

- [Karl Malone Stats, News, Bio | ESPN](#)

[https://www.espn.com/nba/player/\\_id/501/karl-malone](https://www.espn.com/nba/player/_id/501/karl-malone)

KarlMalone · Birthdate: 7/24/1963 · College: Louisiana Tech · Draft Info: 1985: Rd 1, Pk 13 (UTAH).

### Bing Search Results

- [Karl Malone - Wikipedia](#)

[https://en.wikipedia.org/wiki/Karl\\_Malone](https://en.wikipedia.org/wiki/Karl_Malone)

Karl Anthony Malone (born July 24, 1963) [1] is an American former professional basketball player in the National Basketball Association (NBA). Nicknamed "the Mailman", he is considered one of the greatest power forwards in NBA history. [2]

- [Karl Malone | Biography & Accomplishments | Britannica](#)

<https://www.britannica.com/biography/Karl-Malone>

Karl Malone, in full Karl Anthony Malone, byname the Mailman, (born July 24, 1963, Summerfield, Louisiana, U.S.), American basketball player who owns the National Basketball Association (NBA) career record for free throws attempted (13,188) and made (9,787).

- [Karl Malone Stats, Height, Weight, Position, Draft Status and more ..](#)

<https://www.basketball-reference.com/players/m/malonka01.html>

Karl Malone Karl Anthony Malone Twitter : TheDeliverer\_32 (The Mailman, King Karl) Position: Power Forward Shoots: Right 6-9 , 250lb (206cm, 113kg) More bio, uniform, draft, salary info Hall of Fame 14x All Star 14x All-NBA 1985-86 All-Rookie 4x All-Defensive 2x AS MVP 2x MVP NBA 75th Ar

In comparison with google and bing, our results are average. Our search engine does give similar results for some queries with Google/Bing. The major factor that is affecting the performance is the usage of limited amounts of data. Also our search engine is taking a lot of time to generate results with large datasets. So we did realize the limited capacities of local machines in terms of processing speed and memory.

## Conclusion:

Based on the results we have generated. Clustering gives better results compared to page rank and HITS relevance models

## **5. Clustering**

The process of clustering involves categorizing a collection of documents into distinct clusters or subsets. In this part, I implemented the different types of clustering algorithm:

1. Flat Clustering
2. Single Linkage Agglomerative clustering
3. Complete Linkage Agglomerative clustering

### **Flat Clustering**

Flat clustering is a type of clustering method that aims to assign individual data points in a dataset to a particular cluster based on a selected similarity metric.

K-means and DB scan are some of the examples of Flat Clustering algorithms.

### **K-means**

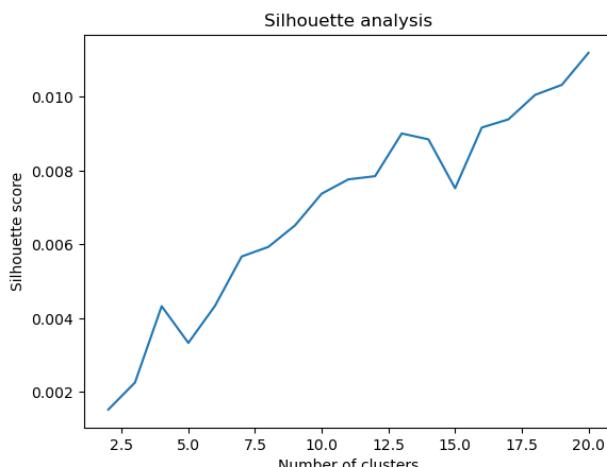
In K-Means algorithm, the primary goal is to reduce the total distance between each data point and its corresponding centroid in order to achieve optimal clustering. It is to be understood that less variation within the clusters will lead to more similar data points within the same cluster.

### **Selection of number of predefined clusters K**

To implement the algorithm effectively, it is crucial to determine the most appropriate number of clusters into which the data can be grouped. Out of the Elbow method and Silhouette method I chose the silhouette method to find the optimal number of clusters as it gives better results than the elbow method.

Silhouette method measures the distance between points in one cluster to points in another cluster known as silhouette coefficients. Coefficient will be in the range of -1 to 1. If the value is -1, the point is very far from another cluster. If it has a value of 0, it implies that it is situated precisely on or in close proximity to the decision boundary that separates two adjacent clusters.

I run the silhouette method for k values ranging from 2 to 20. The results are plotted in the graph below:



K value of 20 is selected as the silhouette coefficient is nearly equal to 0.

### Implementation of K Means

- I implemented kmeans clustering on 60,000 web crawled pages data which was stored in json format.
- The URLs and content of the data were extracted in order to carry out the clustering process.
- The content of the documents was vectorized using Sklearn's TfidfVectorizer function.
- The Pre-defined kmeans function of sklearn's library is used to compute for 20 clusters .
- The result of Kmeans is stored in a csv file with URL and corresponding cluster label.

Below is the sample of kmeans result stored in csv file

link	cluster
<a href="https://en.wikipedia.org/wiki/Ed_Kweller">https://en.wikipedia.org/wiki/Ed_Kweller</a>	4
<a href="https://en.wikipedia.org/wiki/Terry_Furlow">https://en.wikipedia.org/wiki/Terry_Furlow</a>	4
<a href="https://en.wikipedia.org/wiki/Lawrence_Funderburke">https://en.wikipedia.org/wiki/Lawrence_Funderburke</a>	4
<a href="https://en.wikipedia.org/wiki/Homer_Fuller">https://en.wikipedia.org/wiki/Homer_Fuller</a>	4
<a href="https://en.wikipedia.org/wiki/Todd_Fuller">https://en.wikipedia.org/wiki/Todd_Fuller</a>	4
<a href="https://en.wikipedia.org/wiki/Markelle_Fultz">https://en.wikipedia.org/wiki/Markelle_Fultz</a>	4
<a href="https://en.wikipedia.org/wiki/Mike_Lynn_(basketball)">https://en.wikipedia.org/wiki/Mike_Lynn_(basketball)</a>	9
<a href="https://en.wikipedia.org/wiki/Bob_Lytle">https://en.wikipedia.org/wiki/Bob_Lytle</a>	4
<a href="https://en.wikipedia.org/wiki/Larry_Robinson_(basketball,_born_1968)">https://en.wikipedia.org/wiki/Larry_Robinson_(basketball,_born_1968)</a>	9
<a href="https://en.wikipedia.org/wiki/Justin_Robinson_(basketball,_born_1997)">https://en.wikipedia.org/wiki/Justin_Robinson_(basketball,_born_1997)</a>	9
<a href="http://global.nba.com/news/daniel-theis-medical-update/">http://global.nba.com/news/daniel-theis-medical-update/</a>	4
<a href="https://en.wikipedia.org/wiki/John_O%27Boyle">https://en.wikipedia.org/wiki/John_O%27Boyle</a>	4
<a href="https://en.wikipedia.org/wiki/Bob_O%27Brien_(basketball)">https://en.wikipedia.org/wiki/Bob_O%27Brien_(basketball)</a>	9
<a href="https://en.wikipedia.org/wiki/Pete_Myers">https://en.wikipedia.org/wiki/Pete_Myers</a>	4
<a href="https://en.wikipedia.org/wiki/Jaren_Jackson_Jr.">https://en.wikipedia.org/wiki/Jaren_Jackson_Jr.</a>	4
<a href="https://en.wikipedia.org/wiki/Jack_Ozburn">https://en.wikipedia.org/wiki/Jack_Ozburn</a>	4
<a href="https://en.wikipedia.org/wiki/Dick_Nemelka">https://en.wikipedia.org/wiki/Dick_Nemelka</a>	4
<a href="https://en.wikipedia.org/wiki/Jerrod_Mustaf">https://en.wikipedia.org/wiki/Jerrod_Mustaf</a>	4
<a href="https://en.wikipedia.org/wiki/Sviatoslav_Mykhaliiuk">https://en.wikipedia.org/wiki/Sviatoslav_Mykhaliiuk</a>	4
<a href="https://en.wikipedia.org/wiki/Angelo_Musi">https://en.wikipedia.org/wiki/Angelo_Musi</a>	4
<a href="https://en.wikipedia.org/wiki/Toure%27_Murry">https://en.wikipedia.org/wiki/Toure%27_Murry</a>	4
<a href="http://global.nba.com/news/10-most-intriguing-free-agents-ahead-of-2020-21-season/">http://global.nba.com/news/10-most-intriguing-free-agents-ahead-of-2020-21-season/</a>	19
<a href="https://en.wikipedia.org/wiki/Martin_M%C3%BCller">https://en.wikipedia.org/wiki/Martin_M%C3%BCller</a>	4

[https://en.wikipedia.org/wiki/Mike\\_Muscala](https://en.wikipedia.org/wiki/Mike_Muscala)

4

### **Integration of clustering result with relevance model and displaying on UI**

- I worked with Allama Hossain to get the top 20,000 results from the pagerank relevance model and worked with Vaishnavi Kunduru to pass the output of pagerank relevance model (i.e first 20k data) as argument while calling the flat clustering function in Clustering class.
- I integrated both clustering and 10 top relevance model results in such a way that results are organized in a manner where the top-ranked result is part of a cluster, and all other results within the same cluster are subsequently ranked based on their page rank in decreasing order.
- I took the next top ranked result and their clusters pages are displayed in decreasing order until all the results of the relevance model are iterated.
- I returned the output in the list of dictionaries where each dictionary has the title , description and link as keys.
- Vaishnavi Kunduru handled the output in UI such a way that whenever we search a query and select clustering from the dropdown , top 5 documents with link and title are displayed under Flat Clustering title.

### **Agglomerative Clustering**

Among the hierarchical clustering methods, Agglomerative Clustering is widely used. It is also called Bottom up clustering as we will start with small clusters and combine them to generate larger clusters. It is very useful when the number of clusters is not known beforehand .

The algorithm initially considers each data point as an individual cluster. It then merges the two closest clusters using a distance metric, and this process is iteratively repeated until only one cluster remains.

Two methods of Agglomerative clustering are implemented:

1. Single linkage Agglomerative clustering : The distance metric used to measure the distance between two clusters is the minimum distance between any two points belonging to different clusters.
2. Complete linkage Agglomerative clustering : The distance metric used to measure the distance between two clusters is the longest distance between any two points belonging to different clusters.

### **Implementation of Single and Complete linkage Algorithm**

- I implemented single and complete clustering on 20,000 web crawled pages data which was stored in json format .Unable to run for larger data due to the memory limitations and system crashes.
- The URLs and content of the data were extracted in order to carry out the clustering process.
- The content of the documents was vectorized using Sklearn's TfidfVectorizer function.
- Apply cosine similarity function for vector output using the sklearn's library .

- Fastcluster function is used for both single and complete linkage algorithms as it dynamically calculates the number of clusters required.
- Single Linkage : firstly i calculated using linkage function with arguments as single method and euclidean metrics . Later the single function is executed and passed into dendrogram.
- Complete Linkage : i calculated using linkage function with arguments as complete method and euclidean metrics . Later the complete function is executed and passed into dendrogram.
- By using dendrogram it will automatically calculate the number of clusters by observing the number of vertical lines that intersect the horizontal line determines the ideal number of clusters.
- The result of single and complete linkage algorithm is stored in a csv file with URL and corresponding cluster label.

Below is the sample of Single Linkage result stored in csv file

link	cluster
<a href="http://global.nba.com/news/gary-payton-v-horacio-llamas-de-paseo-por-cdmx/">http://global.nba.com/news/gary-payton-v-horacio-llamas-de-paseo-por-cdmx/</a>	2.0
<a href="http://global.nba.com/news/suns-named-one-of-nbas-biggest-offseason-winners/">http://global.nba.com/news/suns-named-one-of-nbas-biggest-offseason-winners/</a>	2.0
<a href="http://global.nba.com/news/postgame-report-magic-vs-sixers-12016/">http://global.nba.com/news/postgame-report-magic-vs-sixers-12016/</a>	2.0
<a href="http://global.nba.com/news/hawks-face-challenges-amidst-busy-november-slate/">http://global.nba.com/news/hawks-face-challenges-amidst-busy-november-slate/</a>	2.0
<a href="http://global.nba.com/news/charlottes-walker-replace-new-yorks-porzingis-team-lebron-2018-nba-star-game/">http://global.nba.com/news/charlottes-walker-replace-new-yorks-porzingis-team-lebron-2018-nba-star-game/</a>	2.0
<a href="http://global.nba.com/news/nba-playoffs-first-round-schedule-2/">http://global.nba.com/news/nba-playoffs-first-round-schedule-2/</a>	2.0
<a href="http://global.nba.com/news/how-kyle-korver-is-making-long-twos-sexy-again/">http://global.nba.com/news/how-kyle-korver-is-making-long-twos-sexy-again/</a>	2.0
<a href="http://global.nba.com/news/shootaround-29-january-report-kings-open-to-trading-gay/">http://global.nba.com/news/shootaround-29-january-report-kings-open-to-trading-gay/</a>	2.0
<a href="http://global.nba.com/news/russell-catches-fire-in-overtime-loss/">http://global.nba.com/news/russell-catches-fire-in-overtime-loss/</a>	2.0
<a href="http://global.nba.com/news/blogtable-top-international-newcomer/">http://global.nba.com/news/blogtable-top-international-newcomer/</a>	2.0
<a href="http://global.nba.com/news/green-involved-in-yet-another-playoff-dustup-2/">http://global.nba.com/news/green-involved-in-yet-another-playoff-dustup-2/</a>	2.0
<a href="http://global.nba.com/news/hornets-kidd-gilchrist-out-for-remainder-of-season-with-torn-labrum/">http://global.nba.com/news/hornets-kidd-gilchrist-out-for-remainder-of-season-with-torn-labrum/</a>	2.0
<a href="http://global.nba.com/news/2018-19-nba-team-previews-northwest-division/">http://global.nba.com/news/2018-19-nba-team-previews-northwest-division/</a>	2.0
<a href="http://global.nba.com/news/dwyane-wade-impressed-with-growth-talent-of-76ers-ben-simmons/">http://global.nba.com/news/dwyane-wade-impressed-with-growth-talent-of-76ers-ben-simmons/</a>	2.0
<a href="http://global.nba.com/news/thunder-vs-minnesota-timberwolves-game-recap-jan-15-2016/">http://global.nba.com/news/thunder-vs-minnesota-timberwolves-game-recap-jan-15-2016/</a>	2.0
<a href="http://global.nba.com/news/power-rankings-celebs-jazz-rolling-into-the-break-2/">http://global.nba.com/news/power-rankings-celebs-jazz-rolling-into-the-break-2/</a>	2.0
<a href="http://global.nba.com/news/if-theres-a-magic-fix-for-pistons-its-getting-the-ball-hopping-on-offense-svg-says/">http://global.nba.com/news/if-theres-a-magic-fix-for-pistons-its-getting-the-ball-hopping-on-offense-svg-says/</a>	2.0
<a href="http://global.nba.com/news/jazz-look-for-road-upset-over-warriors/">http://global.nba.com/news/jazz-look-for-road-upset-over-warriors/</a>	2.0
<a href="http://global.nba.com/news/pistons-open-camp-in-much-better-space/">http://global.nba.com/news/pistons-open-camp-in-much-better-space/</a>	2.0
<a href="http://global.nba.com/news/2-doors-down-the-bond-between-allan-houstons-family-and-muhammad-ali/">http://global.nba.com/news/2-doors-down-the-bond-between-allan-houstons-family-and-muhammad-ali/</a>	2.0
<a href="http://global.nba.com/news/pacers-guard-rodney-stuckey-finds-comfort-zone/">http://global.nba.com/news/pacers-guard-rodney-stuckey-finds-comfort-zone/</a>	2.0
<a href="http://global.nba.com/news/howards-23-points-lead-hornets-past-thunder/">http://global.nba.com/news/howards-23-points-lead-hornets-past-thunder/</a>	2.0
<a href="http://global.nba.com/news/pro%e2%80%a2line-watch-november-21/">http://global.nba.com/news/pro%e2%80%a2line-watch-november-21/</a>	2.0
<a href="http://global.nba.com/news/nuggets-visit-celestics-clash-banged-clubs/">http://global.nba.com/news/nuggets-visit-celestics-clash-banged-clubs/</a>	2.0

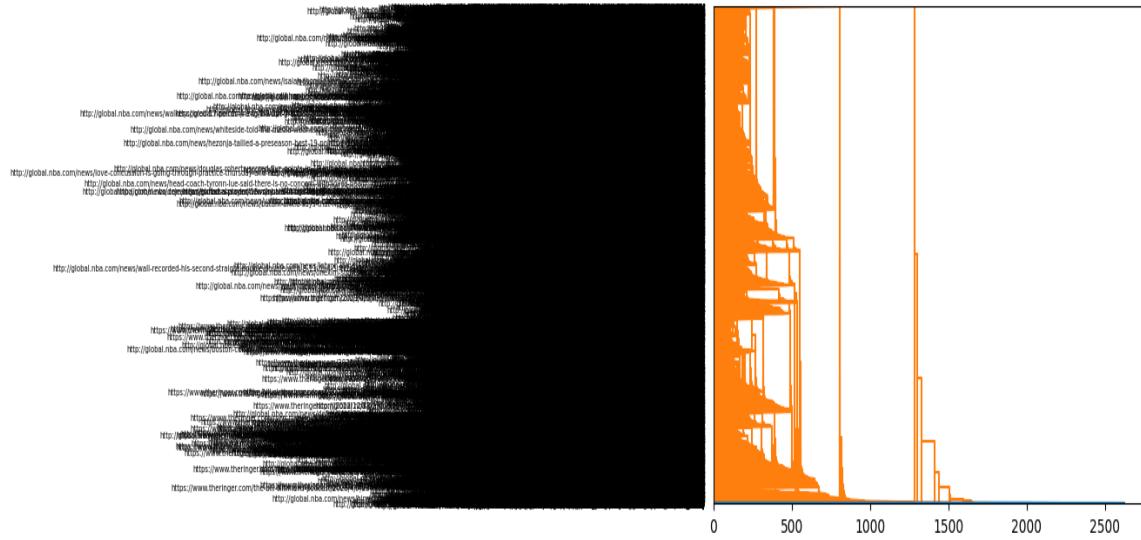
Below is the sample of Complete Linkage result stored in csv file

link	cluster
<a href="http://global.nba.com/news/burke-leads-knicks-past-magic-end-eight-game-losing-streak/">http://global.nba.com/news/burke-leads-knicks-past-magic-end-eight-game-losing-streak/</a>	2
<a href="http://global.nba.com/news/brooks-scores-19-lead-grizzlies-lakers-123-114/">http://global.nba.com/news/brooks-scores-19-lead-grizzlies-lakers-123-114/</a>	2
<a href="http://global.nba.com/news/aldridges-top-15-rankings-houston-rockets-boston-celtics-soaring/">http://global.nba.com/news/aldridges-top-15-rankings-houston-rockets-boston-celtics-soaring/</a>	2
<a href="http://global.nba.com/news/thunder-host-blazers-in-key-west-showdown/">http://global.nba.com/news/thunder-host-blazers-in-key-west-showdown/</a>	2
<a href="http://global.nba.com/news/ingram-21-points-lakers-show-off-depth-beat-hawks/">http://global.nba.com/news/ingram-21-points-lakers-show-off-depth-beat-hawks/</a>	2
<a href="http://global.nba.com/news/jazz-ready-to-launch-season-at-kings/">http://global.nba.com/news/jazz-ready-to-launch-season-at-kings/</a>	2
<a href="http://global.nba.com/news/das-2018-nba-offseason-rankings-the-top-10/">http://global.nba.com/news/das-2018-nba-offseason-rankings-the-top-10/</a>	2
<a href="http://global.nba.com/news/report-memphis-grizzlies-interested-in-free-agent-avery-bradley/">http://global.nba.com/news/report-memphis-grizzlies-interested-in-free-agent-avery-bradley/</a>	2
<a href="http://global.nba.com/news/rockets-rally-spoil-defending-champion-warriors-return/">http://global.nba.com/news/rockets-rally-spoil-defending-champion-warriors-return/</a>	2
<a href="http://global.nba.com/news/antetokounmpo-newcomer-bledsoe-lead-bucks-dallas/">http://global.nba.com/news/antetokounmpo-newcomer-bledsoe-lead-bucks-dallas/</a>	2
<a href="http://global.nba.com/news/lebron-leads-cavaliers-past-celtics-season-opener-102-99/">http://global.nba.com/news/lebron-leads-cavaliers-past-celtics-season-opener-102-99/</a>	2
<a href="http://global.nba.com/news/gordon-hayward-opens-injury-rules-return-2017-18/">http://global.nba.com/news/gordon-hayward-opens-injury-rules-return-2017-18/</a>	2
<a href="http://global.nba.com/news/can-no-1-seeded-toronto-raptors-eastern-conferences-top-threat-2/">http://global.nba.com/news/can-no-1-seeded-toronto-raptors-eastern-conferences-top-threat-2/</a>	2
<a href="http://global.nba.com/news/nuggets-defeat-pelicans-consolation-matchup/">http://global.nba.com/news/nuggets-defeat-pelicans-consolation-matchup/</a>	2
<a href="http://global.nba.com/news/taurean-prince-replacing-malcolm-brogdon-mtn-dew-kickstart-rising-stars/">http://global.nba.com/news/taurean-prince-replacing-malcolm-brogdon-mtn-dew-kickstart-rising-stars/</a>	2
<a href="http://global.nba.com/news/watch-andre-de-grasse-micd-celebrity-game/">http://global.nba.com/news/watch-andre-de-grasse-micd-celebrity-game/</a>	2
<a href="http://global.nba.com/news/kris-dunn-suffers-dislocated-finger-2-4-weeks/">http://global.nba.com/news/kris-dunn-suffers-dislocated-finger-2-4-weeks/</a>	2
<a href="http://global.nba.com/news/pro%e2%80%a2line-watch-december-23/">http://global.nba.com/news/pro%e2%80%a2line-watch-december-23/</a>	2
<a href="http://global.nba.com/news/harden-scores-29-lead-rockets-past-grizzlies/">http://global.nba.com/news/harden-scores-29-lead-rockets-past-grizzlies/</a>	2
<a href="http://global.nba.com/news/rest-day-proves-pivotal-warriors-success-spurs/">http://global.nba.com/news/rest-day-proves-pivotal-warriors-success-spurs/</a>	2
<a href="http://global.nba.com/news/hood-scores-29-points-jazz-beat-spurs/">http://global.nba.com/news/hood-scores-29-points-jazz-beat-spurs/</a>	2
<a href="http://global.nba.com/news/shootaround-august-25/">http://global.nba.com/news/shootaround-august-25/</a>	2
<a href="http://global.nba.com/news/gordon-magic-stop-7-game-skid-beat-pistons-115-106-in-ot/">http://global.nba.com/news/gordon-magic-stop-7-game-skid-beat-pistons-115-106-in-ot/</a>	2
<a href="http://global.nba.com/news/former-boston-celtics-great-jo-jo-white-dead-71/">http://global.nba.com/news/former-boston-celtics-great-jo-jo-white-dead-71/</a>	2

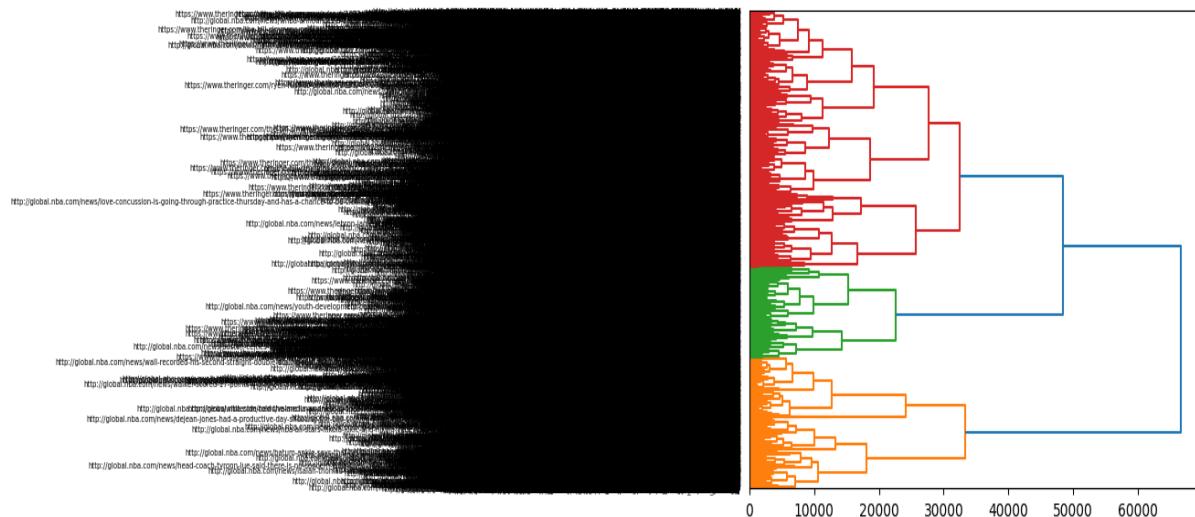
### Number of clusters obtained

The selection of the optimal number of clusters by fastcluster's agglomerative clustering depends on the input data type and the specific distance and linkage metrics employed.

For Single linkage algorithm , number of clusters are 2.Dendrogram is shown as below



For Complete linkage algorithm , number of clusters are 4 and Dendogram is shown below



### **Integration of clustering result with relevance model and displaying on UI**

- Both Single and Complete Linkage code i am able to execute only for 20,000 data due to the memory limitations.
- Pagerank relevance model is performed on 60,000 data by Allama so I worked with Allama Hossian to get the top 20,000 results from pagerank and worked with Vaishnavi Kunduru to pass the output of pagerank relevance model (i.e first 20k data) as argument while calling the single and complete clustering function in Clustering class.
- I integrated both clustering and top relevance model result in such a way that results are organized in a manner where the top-ranked result is part of a cluster , and all other results within the same cluster are subsequently ranked based on their page rank in decreasing order.
- I took the next top ranked result and their clusters are displayed in decreasing order .
- I stopped iterating after 10 top ranked results from 20k relevance output are present in the single and complete clustering 20,000 output files.
- I returned the output in the list of dictionaries where each dictionary has the title , description and link as keys.
- Vaishnavi Kunduru handled the output in UI such a way that whenever we search a query and select clustering from the dropdown , top 5 documents with link and title are displayed under Single and Complete linkage Clustering title.

### **How many queries did you experiment with – such that clustering could be used to improve the results of your search engine (3 points)?**

I experimented with the 50 queries for all the three clustering methods i.e flat clustering , single linkage and complete linkage . Results of search engines were improved as we performed clustering on relevance models.

### **State clearly how many queries you have used to test the impact of the results of each clustering method, how you have generated them and how you have judged the results of your relevance models (3 points).**

10 queries I used to test the impact of the results . These 10 queries which are not part of seed list directly crawled web pages . I tested the advanced queries which are not direct words from seeds and able to get the relevant results.

50 queries I tested earlier are from the seed database.

I observed that page rank results are much more relevant than HITS for most of the queries. The clustering outcomes appear to be meaningful, as they arrange similar outcomes together, prioritizing them over dissimilar outcomes.

**Discuss how you have decided to select the queries for the demonstration of your search engine (2 points)**

The criteria for selection of queries is based on the observation obtained from the type of clusters formed for both the clustering methods used. As shown in the sample subset of the results obtained above, one of the query selection criteria is entering the team name i.e bulls, as it was observed that belonging to the same team i.e bulls were grouped together. Another selection criteria is the player name i.e kobe. For example, relevance models are grouped together while results about the kobe are grouped together.

**Provide three examples of the queries and the results produced by your search engine and the clusters that you have created**

**Query1 : giannis**

<p><b>Results for: <i>giannis</i></b></p> <p><b>Our Search Engine Results</b></p> <p><b>Relevance model</b></p> <ul style="list-style-type: none"><li>• <a href="#">Giannis Antetokounmpo helps hot-shooting Bucks beat 76ers 117-104</a> <a href="https://www.basketballnews.com/stories/giannis-antetokounmpo-helps-hotshooting-bucks-beat-76ers-117104">https://www.basketballnews.com/stories/giannis-antetokounmpo-helps-hotshooting-bucks-beat-76ers-117104</a></li><li>• <a href="#">Bucks rip Heat 128-99 after Giannis exits early with knee issue</a> <a href="https://www.basketballnews.com/stories/bucks-rip-heat-12899-after-giannis-exits-early-with-knee-issue">https://www.basketballnews.com/stories/bucks-rip-heat-12899-after-giannis-exits-early-with-knee-issue</a></li><li>• <a href="#">Damian Lillard, Giannis Antetokounmpo named NBA's Players of the Week</a> <a href="https://www.basketballnews.com/stories/damian-lillard-giannis-antetokounmpo-named-nbas-players-of-the-week">https://www.basketballnews.com/stories/damian-lillard-giannis-antetokounmpo-named-nbas-players-of-the-week</a></li></ul> <p><b>Flat Clustering</b></p> <ul style="list-style-type: none"><li>• <a href="#">Giannis Antetokounmpo helps hot-shooting Bucks beat 76ers 117-104</a> <a href="https://www.basketballnews.com/stories/giannis-antetokounmpo-helps-hotshooting-bucks-beat-76ers-117104">https://www.basketballnews.com/stories/giannis-antetokounmpo-helps-hotshooting-bucks-beat-76ers-117104</a></li><li>• <a href="#">Bucks rip Heat 128-99 after Giannis exits early with knee issue</a> <a href="https://www.basketballnews.com/stories/bucks-rip-heat-12899-after-giannis-exits-early-with-knee-issue">https://www.basketballnews.com/stories/bucks-rip-heat-12899-after-giannis-exits-early-with-knee-issue</a></li><li>• <a href="#">Giannis Antetokounmpo open to playing for Bulls 'down the line'</a> <a href="https://www.basketballnews.com/stories/giannis-antetokounmpo-open-to-playing-for-bulls-down-the-line">https://www.basketballnews.com/stories/giannis-antetokounmpo-open-to-playing-for-bulls-down-the-line</a></li><li>• <a href="#">Bucks rally late in Giannis Antetokounmpo's return for 117-113 win</a> <a href="https://www.basketballnews.com/stories/bucks-rally-late-in-giannis-antetokounmpo-return-for-117113-win">https://www.basketballnews.com/stories/bucks-rally-late-in-giannis-antetokounmpo-return-for-117113-win</a></li><li>• <a href="#">Giannis Antetokounmpo (COVID) expected to play vs. Celtics</a> <a href="https://www.basketballnews.com/stories/giannis-antetokounmpo-covid-expected-to-play-vs-celestial">https://www.basketballnews.com/stories/giannis-antetokounmpo-covid-expected-to-play-vs-celestial</a></li></ul>
---

#### **Single Clustering**

- [Giannis Antetokounmpo helps hot-shooting Bucks beat 76ers 117-104](#)  
<https://www.basketballnews.com/stories/giannis-antetokounmpo-helps-hotshooting-bucks-beat-76ers-117104>
- [Tatum scores All-Star record 55, wins MVP as Team Giannis wins 184-175](#)  
<https://www.basketballnews.com/stories/tatum-scores-allstar-record-55-wins-mvp-as-team-giannis-wins-184175>
- [Giannis Antetokounmpo \(COVID\) expected to play vs. Celtics](#)  
<https://www.basketballnews.com/stories/giannis-antetokounmpo-covid-expected-to-play-vs-celecs>
- [Damian Lillard, Giannis Antetokounmpo named NBA Players of Week](#)  
<https://www.basketballnews.com/stories/damian-lillard-giannis-antetokounmpo-named-nba-players-of-week>
- [Bucks rally late in Giannis Antetokounmpo's return for 117-113 win](#)  
<https://www.basketballnews.com/stories/bucks-rally-late-in-giannis-antetokounmpos-return-for-117113-win>

#### **Complete Clustering**

- [Giannis Antetokounmpo helps hot-shooting Bucks beat 76ers 117-104](#)  
<https://www.basketballnews.com/stories/giannis-antetokounmpo-helps-hotshooting-bucks-beat-76ers-117104>
- [Giannis Antetokounmpo open to playing for Bulls 'down the line'](#)  
<https://www.basketballnews.com/stories/giannis-antetokounmpo-open-to-playing-for-bulls-down-the-line>
- [Bucks rally late in Giannis Antetokounmpo's return for 117-113 win](#)  
<https://www.basketballnews.com/stories/bucks-rally-late-in-giannis-antetokounmpos-return-for-117113-win>
- [Celtics' Joe Mazzulla to coach Team Giannis in 2023 NBA All-Star Game](#)  
<https://www.basketballnews.com/stories/celtics-joe-mazzulla-to-coach-team-giannis-in-2023-nba-allstar-game>
- [Giannis limps, Greece rolls into round of 16 at EuroBasket](#)  
<https://www.basketballnews.com/stories/giannis-limbs-greece-rolls-into-round-of-16-at-eurobasket>

#### **Summary :**

From this we can say that if we enter the player name i.e kobe. relevance models are grouped together for all kobe relevant results i.e Named players of the week, highest scores of kobe etc. are grouped.

## Query 2: lebron

Results for: **lebron**

**Our Search Engine Results**

**Relevance model**

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)  
<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>
- [LeBron James closes in on NBA record, Lakers falls to Pelicans](#)  
<https://www.basketballnews.com/stories/lebron-james-closes-in-on-nba-record-lakers-falls-to-pelicans>
- ['Im not going anywhere': LeBron James plans to play 'a few more years'](#)  
<https://www.basketballnews.com/stories/im-not-going-anywhere-lebron-james-plans-to-play-a-few-more-years>

**Flat Clustering**

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)  
<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>
- [LeBron, Anthony Davis get new-look Lakers past Pelicans](#)  
<https://www.basketballnews.com/stories/lebron-anthony-davis-get-newlook-lakers-past-pelicans>
- [How LeBron James, Anthony Davis recruited Russell Westbrook to Lakers](#)  
<https://www.basketballnews.com/stories/how-lebron-james-anthony-davis-recruited-russell-westbrook-to-lakers>

**Single Clustering**

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)  
<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>
- [LeBron James passes Kareem Abdul-Jabbar as NBA all-time scoring leader](#)  
<https://www.basketballnews.com/stories/nba-breaking-news-lebron-james-passes-kareem-abduljabbar-as-nba-all-time-scoring-leader-los-angeles-lakers-individual-record-akron-ohio-cleveland-cavaliers-miami-heat>
- [Referees' union says LeBron James was fouled on controversial play](#)  
<https://www.basketballnews.com/stories/referees-union-says-lebron-james-was-fouled-on-controversial-play>
- [Scan on LeBron James' sore left foot shows no serious damage](#)  
<https://www.basketballnews.com/stories/scan-on-lebron-james-sore-left-foot-shows-no-serious-damage>
- [LeBron James climbs to fourth place on NBA's career assists list](#)  
<https://www.basketballnews.com/stories/lebron-james-climbs-to-fourth-place-on-nbas-career-assists-list>

**Complete Clustering**

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)  
<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>
- [LeBron James climbs to fourth place on NBA's career assists list](#)  
<https://www.basketballnews.com/stories/lebron-james-climbs-to-fourth-place-on-nbas-career-assists-list>
- [LeBron James passes Kareem Abdul-Jabbar as NBA all-time scoring leader](#)  
<https://www.basketballnews.com/stories/nba-breaking-news-lebron-james-passes-kareem-abduljabbar-as-nba-all-time-scoring-leader-los-angeles-lakers-individual-record-akron-ohio-cleveland-cavaliers-miami-heat>
- [What LeBron passing Kareem means for NBA, GOAT debate](#)  
<https://www.basketballnews.com/stories/what-lebron-passing-kareem-means-for-nba-goat-debate>
- [Scan on LeBron James' sore left foot shows no serious damage](#)  
<https://www.basketballnews.com/stories/scan-on-lebron-james-sore-left-foot-shows-no-serious-damage>

## Summary :

From this we can say that if we enter the player name i.e lebron. relevance models are grouped together for all lebron relevant results i.e Named players of the week,Comments on this play, highest scores and titles etc. are grouped.

### Query 3 : mvp

Results for: *mvp*

#### Our Search Engine Results

##### Relevance model

- [Las Vegas Aces star A'ja Wilson adds second WNBA MVP to her collection](#)  
<https://www.basketballnews.com/stories/las-vegas-aces-star-aja-wilson-adds-second-wnba-mvp-to-her-collection>
- [Warriors star Stephen Curry unanimously named 2022 NBA Finals MVP](#)  
<https://www.basketballnews.com/stories/golden-state-warriors-star-stephen-curry-named-2022-nba-finals-mvp-steve-kerr>
- [Tatum scores All-Star record 55, wins MVP as Team Giannis wins 184-175](#)  
<https://www.basketballnews.com/stories/tatum-scores-allstar-record-55-wins-mvp-as-team-giannis-wins-184175>

##### Flat Clustering

- [Las Vegas Aces star A'ja Wilson adds second WNBA MVP to her collection](#)  
<https://www.basketballnews.com/stories/las-vegas-aces-star-aja-wilson-adds-second-wnba-mvp-to-her-collection>
- [Tatum scores All-Star record 55, wins MVP as Team Giannis wins 184-175](#)  
<https://www.basketballnews.com/stories/tatum-scores-allstar-record-55-wins-mvp-as-team-giannis-wins-184175>
- [Magic signing G League Finals MVP Devin Cannady to 10-day contract](#)  
<https://www.basketballnews.com/stories/magic-signing-g-league-finals-mvp-devin-cannady-to-10day-contract>
- [Dishes and Dimes: The real problem with MVP voting](#)  
<https://www.basketballnews.com/stories/nba-podcast-dishes-dimes-toronto-raptors-los-angeles-clippers-heat-celebs-lakers-kyle-lowry-pascal-siakam-nick-nurse>
- [Why Nikola Jokic deserves back-to-back MVPs](#)  
<https://www.basketballnews.com/stories/why-nikola-jokic-deserves-backtoback-mvps>

#### **Single Clustering**

- [Las Vegas Aces star A'ja Wilson adds second WNBA MVP to her collection](#)  
<https://www.basketballnews.com/stories/las-vegas-aces-star-aja-wilson-adds-second-wnba-mvp-to-her-collection>
- [Dishes and Dimes: The real problem with MVP voting](#)  
<https://www.basketballnews.com/stories/nba-podcast-dishes-dimes-toronto-raptors-los-angeles-clippers-heat-celebs-lakers-kyle-lowry-pascal-siakam-nick-nurse>
- [LeBron on MVP voting: 'It pissed me off'](#)  
<https://www.basketballnews.com/stories/nba-rumors-los-angeles-lakers-lebron-james-mvp-voting-pissed-me-off-giannis-antetokounmpo>
- [Why Nikola Jokic deserves back-to-back MVPs](#)  
<https://www.basketballnews.com/stories/why-nikola-jokic-deserves-backtoback-mvps>
- [Russell Westbrook's MVP odds range from 25-1 to 85-1](#)  
<https://www.basketballnews.com/stories/russell-westbrooks-mvp-odds-range-from-251-to-851>

#### **Complete Clustering**

- [Las Vegas Aces star A'ja Wilson adds second WNBA MVP to her collection](#)  
<https://www.basketballnews.com/stories/las-vegas-aces-star-aja-wilson-adds-second-wnba-mvp-to-her-collection>
- [Jayson Tatum weighs in on the NBA's wide-open MVP race](#)  
<https://www.basketballnews.com/stories/jayson-tatum-weighs-in-on-the-nbas-wideopen-mvp-race>
- [LeBron on MVP voting: 'It pissed me off'](#)  
<https://www.basketballnews.com/stories/nba-rumors-los-angeles-lakers-lebron-james-mvp-voting-pissed-me-off-giannis-antetokounmpo>
- [Russell Westbrook's MVP odds range from 25-1 to 85-1](#)  
<https://www.basketballnews.com/stories/russell-westbrooks-mvp-odds-range-from-251-to-851>
- [Magic signing G League Finals MVP Devin Cannady to 10-day contract](#)  
<https://www.basketballnews.com/stories/magic-signing-g-league-finals-mvp-devin-cannady-to-10day-contract>

#### **Summary :**

In this instead of querying by player or team name , queried by mvp such that all the information about the voting of mvp , results of mvp , players won the mvp are grouped

#### **Summary of clustering**

Kmeans provides faster results of clustering for large data sets compared to single and complete linkage .

Single linkage results are faster but it cannot perform accurately when there is a lot of noise in the data .

That is the reason we got 2 clusters for single clustering.

Complete linkage results perform more accurately when compared with single linkage for noisy data , That's the reason we got 4 clusters for complete clustering.

**6 Query Expansion:****Rocchio**

To test our algorithm, we tried out various combinations of queries from the assumption of user:

- Sufficient knowledge of domain (Clippers)
- Not enough knowledge of domain (Nbaa)
- Spelling errors (Bston)
- Short forms for names (steph instead of stephen)

**Player names:**

- 1) Davis
- 2) Durant
- 3) Jimmy
- 4) Sam Jones

**Team names:**

- 5) Boston Celtics
- 6) Clippers
- 7) Dallas Mavericks
- 8) Pacers

**Basketball related words:**

- 9) FIBA
- 10) college
- 11) 2016 champions
- 12) injury

**Queries with mistakes:**

- 13) jams
- 14) memphis adam
- 15) Bston
- 16) nbaa

**Ambiguous queries:**

- 17) Steph Curry
- 18) Stephen Curry
- 19) Kobe Bryant
- 20) Bryant

Reiterating, this provides a good understanding of how adaptable our algorithm is to:

- Handling different search terms and multiple words
- Mistakes
- Nicknames
- Two players having same last name

The different queries for testing the Rocchio algorithm is as follows:

Kobe Bryant:

Relevant:

[https://www.espn.com/nba/story/\\_/id/13815620/losangeles-lakers-the-kobe-bryant-issue](https://www.espn.com/nba/story/_/id/13815620/losangeles-lakers-the-kobe-bryant-issue)

[https://www.espn.com/nba/story/\\_/id/20149584/does-kyrie-irving-kobe-bryant-lebron-james-problem-nba](https://www.espn.com/nba/story/_/id/20149584/does-kyrie-irving-kobe-bryant-lebron-james-problem-nba)

Irrelevant: (Includes results about thomas bryant)

[https://www.espn.com/nba/story/\\_/id/35623232/sources-lakers-trading-center-thomas-bryant-nuggets](https://www.espn.com/nba/story/_/id/35623232/sources-lakers-trading-center-thomas-bryant-nuggets)

<https://www.basketballnews.com/stories/nba-news-thomas-bryant-fined-45000-dollars-repeated-inappropriate-contact-game-official-washington-wizards-detroit-pistons>

2016 champions:

Relevant:

[https://www.espn.com/nba/story/\\_/page/playoffindex16/nba-playoffs-coverage](https://www.espn.com/nba/story/_/page/playoffindex16/nba-playoffs-coverage)

[https://www.espn.com/nba/story/\\_/page/GSWCLE-15859161/nba-playoffs-2016-experts-predictions-golden-state-warriors-cleveland-cavaliers-nba-finals](https://www.espn.com/nba/story/_/page/GSWCLE-15859161/nba-playoffs-2016-experts-predictions-golden-state-warriors-cleveland-cavaliers-nba-finals)

Irrelevant: (Includes results focused on last word ignoring 2016)

<https://www.basketballnews.com/stories/jill-biden-wants-champions-lsu-and-iowa-at-white-house>

<https://www.basketballnews.com/stories/new-champion-uconn-to-headline-2024-maui-invitational-field>

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

- The outputs were not correct at times when results included two or more words since the complexity increases.
- For queries with spelling mistakes no output was returned.
- It is important to consider the trade off between alpha, beta and gamma values. The gamma value should be less than beta to provide more importance to positive feedback

We do not consider the Rocchio algorithm in general as there are a lot of assumptions we make about users that need to be satisfied to get good results. The user should not have any spelling mistakes, word mismatch. The relevance documents should also have even term distribution.

Query	Expanded query
Larry	Larry Bird
Durant	Durant stats
Boston Celtics	Boston Celtics team
Clippers	Clippers los
Pacers	Pacers vs
FIBA	FIBA rank
college	College game
2016 champions	2016 champions 2016
jams	Jams lebron
Memphis adam	Memphis adam silver
Bston	No result
Stephen Curry	Stephen Curry game
Kobe Bryant	Kobe Bryant stats
Kobe	Kobe bryant

### **Pseudo relevance feedback:**

50 queries chosen for pseudo-relevance feedback

Similar to Rocchio, To test our algorithm, we tried out various combinations of queries from the assumption of user:

- Sufficient knowledge of domain
- Not enough knowledge of domain
- Spelling errors
- Short forms for names

<b>Query</b>	<b>Expanded query</b>
NBA	NBA players
Lakers	Lakers Los
Russell	Russell Westbrook
Lebron	Lebron Jam
MVP	MVP win kia
Chicago	Chicago guard
College	College Season
Giannis	Giannis bucks
Kobe	Kobe Bryant
Abdul Kareem	Abdul Kareem Jabbar
NCAA	NCAA team
Stars	Stars announced
Stephen	Stephen steph
Curry	Curry steph
Women	Women game
Coach	Coach Basketball
Kyrie	Kyrie Irving
Gasol Spain	Gasol Pain Pau
Phoenix	Phoenix Suns
WNBA	WNBA games
Trae	Trae Hawk
Nikola Denver	Nikola Denver Jovic
Poole	Poole Green
Rankings	Rankings Net
Victor	Victor Oladipo

Antetokounmpo	Antetokounmpo bucks
Butler heat	Butler heat jimmy
Clippers los	Clippers los angeles
Nets kevin	Nets kevin Brooklyn
Mens	Mens Basketball
Luka Dallas	Luka Dallas Mavericks
Injury	Injury Source
Hornets raptors	Hornets raptors Toronto
Grizzlies morant	Grizzlies morant ja
Dwight magic	Dwight magic howard
Durant	Durant nets
Detroit	Detroit Pistons
Curry	Curry stephen
Warriors	Warriors curry steph vs
Playoffs	Playoffs 2022 games
Shaq	Shaq Lakers Kobe
MVP	MVP win kia
Miami Heat	Miami Heat vs game
baktbll	No result
westbrook	Westbrook lakers
lebrun	No result
Magic johnson	Magic johnson stanley orlando
Magic jonson	No result
Johnson	Johnson Magic basketball
2016 nba	2016 nba nba relive kris

- For all three cluster methods, 3 separate queries are analyzed.

- The local document set in my case is just the ID of Data CSV dataframe. The top 500 relevant documents from the relevance model are used for each query.
- The local vocabulary, stem and vocab of stems are calculated for 3 clusters. A combination of Lancaster Stemmer and Porter Stemmer were used. The top few results are given for each query as it is impossible to print all values.
- The normalized correlation scores are calculated, sorted to get the expanded query terms
- For most of the observations, the Associative cluster gave good results, followed by metric and then scalar gave poor results. We are only showcasing Associative cluster results in the UI.

## **ASSOCIATIVE CLUSTERING:**

### **Query 1:**

NBA

#### **Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

#### **Local Vocabulary:**

```
[lamarus', 'aldrige', '7', 'time', 'nba', 'star', 'announces', 'retirement']
[gradey', 'dick', 'leave', 'kansas', '2023', 'nba', 'draft', 'one', 'season']
[lebron', 'james', 'closes', 'nba', 'record', 'lakers', 'falls', 'pelicans']
[ucla', 'guard', 'jaylen', 'clark', 'declares', '2023', 'nba', 'draft']
[tracking', 'data', 'shows', 'nba', 'players', 'run', 'miles', 'per', 'game']
[nba', 'roundtable', 'team', 'desperately', 'needs', 'make', 'trade']
[uconn', 'guard', 'jordan', 'hawkins', 'declares', '2023', 'nba', 'draft']
[lebron', 'james', 'climbs', 'fourth', 'place', 'nbas', 'career', 'assists', 'list']
[embiid', 'vs', 'doncic', 'scoring', 'title', 'nbas', 'final', 'week']
[nba', 'announces', '28', 'players', 'selected', '2023', 'jordan', 'rising', 'stars', 'games']
```

#### **Local stems:**

```
'lamarus', 'aldrige', '7', 'time', 'nba', 'star', 'announces', 'retirement', 'gradey', 'dick', 'leave', 'kansas', '2023',
'draft', 'one', 'season', 'lebron', 'james', 'closes', 'record', 'lakers', 'falls', 'pelicans', 'ucla', 'guard', 'jaylen',
'clark', 'declares', 'tracking', 'data', 'shows', 'players', 'run', 'miles', 'per', 'game', 'roundtable', 'team',
'desperately', 'needs', 'make', 'trade', 'uconn', 'jordan', 'hawkins', 'climbs', 'fourth', 'place', 'career', 'assists',
'list', 'embiid', 'vs', 'doncic', 'scoring', 'title', 'final', 'week', '28', 'selected', 'rising', 'duke', 'freshman',
'dereck', 'lively', 'ii', 'says', 'hell', 'enter', 'passing', 'kareem', 'means', 'goat', 'debate', '2022', 'free', 'agency',
'notable', 'still', 'available', 'remaining', 'agents', 'unsigned', 'sets', 'covid', '19', 'protocols', '23', 'stats',
'notebook'
```

**Vocabulary of stems:**

```
{'nba': {'nba', 'nbas'}, 'star': {'star', 'stars'}, 'annount': {'announces', 'announced'}, 'retir': {'retirement', 'retired', 'retires'}, 'draft': {'draft', 'drafts'}, 'season': {'seasons', 'season'}, 'jam': {'james', 'jamal'}, 'clos': {'close', 'closes'}, 'lak': {'lakers', 'laker', 'lake'}, 'fal': {'fall', 'falls'}, 'guard': {'guard', 'guards', 'guarded'}, 'track': {'tracker', 'tracking'}, 'dat': {'data', 'dates'}, 'show': {'showed', 'shows', 'show', 'showing'}, 'play': {'play', 'players', 'player'}, 'mil': {'miles', 'million', 'miller', 'mills'}, 'gam': {'games', 'game'}, 'team': {'teams', 'team'}, 'desp': {'desperately', 'desperate'}, 'nee': {'needs', 'need'}, 'mak': {'making', 'make'}, 'makers': {'makers'}, 'trad': {'trade', 'trading', 'trades', 'traded'}, 'jord': {'jordan', 'jordans'}, 'four': {'fourth', 'four'}, 'car': {'cares', 'career'}, 'assist': {'assist', 'assists'}, 'list': {'list', 'listed'}, 'dont': {'doncie', 'dont'}, 'scor': {'scorer', 'scores', 'scorers', 'scoring'}, 'titl': {'titles', 'title'}, 'fin': {'fines', 'final', 'finalists', 'finals'}, 'select': {'selected', 'select'}, 'ris': {'rise', 'risers', 'rising'}, 'duk': {'duke', 'dukes'}, 'ent': {'enter', 'entering', 'entered', 'enters'}, 'pass': {'passing', 'passes'}}
```

**Correlations:**

```
[('nba', 'nba', 0.3333333333333333),  
 ('draft', 'nba', 0.1163141993957704),  
 ('2022', 'nba', 0.10664605873261206),  
 ('play', 'nba', 0.10124610591900311),  
 ('team', 'nba', 0.07345575959933222),  
 ('players', 'nba', 0.06335616438356165),  
 ('league', 'nba', 0.057291666666666664),  
 ('bet', 'nba', 0.05574912891986063),  
 ('2021', 'nba', 0.054385964912280704),  
 ('game', 'nba', 0.04946996466431095),  
 ('best', 'nba', 0.04946996466431095),  
 ('bets', 'nba', 0.04946996466431095),  
 ('star', 'nba', 0.047872340425531915),
```

**Cluster Selection:**

Associative cluster is used.

**Expanded Query Result:**

NBA 2022 play team

**Query 2:**

Giannis

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

### **Local Vocabulary:**

- ['giannis', 'antetokounmpo', 'helps', 'hot', 'shooting', 'bucks', 'beat', '76ers', '117', '104']
- ['bucks', 'rip', 'heat', '128', '99', 'giannis', 'exits', 'early', 'knee', 'issue']
- ['damian', 'lillard', 'giannis', 'antetokounmpo', 'named', 'nbas', 'players', 'week']
- ['giannis', 'limps', 'greece', 'rolls', 'round', '16', 'eurobasket']
- ['celtics', 'joe', 'mazzulla', 'coach', 'team', 'giannis', '2023', 'nba', 'star', 'game']
- ['damian', 'lillard', 'giannis', 'antetokounmpo', 'named', 'nba', 'players', 'week']
- ['giannis', 'antetokounmpo', 'open', 'playing', 'bulls', 'line']
- ['atum', 'scores', 'star', 'record', '55', 'wins', 'mvp', 'team', 'giannis', 'wins', '184', '175']
- ['bucks', 'rally', 'late', 'giannis', 'antetokounmpo', 'return', '117', '113', 'win']
- ['giannis', 'antetokounmpo', 'covid', 'expected', 'play', 'vs', 'celtics']

### **Local stems:**

- ['giannis', 'antetokounmpo', 'helps', 'hot', 'shooting', 'bucks', 'beat', '76ers', '117', '104', 'rip', 'heat', '128', '99', 'exits', 'early', 'knee', 'issue', 'damian', 'lillard', 'named', 'nbas', 'players', 'week', 'limps', 'greece', 'rolls', 'round', '16', 'eurobasket', 'celtics', 'joe', 'mazzulla', 'coach', 'team', '2023', 'nba', 'star', 'game', 'open', 'playing', 'bulls', 'line', 'atum', 'scores', 'record', '55', 'wins', 'mvp', '184', '175', 'rally', 'late', 'return', '113', 'win', 'covid', 'expected', 'play', 'vs', 'denies', 'fourth', 'triple', 'double', 'stirs', 'talks', 'stat', 'padding', 'richard', 'jefferson', 'unfair', 'compare', 'anthony', 'davis', 'prepared', 'continue', 'reign', 'terror', 'suns', 'hack', 'whack', 'know', 'hell', 'keep', 'coming', 'buck', 'bud', 'verge', 'another', 'self', 'inflicted', 'collapse', 'upgraded']

### **Vocabulary of stems:**

```
{'antetokounmpo': {'antetokounmpo': 'antetokounmpo'}, 'help': {'helps': 'help'}, 'shoot': {'shooter', 'shooting'}, 'buck': {'bucks': 'buck'}, 'beat': {'beater': 'beat'}, 'dam': {'damage': 'damian'}, 'nam': {'named', 'name'}, 'nba': {'nba': 'nbas'}, 'play': {'play': 'players', 'player', 'playing'}, 'coach': {'coach': 'coaching'}, 'team': {'teams': 'team', 'teammates'}, 'star': {'star': 'stars'}, 'gam': {'games': 'game'}, 'op': {'opener': 'open'}, 'scor': {'scored': 'scores', 'scoring'}, 'record': {'records': 'record'}, 'win': {'win': 'wins', 'winning', 'winner'}, 'ral': {'rally': 'rallies'}, 'return': {'returns': 'return'}, 'four': {'fourth': 'four'}, 'talk': {'talk': 'talks'}, 'stat': {'status': 'stat', 'stats', 'state'}, 'continu': {'continues': 'continue'}, 'hack': {'hack': 'hacked'}, 'know': {'know', 'knows'}, 'keep': {'keeps': 'keep'}, 'com': {'coming': 'comment', 'comments'}, 'playoff': {'playoffs', 'playoff'}, 'quest': {'questionable', 'question', 'questions'}, 'fin': {'fine': 'finals', 'fined', 'finalists', 'fines'}, 'lebron': {'lebron': 'lebrons'}, 'lead': {'lead': 'leading', 'leads'}, 'get': {'getting': 'gets', 'get'}}
```

### **Correlations:**

- [('giannis', 'giannis', 0.3333333333333333),
- ('antetokounmpo', 'giannis', 0.2698268003646308),
- ('buck', 'giannis', 0.21548821548821548),
- ('bucks', 'giannis', 0.21420518602029312),

('nba', 'giannis', 0.1602136181575434),  
('game', 'giannis', 0.11297709923664122),  
('vs', 'giannis', 0.10581222056631892),  
('star', 'giannis', 0.0928),  
('play', 'giannis', 0.08723747980613894),  
('win', 'giannis', 0.07653910149750416),  
('team', 'giannis', 0.07455429497568881),  
('lebron', 'giannis', 0.06814310051107325),  
('milwaukee', 'giannis', 0.06666666666666667),  
('mvp', 'giannis', 0.06643952299829642),

**Cluster Selection:**

Associative cluster is considered.

**Expanded Query Result:**

Giannis antetokounmpo buck bucks nba

**Query 3:**

Lakers

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

**Local Vocabulary:**

['davis', '40', 'lebron', 'triple', 'double', 'lakers', 'beat', 'rockets']  
['james', 'triple', 'double', 'leads', 'lakers', 'past', 'knicks', 'ot']  
['lebron', 'james', 'closes', 'nba', 'record', 'lakers', 'falls', 'pelicans']  
['anthony', 'davis', 'leads', 'rally', '27', 'lakers', 'beat', 'mavs', '111', '108']  
['irving', '26', 'nets', 'beat', 'lakers', '121', '104', 'james', 'davis', 'sit']  
['russell', 'westbrook', 'joining', 'clippers', 'leaving', 'lakers']  
['juan', 'toscano', 'anderson', 'agrees', 'deal', 'lakers']  
['troy', 'brown', 'jr', 'signs', 'one', 'year', 'deal', 'lakers']  
['lakers', 'lonnie', 'walker', 'agree', 'one', 'year', 'deal']  
['lakers', 'sign', 'center', 'damian', 'jones', 'two', 'year', 'deal']

**Local stems:**

['davis', '40', 'lebron', 'triple', 'double', 'lakers', 'beat', 'rockets', 'james', 'leads', 'past', 'knicks', 'ot', 'closes', 'nba', 'record', 'falls', 'pelicans', 'anthony', 'rally', '27', 'mavs', '111', '108', 'irving', '26', 'nets', '121', '104', 'sit', 'russell', 'westbrook', 'joining', 'clippers', 'leaving', 'juan', 'toscano', 'anderson', 'agrees', 'deal', 'troy', 'brown', 'jr', 'signs', 'one', 'year', 'lonnie', 'walker', 'agree', 'sign', 'center', 'damian', 'jones', 'two', 'stein', 'heat', 'interested', 'kyrie', 'explore', 'trading', 'trade', 'likely', 'malik', 'monk', 'open', 'taking', 'less', 'money', 'return', 'wednesday', 'night', 'thomas', 'bryant', 'discussing', 'jazz', 'top', 'choice', 'free', 'agency', 'scores', '21', 'points', 'defeat', 'spurs', 'shines', 'wolves', 'may', 'found', 'right', 'formula', 'following', 'deadline', 'moves', 'begins', 'ramp', 'process', 'could', 'play', 'weeks', 'bucks', 'win', '9th', 'straight', 'hold', '115', '106', 'frank', 'vogel', 'contract', 'extension', 'austin', 'reaves', 'rising', 'challenges', 'push', 'luck', 'lack', 'exploring', 'blazers', 'three', 'point', 'offense', 'vs', 'get', 'new', 'look', 'pau', 'gasol', 'emotional', 'retire', '16', 'jersey', 'announcer', 'asked', 'suns', 'put', 'us', 'misery', 'take', '1', '0', 'lead', 'nuggets', 'wcf', 'shams', 'interviewed', 'mark', 'jackson']

### Vocabulary of stems:

beat': {'beater', 'beat', 'beating'}, 'lead': {'lead', 'leading', 'leads', 'leadership'}, 'clos': {'closing', 'closes'}, 'nba': {'nba', 'nbas'}, 'record': {'records', 'record'}, 'irv': {'irving', 'irvings'}, 'sit': {'sit', 'sits'}, 'russel': {'russell', 'russells'}, 'westbrook': {'westbrook', 'westbrooks'}, 'join': {'joins', 'joining', 'join', 'joined'}, 'leav': {'leaves', 'leaving', 'leave'}, 'agr': {'agrees', 'agree'}, 'deal': {'deal', 'deals'}, 'sign': {'sign', 'signing'}, 'signs', 'significant'}, 'on': {'one', 'oneal'}, 'year': {'years', 'year'}, 'cent': {'centers', 'center'}, 'interest': {'interest', 'interested'}, 'expl': {'explore', 'exploring'}, 'trad': {'trading', 'trades', 'trade', 'traded'}, 'lik': {'like', 'likely'}, 'monk': {'monks', 'monk'}, 'op': {'opens', 'open', 'opening', 'opener', 'operate'}, 'tak': {'taking', 'take', 'takes'}, 'return': {'returns', 'return', 'returning'}, 'night': {'night', 'nights'}, 'discuss': {'discussions', 'discusses', 'discussing'}, 'ag': {'agency', 'agent', 'agents'}, 'scor': {'scorer', 'scores', 'score'}, 'scoring': 'point': {'pointer', 'point', 'points'}, 'right': {'rights', 'right'}, 'mov': {'move', 'moves'}, 'play': {'play', 'plays', 'player', 'players', 'played', 'playing'}, 'week': {'week', 'weeks'}, 'win': {'winning', 'win', 'wins', 'winner', 'winners'}, 'vogel': {'vogels', 'vogel'}, 'contract': {'contracts', 'contract'}, 'extend': {'extend', 'extension'}

### Correlations:

[('lakers', 'lakers', 0.3333333333333333),  
 ('lebron', 'lakers', 0.168848167539267),  
 ('james', 'lakers', 0.16312997347480107),  
 ('nba', 'lakers', 0.13381294964028778),  
 ('vs', 'lakers', 0.12),  
 ('game', 'lakers', 0.09921259842519685),  
 ('westbrook', 'lakers', 0.09455128205128205),  
 ('russell', 'lakers', 0.09294871794871795),  
 ('trade', 'lakers', 0.0928),  
 ('time', 'lakers', 0.0864600326264274),  
 ('play', 'lakers', 0.08482871125611746),  
 ('davis', 'lakers', 0.07960199004975124),  
 ('anthony', 'lakers', 0.07394957983193277),

('live', 'lakers', 0.07251264755480608),

**Cluster Selection:**

Associative cluster

**Expanded Query Result:**

Lakers lebron james nba vs

**METRIC CLUSTERING:**

**Query 1:**

Lebron

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

**Local Vocabulary:**

['davis', 'lebron', 'triple', 'double', 'lakers', 'beat', 'rockets'], ['lebron', 'james', 'closes', 'nba', 'record', 'lakers', 'falls', 'pelicans'], ['im', 'going', 'anywhere', 'lebron', 'james', 'plans', 'play', 'years'], ['lebron', 'james', 'climbs', 'fourth', 'place', 'nbas', 'career', 'assists', 'list'], ['lebron', 'passing', 'kareem', 'means', 'nba', 'goat', 'debate'], ['samaki', 'walker', 'missed', 'foul', 'call', 'lebron', 'affects', 'nbas', 'integrity'], ['referees', 'union', 'says', 'lebron', 'james', 'fouled', 'controversial', 'play'], ['scan', 'lebron', 'james', 'sore', 'left', 'foot', 'shows', 'serious', 'damage'], ['lebron', 'james', 'passes', 'kareem', 'abdul', 'jabbar', 'nba', 'time', 'scoring', 'leader'], ['lebron', 'passing', 'kareem', 'means', 'nba', 'greatest', 'debate'], ['battier', 'love', 'ham', 'highlight', 'lebrons', 'miami', 'cleveland', 'la', 'stints'], ['days', 'later', 'lebron', 'still', 'feels', 'scoring', 'record', 'surreal']

**Local stems:**

['davis', '40', 'lebron', 'triple', 'double', 'lakers', 'beat', 'rockets', 'james', 'leads', 'past', 'knicks', 'ot', 'closes', 'nba', 'record', 'falls', 'pelicans', 'anthony', 'rally', '27', 'mavs', '111', '108', 'irving', '26', 'nets', '121', '104', 'sit', 'russell', 'westbrook', 'joining', 'clippers', 'leaving', 'juan', 'toscano', 'anderson', 'agrees', 'deal', 'troy', 'brown', 'jr', 'signs', 'one', 'year', 'lonnie', 'walker', 'agree', 'sign', 'center', 'damian', 'jones', 'two', 'stein', 'heat', 'interested', 'kyrie', 'explore', 'trading', 'trade', 'likely', 'malik', 'monk', 'open', 'taking', 'less', 'money',

'return', 'wednesday', 'night', 'thomas', 'bryant', 'discussing', 'jazz', 'top', 'choice', 'free', 'agency', 'scores', '21', 'points', 'defeat', 'spurs', 'shines', 'wolves', 'may', 'found', 'right', 'formula', 'following', 'deadline', 'moves', 'begins', 'ramp', 'process', 'could', 'play', 'weeks', 'bucks', 'win', '9th', 'straight', 'hold', '115', '106', 'frank', 'vogel', 'contract', 'extension', 'austin', 'reaves', 'rising', 'challenges', 'push', 'luck', 'lack', 'exploring', 'blazers', 'three', 'point', 'offense', 'vs', 'get', 'new', 'look', 'pau', 'gasol', 'emotional', 'retire', '16', 'jersey', 'announcer', 'asked', 'suns', 'put', 'us', 'misery', 'take', '1', '0', 'lead', 'nuggets', 'wcf', 'shams', 'interviewed', 'mark', 'jackson'

### Vocabulary of stems:

'lebron': {'lebron', 'lebrons'}, 'tripl': {'triple'}, 'doubl': {'double'}, 'laker': {'lakers'}, 'beat': {'beat', 'beats'}, 'rocket': {'rockets'}, 'jame': {'james'}, 'close': {'close', 'closing', 'closes', 'closed'}, 'nba': {'nba', 'nbas'}, 'record': {'records', 'record'}, 'fall': {'falls'}, 'pelican': {'pelicans'}, 'im': {'im'}, 'go': {'go', 'going'}, 'anywher': {'anywhere'}, 'plan': {'plan', 'plans'}, 'play': {'play', 'played', 'plays', 'playing'}, 'year': {'years', 'year'}, 'climb': {'climbs'}, 'fourth': {'fourth'}, 'place': {'place'}, 'career': {'career'}, 'assist': {'assist', 'assists'}, 'list': {'list'}, 'pass': {'passing', 'pass', 'passes'}, 'kareem': {'kareem'}, 'mean': {'meaning', 'mean', 'means'}, 'goat': {'goats', 'goat'}, 'debat': {'debate'}, 'samaki': {'samaki'}, 'walker': {'walker'}, 'miss': {'missed', 'miss', 'misses', 'missing'}, 'foul': {'foul', 'fouled'}, 'call': {'calls', 'call'}, 'affect': {'affects'}, 'integr': {'integrity'}, 'refere': {'referees'}, 'union': {'union'}, 'say': {'says', 'saying', 'say'}, 'controversi': {'controversy', 'controversial'}, 'scan': {'scan'}, 'sore': {'sore', 'sorenness'}, 'left': {'left'}, 'foot': {'foot'}, 'show': {'shows', 'show', 'showing'}, 'seriou': {'serious'}, 'damag': {'damage'}, 'abdul': {'abdul'}, 'jabbar': {'jabbar', 'jabbars'}, 'time': {'time', 'times'}, 'score': {'scores', 'score', 'scoring'}, 'leader': {'leader', 'leaders'}

### Correlations:

Lebron vs 2.499999999999999  
 Lebron james 1.1666666666666665  
 Lebron team 0.722222222222222  
 Lebron star 0.4166666666666663  
 Lebron head 0.333333333333333  
 Lebron st 0.333333333333333

### Cluster Selection:

Metric cluster provides good results, but still associative gave better results so using that.

### Expanded Query Result:

'Lebron vs james team'

### Query 2:

## WNBA

### **Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

### **Local Vocabulary:**

```
[['wnba', 'free', 'agency', 'versatile', 'center', 'azura', 'stevens', 'join', 'la', 'sparks']
 ['wnba', 'announces', 'expansion', 'charter', 'flights', 'program']
 ['wnba', 'draft', 'full', 'list', 'prospects', 'entered']
 ['sun', 'rally', 'th', 'beat', 'sky', 'advance', 'wnba', 'finals']
 ['wnba', 'first', 'round', 'drew', 'viewers', 'years']
 ['dewanna', 'bonner', 'scores', 'sun', 'beat', 'sky', 'game', 'wnba', 'semis']
 ['jewell', 'loyd', 'lifts', 'storm', 'aces', 'wnba', 'semis', 'opener']
 ['las', 'vegas', 'aces', 'star', 'aja', 'wilson', 'adds', 'second', 'wnba', 'mvp', 'collection']
 ['wnba', 'playoffs', 'previewing', 'second', 'round', 'matchups']
 ['griner', 'jail', 'wnba', 'players', 'skip', 'russia', 'offseason']
 ['wnba', 'legend', 'sue', 'bird', 'hopes', 'go', 'top', 'th', 'title']
 ['wnba', 'awards', 'roty', 'rookie', 'team', 'coty', 'improved']
 ['aces', 'becky', 'hammon', 'wins', 'wnba', 'coach', 'year', 'award']
 ['wnba', 'playoffs', 'one', 'key', 'stat', 'know', 'team']
 ['wnbas', 'brittney', 'griner', 'appeals', 'russian', 'prison', 'sentence']]
```

### **Local stems:**

```
{'defensive', 'heads', 'choosing', 'gigi', 'offseason', 'sentences', 'wrapping', 'betting', 'detainment',
 'executive', 'wilson', 'notebook', 'freed', 'gets', 'hamby', 'opt', 'jasmine', 'viewed', 'tragic', 'kings', 'fears',
 'help', 'scouting', 'brown', 'ap', 'raise', 'extends', 'champ', 'improving', 'enter', 'black', 'ot', 'entering', 'bria',
 'collection', 'awards', 'sucker', 'expansion', 'earns', 'release', 'championship', 'situation', 'new', 'ok', 'edition',
 'highest', 'playoffs', 'surprising', 'solution', 'dedicated', 'advance', 'clubs', 'lottery', 'arenas', 'drops', 'mark',
 'times', 'nets', 'contracts', 'z', 'capital', 'brady', 'shouldnt', 'format', 'double', 'griner', 'showed', 'ever', 'office',
 'ownership', 'sharpshooters', 'constantly'}
```

### **Vocabulary of stems:**

```
{'wnba': {'wnba', 'wnbas'}, 'free': {'free'}, 'agenc': {'agency'}, 'versatil': {'versatile'}, 'center': {'center'},
 'azura': {'azura'}, 'steven': {'stevens'}, 'join': {'joining', 'join'}, 'la': {'la', 'las'}, 'spark': {'sparks'}, 'announc':
 {'announces', 'announce', 'announced'}, 'expans': {'expansion'}, 'charter': {'charter'}, 'flight': {'flights'},
 'program': {'program'}, 'draft': {'draft', 'drafting', 'drafted'}, 'full': {'full'}, 'list': {'list', 'listing'}, 'prospect':
 {'prospects'}, 'enter': {'enter', 'entering', 'entered'}, 'sun': {'suns', 'sun'}, 'ralli': {'rally'}}
```

**Correlations:**

WNBA history 0.333333333333333  
WNBA brittney 0.333333333333333  
WNBA lottery 0.16666666666666666  
WNBA viewed 0.16666666666666666  
WNBA star 0.16666666666666666  
WNBA new 0.16666666666666666

**Cluster Selection:**

Metric cluster itself can be used. Although the results after the first few were not good.

**Expanded Query Result:**

WNBA history brittney

**Query 3:**

Jordan

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

**Local Vocabulary:**

['uconn', 'guard', 'jordan', 'hawkins', 'declares', 'nba', 'draft']  
['nba', 'announces', 'players', 'selected', 'jordan', 'rising', 'stars', 'games']  
['michael', 'jordan', 'considering', 'sale', 'hornets', 'deal', 'imminent']  
['jordan', 'poole', 'talks', 'role', 'warriors', 'coming', 'bench']  
['jordan', 'poole', 'turn', 'around', 'celtics', 'nba', 'finals']  
['wizards', 'waive', 'barton', 'convert', 'jordan', 'goodwin', 'standard', 'contract']  
['michael', 'jordan', 'start', 'nascar', 'team']  
['michael', 'jordan', 'launching', 'fan', 'engagement', 'app', 'called', 'heir']  
['kevin', 'garnett', 'different', 'level', 'respect', 'lebron', 'jordan']  
['bulls', 'sign', 'jordan', 'bell', 'hardship', 'contract']  
['jordan', 'clarkson', 'fined', 'k', 'making', 'contact', 'official']  
['michael', 'jordan', 'stars', 'teaming', 'hurts', 'nba']  
['mj', 'jordan', 'brand', 'important']

[**'jordan', 'small', 'markets', 'must', 'build', 'via', 'draft'**]  
 [**'michael', 'jordan', 'dont', 'know', 'couldve', 'survived', 'twitter', 'era'**]

### **Local stems:**

{'low', 'preston', 'gifted', 'last', 'kemba', 'x', 'wow', 'introducing', 'signature', 'medical', 'dr', 'black', 'reviewing', 'awards', 'vie', 'aka', 'championship', 'new', 'lee', 'shuts', 'trevor', 'ability', 'beverly', 'greats', 'griner', 'voted', 'franchises', 'oj', 'bad', 'airness', 'cheerios', 'catches', 'nick', 'young', 'among', 'disputes', 'college', 'must', 'practice', 'goal', 'isiah', 'potential', 'tough', 'hops', 'cant', 'horrible', 'becomes', 'coach', 'labor', 'jarred', 'trying', 'record', 'wouldve', 'cross', 'wants', 'stock', 'pt', 'candidate', 'tennessees', 'left', 'failed', 'blames', 'kids', 'made', 'night', 'bone', 'pairs', 'ads', 'respect', 'mid', 'poole', 'parker', 'fights', 'charles', 'ie', 'score', 'pacers', 'priceless', 'hates', 'like', 'revenue', 'called', 'trey', 'hilarious', 'iv', 'landing', 'playable', 'posterizes', 'brittney', 'drunk', 'remembering', 'courts', 'kaman', 'mock', 'sources', 'shoes', 'appearance', 'criticism', 'tnt', 'sean', 'start', 'blank', 'em', 'thief', 'tmz', 'look', 'thrilling', 'amazing', 'rebounds', 'drexler', 'turn', 'bulls', 'grant'}

### **Vocabulary of stems:**

{'uconn': {'uconn'}, 'guard': {'guard', 'guards'}, 'jordan': {'jordan', 'jordans'}, 'hawkin': {'hawkins'}, 'declar': {'declares'}, 'nba': {'nba', 'nbas'}, 'draft': {'draft'}, 'announc': {'announces', 'announced'}, 'player': {'players', 'player'}, 'select': {'selected'}, 'rise': {'rising'}, 'star': {'stars', 'star'}, 'game': {'games', 'game'}, 'michael': {'michael'}, 'consid': {'consider', 'considering'}, 'sale': {'sale'}, 'hornet': {'hornets'}, 'deal': {'deal', 'deals'}, 'immin': {'imminent'}, 'pool': {'poole', 'pooles'}, 'talk': {'talk', 'talking', 'talks'}, 'role': {'role'}, 'warrior': {'warriors'}, 'come': {'coming', 'comes', 'come'}, 'bench': {'bench'}, 'turn': {'turns', 'turn'}, 'around': {'around'}, 'celtic': {'celtics'}, 'final': {'finally', 'final', 'finals'}, 'wizard': {'wizards'}, 'waiv': {'waive'}, 'barton': {'barton'}, 'convert': {'convert'}, 'goodwin': {'goodwin'}, 'standard': {'standard'}, 'contract': {'contracts', 'contract'}, 'start': {'starting', 'start'}, 'nascar': {'nascar'}, 'team': {'teaming', 'team'}, 'launch': {'launching', 'launch'}, 'fan': {'fans', 'fan'}}

### **Correlations:**

Jordan michael 1.433333333333333  
 Jordan game 0.333333333333333  
 Jordan value 0.333333333333333  
 Jordan prime 0.333333333333333  
 Jordan shoe 0.1666666666666666  
 Jordan nba 0.1666666666666666

### **Cluster Selection:**

Associative is chosen due to better results.

**Expanded Query Result:**

Jordan Michael game

**SCALAR CLUSTERING:**

**Query 1:**

Warriors

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

**Local Vocabulary:**

```
['jamal', 'murray', 'scores', '26', 'nuggets', 'hold', 'warriors', '112', '110']
['kings', 'rest', 'deaaron', 'fox', 'domantas', 'sabonis', 'kevin', 'huerter', 'vs', 'warriors']
['mavs', 'owner', 'mark', 'cuban', 'plans', 'protest', 'free', 'bucket', 'warriors']
['luka', 'doncic', 'back', 'kyrie', 'irving', 'mavericks', 'face', 'warriors']
['warriors', 'gary', 'payton', 'ii', 'begun', 'court', 'activities', 'intensify']
['jordan', 'poole', 'talks', 'role', 'warriors', 'coming', 'bench']
['warriors', 'sixers', 'never', 'close', 'ben', 'simmons', 'trade', 'talks']
['stephen', 'curry', 'draymond', 'green', 'still', 'define', 'warriors', 'greatness']
['thompson', 'scores', '42', 'points', '12', '3s', 'warriors', 'beat', 'thunder']
['aint', 'done', 'nba', 'champion', 'warriors', 'already', 'looking', 'ahead']
```

**Local stems:**

```
['jamal', 'murray', 'scores', '26', 'nuggets', 'hold', 'warriors', '112', '110', 'kings', 'rest', 'deaaron', 'fox',
'domantas', 'sabonis', 'kevin', 'huerter', 'vs', 'mavs', 'owner', 'mark', 'cuban', 'plans', 'protest', 'free', 'bucket',
'luka', 'doncic', 'back', 'kyrie', 'irving', 'mavericks', 'face', 'gary', 'payton', 'ii', 'begun', 'court', 'activities',
'intensify', 'jordan', 'poole', 'talks', 'role', 'coming', 'bench', 'sixers', 'never', 'close', 'ben', 'simmons', 'trade',
'stephen', 'curry', 'draymond', 'green', 'still', 'define', 'greatness', 'thompson', '42', 'points', '12', '3s', 'beat',
'thunder', 'aint', 'done', 'nba', 'champion', 'already', 'looking', 'ahead', 'hawks', '119', 'golden', 'states']
```

**Vocabulary of stems:**

```
'trad': {'trades', 'trade', 'trading'}, 'steph': {'stephen', 'stephs', 'steph'}, 'curry': {'currys', 'curry'}, 'green':
{'green', 'greens'}, 'gre': {'greatness', 'great'}, 'thompson': {'thompson', 'thompsons'}, 'point': {'point',
'pointers', 'points'}, 'beat': {'beat', 'beating', 'beats'}, 'nba': {'nba', 'nbas'}, 'champ': {'champions',
```

'championship', 'champion', 'championships', 'champs'}, 'look': {'looks', 'looking'}, 'stat': {'status', 'stats', 'states', 'state'}, 'break': {'break', 'breaking'}, 'play': {'play', 'plays', 'player', 'players', 'playing'}, 'gam': {'games', 'game'}, 'vict': {'victory', 'victor'}, 'star': {'star', 'stars'}, 'fin': {'financial', 'final', 'fine', 'finish', 'finals'},

**Correlations:**

[(jordans, 'warriors', 0.3333333333333333),  
(loss, 'warriors', 0.14972527472527472),  
('steph', 'warriors', 0.14484679665738162),  
('vs', 'warriors', 0.13934426229508196),  
('game', 'warriors', 0.13275613275613277),  
('stephen', 'warriors', 0.13043478260869565),  
('golden', 'warriors', 0.11940298507462686),

**Cluster Selection:**

Scalar does not provide good results. So Associative is chosen.

**Expanded Query Result:**

Warriors jordans loss steph

**Query 2:**

Playoffs

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

**Local Vocabulary:**

['knicks', 'clinch', 'playoff', 'berth', '118', '109', 'win', 'wizards']  
['nets', 'hold', 'jazz', '111', '110', 'move', 'closer', 'playoff', 'berth']  
['mavericks', 'sit', 'kyrie', 'irving', '4', 'others', 'playoff', 'hopes', 'still', 'alive']  
['success', 'starved', 'kings', 'close', 'long', 'awaited', 'playoff', 'berth']  
['nets', 'beat', 'pistons', '123', '108', 'close', 'playoff', 'berth']  
['grizzlies', 'earn', 'playoff', 'berth', 'kennard', 'makes', '10', '3', 'pointers']  
['2022', 'wnba', 'playoffs', 'previewing', 'second', 'round', 'matchups']  
['2022', 'wnba', 'playoffs', 'one', 'key', 'stat', 'know', 'team']

['making', 'oklahoma', 'city', 'thunders', 'case', 'top', '6', 'nba', 'playoff', 'seed']  
 ['austin', 'reaves', 'rising', 'challenges', 'lakers', 'playoff', 'push']

### Local stems:

front', 'crowd', 'stadium', 'attendance', 'shaking', 'rust', 'white', 'house', 'biden', 'modern', 'call', 'shorter', 'sitting', 'jersey', 'half', 'blasts', 'execution', 'domino', 'effect', 'picture', 'transform', 'enough', 'land', 'ship', 'cost', 'cutting', 'parlay', 'things', 'forgot', 'dfs', 'survived', 'put', 'position', 'claps', 'dillon', 'brooks', 'sure', 'result', 'shoots', 'lights', 'splash', 'pre', 'self', 'tyus', 'jones', 'earns', 'puzzle', 'complete', 'trolls', 'solve', 'lose', 'woes', 'thrilling', 'rejected', 'claim', 'controversial', 'unhappy', 'collective', 'bargaining', 'family', 'matters', 'storyline', 'series', 'wears', 'sleeve', 'arm', 'wanted', 'bracket', 'draws', 'parallels', 'steal', 'emotional', 'aims', 'los', 'angeles', 'surges', 'voting', 'big', 'failures', 'failing', 'attempt', 'relives', 'online', 'rocket', 'science', 'force', 'physical', 'installed', 'favorites', 'seven', 'dynasty', 'brink', 'unraveling', 'stakes', 'hoops', 'necessary', 'suffocating', 'wills', 'frustrated', 'pulled', 'understands', 'roll', 'men', 'wait', 'mode', 'caught', 'scrum', 'poised', 'foundering', 'lock', 'besides', 'stop', 'third', 'singular', 'focus', 'nemeses', 'seals', 'joy', 'secret', 'rage', 'ignites', 'francisco', 'share', 'crazier', 'famous', 'decimated', 'unfinished', 'business', 'torn

### Vocabulary of stems:

'pick': {'pick', 'picks'}, 'er': {'era', 'eric'}, 'discuss': {'discussed', 'discusses'}, 'ag': {'agency', 'ages', 'agent'}, 'award': {'awards', 'award'}, 'pac': {'pacific', 'pacers'}, 'interest': {'interest', 'interested'}, 'hir': {'hiring', 'hire'}, 'coach': {'coach', 'coaches'}, 'vet': {'veteran', 'veterans'}, 'say': {'says', 'say'}, 'test': {'test', 'tests', 'tested'}, 'posit': {'positive', 'position'}, 'wisem': {'wiseman', 'wisemans'}, 'lin': {'lins', 'lin'}, 'day': {'day', 'days'}, 'valu': {'values', 'value', 'valuable'}, 'franch': {'franchise', 'franchises'}, 'show': {'shows', 'show', 'showing'}, 'join': {'joining', 'join'}, 'dur': {'durant', 'durants'}, 'set': {'set', 'sets'}, 'draw': {'drawing', 'draws'}, 'allow': {'allow', 'allows'}, 'fan': {'fans', 'fan'}, 'miss': {'miss', 'misses'}, 'town': {'towns', 'town'}, 'rocket': {'rockets', 'rocket'}, 'expect': {'expected', 'expect'}, 'hap': {'happened', 'happens'}, 'rep': {'repeat', 'repeats'}, 'find': {'find', 'finding'}, 'deny': {'denied', 'denies'}, 'assign': {'assigned', 'assignment'}, 'us': {'us', 'use'}, 'rost': {'rosters', 'roster'}, 'mil': {'million', 'miller'}, 'punch': {'punches', 'punching', 'punch'}, 'stev': {'steve', 'stevens'}, 'ker': {'kerr', 'kerrs'}, 'sens': {'sensation', 'sense'}, 'start': {'starting', 'start', 'starts'}, 'new': {'new', 'newly', 'news'}, 'fal': {'fall', 'falls'}, 'clear': {'cleared', 'clears', 'clearance'}, 'hit': {'hitting', 'hit', 'hits'}, 'end': {'end', 'ends'}, 'report': {'report', 'reportedly'}, 'avoid': {'avoid', 'avoiding'}

### Correlations:

('playoff', 'playoffs', 0.26549491211840887),  
 ('2022', 'playoffs', 0.13533834586466165),  
 ('2021', 'playoffs', 0.1220159151193634),  
 ('round', 'playoffs', 0.11081794195250659),  
 ('series', 'playoffs', 0.1),  
 ('game', 'playoffs', 0.0859375),  
 ('first', 'playoffs', 0.08262108262108261),  
 ('vs', 'playoffs', 0.07584269662921349),

**Cluster Selection:**

Associative

**Expanded Query Result:**

Playoffs playoff 2022 2021

**Query 3:**

Shaq

**Local Document set:**

Index of data which is used as a Pandas DataFrame. (1-500)

**Local Vocabulary:**

```
['shaq', 'thanks', 'kareem', 'dr', 'j', 'magic', 'barkley', 'passionate', 'speech']
['shaq', 'shares', 'dealt', 'crazy', 'fans', 'sacramento']
['free', 'agent', 'guard', 'shaq', 'harrison', 'signs', 'training', 'camp', 'deal', '76ers']
['offered', 'pay', 'dads', 'funeral', 'shaq', 'best', 'star', 'teammate', 'ever']
['arenas', 'shaq', 'barkley', 'hypocrites', 'calling', 'ben', 'simmons', 'diva']
['shaq', 'barkleys', 'misinformed', 'comments', 'disrespectful']
['posecast', 'udonis', 'haslem', 'heat', 'culture', 'playing', 'shaq', 'makes', 'lebron', 'great']
['shaq', 'chuck', 'inside', 'nba', 'compromise']
['damian', 'lillard', 'raps', 'back', 'shaq', 'kobe', 'rings']
['lakers', 'sign', 'tristan', 'thompson', 'shaq', 'harrison', 'nba', 'playoffs']
```

**Local stems:**

```
stopper', 'defend', 'better', 'anyone', 'spencer', 'dinwiddie', 'ejected', 'elbowing', 'watch', 'channel', 'live',
'stream', 'tuesday', 'sporting', 'podcast', 'shohei', 'ohtani', 'novak', 'djokovic', 'wimbledon', 'overall', 'las',
'vegas', 'summer', 'johnny', 'future', 'among', 'biggest', 'losers', 'weighs', 'possible', 'burning', 'hypothetical',
'reunion', 'werent', 'williams', 'reflects', 'vaccine', 'didnt', 'like', 'wasnt', 'choice', 'divincenzo', 'bring', 'agent',
'bolster', 'mac', 'mcclung', 'youtube', 'sensation', 'espy', 'home', 'pays', 'tribute', 'bryant', 'espy', 'every',
'day', 'fined', 'luxury', 'unfair', 'system', 'michael', 'tried', 'recruit', 'demar', 'derozan', 'offseason', 'threw',
'wild', 'pitch', 'oakland', 'athletics', 'details', 'mainstay', 'max', 'explore', 'shouldnt', 'fly', 'radar', 'candidates',
'financial', 'worth', 'tough', 'decision', 'looming', 'wood', 'value', 'isnt', 'five', 'according', 'opening', 'host',
'holiday', 'schedule', 'easiest', 'final', 'prediction', 'dates', 'national', 'champs', 'rumors', 'supercharged', 'took',
'defensive'
```

**Vocabulary of stems:**

fight': {'fights', 'fight', 'fighting'}, 'fil': {'file', 'fill'}, 'ad': {'ad', 'adds'}, 'surg': {'surging', 'surges'}, 'put': {'put', 'puts'}, 'preview': {'preview', 'previews'}, 'watch': {'watch', 'watching'}, 'dram': {'dramatic', 'drama'}, 'nee': {'needs', 'need'}, 'pay': {'paying', 'pays'}, 'want': {'wants', 'wanted'}, 'shop': {'shop', 'shopping'}, 'disciplin': {'disciplining', 'discipline'}, 'cal': {'calls', 'call'}, 'sal': {'sales', 'sale'}, 'rel': {'relishes', 'relives'}, 'shot': {'shot', 'shots'}, 'forc': {'forces', 'force'}, 'confirm': {'confirms', 'confirm'}, 'hurt': {'hurt', 'hurts'}, 'lebron': {'lebron', 'lebrons'}, 'shin': {'shine', 'shines'}, 'push': {'push', 'pushing', 'pushed'}, 'ey': {'eyes', 'eye'}, 'playoff': {'playoffs', 'playoff'}, 'stop': {'stopper', 'stop'}, 'defend': {'defend', 'defends', 'defending', 'defensive'}, 'bet': {'bets', 'better'}, 'fut': {'future', 'futures'}, 'espy': {'espy', 'espy'}, 'detail': {'detail', 'details'}}

**Correlations:**

('nba', 'shaq', 0.10154525386313466),  
 ('kobe', 'shaq', 0.08525345622119816),  
 ('barkley', 'shaq', 0.07746478873239436),  
 ('say', 'shaq', 0.07746478873239436),  
 ('best', 'shaq', 0.056790123456790124),  
 ('inside', 'shaq', 0.05609756097560976),  
 ('lakers', 'shaq', 0.04773869346733668),

**Cluster Selection:**

Associative cluster is chosen because it has more relevant results.

**Expanded Query Result:**

Shaq nba kobe barkley

**Results of the search engine on your expanded queries**

The following images showcase the EXPANDED query results of our search engine. There are several categories in which a user might search a query related to basketball:

1. Player name
2. Team name
3. Basketball related news
4. Match score

Query expansion of all of these queries might include things like:

1. Expanding player name
2. Expanding to team
3. Expanding to a particular game
4. Expanding to most searched term

Expanding player name:

Kyrie (a player) name gets expanded to include his last name Irving.

Type your query...	--Choose one option--	Search
--------------------	-----------------------	--------

**Results for: *kyrie***

## Our Search Engine Results

### Query expansion

kyrie irving

- [Irving has 26, Nets beat Lakers 121-104 as James, Davis sit](#)

<https://www.basketballnews.com/stories/irving-has-26-nets-beat-lakers-121104-as-james-davis-sit>

- [Woj: Kevin Durant will request trade if Kyrie leaves Brooklyn](#)

<https://www.basketballnews.com/stories/woj-kevin-durant-will-request-trade-if-kyrie-leaves-brooklyn>

- [Harden isn't surprised Durant, Irving fled Nets' 'dysfunction'](#)

<https://www.basketballnews.com/stories/harden-not-surprised-durant-irving-fled-nets-dysfunction>

---

This even includes players with middle names.

Abdul Kareem (a player) name gets expanded to include his last name Jabbar.

Type your query... --Choose one option-- Search

Results for: *abdur kareem*

#### Our Search Engine Results

##### Query expansion

abdul kareem jabbar

- [What LeBron passing Kareem means for NBA\\_Goat debate](#)

<https://www.basketballnews.com/stories/what-lebron-passing-kareem-means-for-nba-goat-debate>

- [LeBron James passes Kareem Abdul-Jabbar as NBA all-time scoring leader](#)

<https://www.basketballnews.com/stories/nba-breaking-news-lebron-james-passes-kareem-abduljabbar-as-nba-all-time-scoring-leader-los-angeles-lakers-individual-record-akron-ohio-cleveland-cavaliers-miami-heat>

- [What LeBron passing Kareem means to NBA\\_greatest debate](#)

<https://www.basketballnews.com/stories/what-lebron-passing-kareem-means-to-nba-greatest-debate>

---

Luka gets expanded to include Luka Doncic

Type your query... --Choose one option-- Search

Results for: *luka*

#### Our Search Engine Results

##### Query expansion

luka doncic

- [It's Embiid vs. Doncic for scoring title in NBA's final week](#)

<https://www.basketballnews.com/stories/its-embiid-vs-doncic-for-scoring-title-in-nbas-final-week>

- [Doncic scores 53 in return, Mavericks beat Pistons 111-105](#)

<https://www.basketballnews.com/stories/doncic-scores-53-in-return-mavericks-beat-pistons-111105>

- [Green, Hardy help Mavs surprise Jazz 124-111 without Doncic](#)

<https://www.basketballnews.com/stories/green-hardy-help-mavs-surprise-jazz-124111-without-doncic>

---

Sometimes the search query expands to the team name instead of the player. Antetokounmpo, a player's name, expands to Antetokounmpo Bucks which is the team he plays for.

**Results for: *antetokounmpo***

## Our Search Engine Results

### Query expansion

antetokounmpo bucks

- [Bucks beat Hornets 124-115 for 5th consecutive victory](#)

<https://www.basketballnews.com/stories/bucks-beat-hornets-124115-for-5th-consecutive-victory>

- [Bucks rip Heat 128-99 after Giannis exits early with knee issue](#)

<https://www.basketballnews.com/stories/bucks-rip-heat-12899-after-giannis-exits-early-with-knee-issue>

- [Bucks outlast Suns 104-101 for 14th consecutive victory](#)

<https://www.basketballnews.com/stories/bucks-outlast-suns-104101-for-14th-consecutive-victory>

---

Searching last name Curry brings the first name of the player Stephen

**Results for: *curry***

## Our Search Engine Results

### Query expansion

curry stephen

- [Steph Curry resumes on-court work, still sidelined by leg injury](#)

<https://www.basketballnews.com/stories/curry-resumes-oncourt-work-still-sidelined-by-leg-injury>

- [Steph Curry aims to inspire with 'I Have a Superpower' book](#)

<https://www.basketballnews.com/stories/steph-curry-aims-to-inspire-with-i-have-a-superpower-book>

- [Steph Curry has generated 1.4 billion video views across NBA socials](#)

<https://www.basketballnews.com/stories/steph-curry-has-generated-14-billion-video-views-across-nba-socials>

---

**Expanding to team name**

Sometimes the query expands to team name. This indicates that some times the most relevant expansion is the team and the next result would be the last name.

**Results for: durant**

## Our Search Engine Results

### Query expansion

durant nets

- [Kevin Durant scores 35 points, Suns roll past Thunder](#)

<https://www.basketballnews.com/stories/kevin-durant-scores-35-points-suns-roll-past-thunder>

- [Back home in Canada, Andrew Nembhard nets 25 as Pacers top Raptors](#)

<https://www.basketballnews.com/stories/back-home-in-canada-andrew-nembhard-nets-25-as-pacers-top-raptors>

- [Nets hold off Jazz 111-110, move closer to playoff berth](#)

<https://www.basketballnews.com/stories/nets-hold-off-jazz-111110-move-closer-to-playoff-berth>

---

Giannis expands to bucks (Team) and Luka Dallas expands to Mavericks (complete team name)

**Results for: giannis**

## Our Search Engine Results

### Query expansion

giannis bucks

- [Bucks beat Hornets 124-115 for 5th consecutive victory](#)

<https://www.basketballnews.com/stories/bucks-beat-hornets-124115-for-5th-consecutive-victory>

- [Antetokounmpo's triple-double lifts Bucks over Wizards](#)

<https://www.basketballnews.com/stories/antetokounmpos-tripledouble-lifts-bucks-over-wizards>

- [Bucks outlast Suns 104-101 for 14th consecutive victory](#)

<https://www.basketballnews.com/stories/bucks-outlast-suns-104101-for-14th-consecutive-victory>

---

**Results for: luka dallas**

### Our Search Engine Results

#### Query expansion

luka dallas mavericks

- [Nicolas Batum's block lifts France over Luka Doncic, Slovenia 90-89](#)

<https://www.basketballnews.com/stories/nicolas-batum-block-lifts-france-over-luka-doncic-slovenia-90-89-2020-tokyo-olympics>

- [Trae Young lifts Hawks past Mavericks in overtime, 132-130](#)

<https://www.basketballnews.com/stories/trae-young-lifts-hawks-past-mavericks-in-overtime-132130>

- [Mavericks sit Kyrie Irving, 4 others with playoff hopes still alive](#)

<https://www.basketballnews.com/stories/mavericks-sit-kyrie-irving-4-others-with-playoff-hopes-still-alive>

---

---

Searching phoenix return Phoenix Suns

**Results for: phoenix**

### Our Search Engine Results

#### Query expansion

phoenix suns

- [Kevin Durant scores 35 points, Suns roll past Thunder](#)

<https://www.basketballnews.com/stories/kevin-durant-scores-35-points-suns-roll-past-thunder>

- [Mikal Bridges scores 25, leads Suns to 106-94 win over Celtics](#)

<https://www.basketballnews.com/stories/mikal-bridges-scores-25-leads-suns-to-10694-win-over-celestics>

- [Devin Booker scores 25 points, Suns beat Thunder 124-115](#)

<https://www.basketballnews.com/stories/devin-booker-scores-25-points-suns-beat-thunder-124115>

---

---

Detroit gives Pistons as result

**Results for: *detroit***

## Our Search Engine Results

### Query expansion

detroit pistons

- [Nets beat Pistons 123-108, close in on playoff berth](#)

<https://www.basketballnews.com/stories/nets-beat-pistons-123108-close-in-on-playoff-berth>

- [Jimmy Butler leads Heat to late win against Pistons, 118-105](#)

<https://www.basketballnews.com/stories/jimmy-butler-leads-heat-to-late-win-against-pistons-118105>

- [Doncic scores 53 in return, Mavericks beat Pistons 111-105](#)

<https://www.basketballnews.com/stories/doncic-scores-53-in-return-mavericks-beat-pistons-111105>

---

For several of the following queries which are basketball related organizations or events, QE is done.

**Results for: *nets kevin***

## Our Search Engine Results

### Query expansion

nets kevin brooklyn

- [Kevin Durant scores 35 points, Suns roll past Thunder](#)

<https://www.basketballnews.com/stories/kevin-durant-scores-35-points-suns-roll-past-thunder>

- [Hawks trade Kevin Huerter to Kings for Justin Holiday, Mo Harkless](#)

<https://www.basketballnews.com/stories/hawks-trade-kevin-huerter-to-kings-for-justin-holiday-mo-harkless-nba-offseason-trade>

- [Back home in Canada, Andrew Nembhard nets 25 as Pacers top Raptors](#)

<https://www.basketballnews.com/stories/back-home-in-canada-andrew-nembhard-nets-25-as-pacers-top-raptors>

---

Type your query...

Results for: *ncaa record*

## Our Search Engine Results

### Query expansion

ncaa record tournament

- [The longest an NCAA bracket has ever stayed perfect](#)

[https://www.ncaa.com/news/basketball-men/bracketiq/2023-03-17/longest-ncaa-bracket-has-ever-stayed-perfect?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/bracketiq/2023-03-17/longest-ncaa-bracket-has-ever-stayed-perfect?utm_campaign=inline-article)

- [Boston sets record, No. 1 Gamecocks top Arkansas 92-46](#)

<https://www.basketballnews.com/stories/boston-sets-record-no-1-gamecocks-top-arkansas-9246>

- [LeBron James closes in on NBA record, Lakers falls to Pelicans](#)

<https://www.basketballnews.com/stories/lebron-james-closes-in-on-nba-record-lakers-falls-to-pelicans>

---

Type your query...

Results for: *ncaa*

## Our Search Engine Results

### Query expansion

ncaa team

- [NCAA DII bracket 2019: Printable DII men's basketball tournament bracket .PDF](#)

[https://www.ncaa.com/news/basketball-men/article/2019-03-20/ncaa-dii-bracket-2019-printable-dii-mens-basketball?utm\\_campaign=d2-inline](https://www.ncaa.com/news/basketball-men/article/2019-03-20/ncaa-dii-bracket-2019-printable-dii-mens-basketball?utm_campaign=d2-inline)

- [These 5 teams could challenge Northwest Missouri State in the MIAA](#)

[https://www.ncaa.com/news/basketball-men/article/2020-11-12/5-teams-who-could-challenge-northwest-missouri-state-miaa?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2020-11-12/5-teams-who-could-challenge-northwest-missouri-state-miaa?utm_campaign=inline-article)

- [8 ways the 1982 NCAA tournament was different than it is today](#)

[https://www.ncaa.com/news/basketball-men/article/2020-04-21/8-ways-1982-ncaa-tournament-was-different-it-today?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2020-04-21/8-ways-1982-ncaa-tournament-was-different-it-today?utm_campaign=inline-article)

---

**Results for: nba**

## Our Search Engine Results

### Query expansion

nba players

- [College basketball: No. 3 Gonzaga's Hachimura thriving as go-to player](#)  
<https://www.ncaa.com/news/basketball-men/article/2018-11-23/college-basketball-no-3-gonzagas-hachimura-thriving-go>
  - [Tennessee's Grant Williams wins March Madness' Preseason Player of the Year](#)  
<https://www.ncaa.com/news/basketball-men/article/2018-10-26/tennessees-grant-williams-wins-march-madness-preseason>
  - [Ja Morant and R.J. Barrett are expected to be the No. 2 and No. 3 picks in the NBA Draft. Let's break down their college games](#)  
<https://www.ncaa.com/news/basketball-men/article/2019-06-19/ja-morant-and-rj-barrett-are-expected-be-no-2-and-no-3-picks>
- 

**Results for: mens**

## Our Search Engine Results

### Query expansion

mens basketball

- [Saturday's college basketball upsets \(and near-upsets\) showed again why pandemonium is this season's norm](#)  
[https://www.ncaa.com/news/basketball-men/article/2020-01-19/saturdays-college-basketball-upsets-and-near-upsets-showed?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2020-01-19/saturdays-college-basketball-upsets-and-near-upsets-showed?utm_campaign=inline-article)
  - [Kansas vs. Gonzaga predicted, the Maui Invitational winner and 5 more college basketball games to watch this week](#)  
<https://www.ncaa.com/news/basketball-men/article/2020-11-26/kansas-vs-gonzaga-predicted-maui-invitational-winner-and-5-more-college>
  - [10 college basketball players primed for breakout seasons in 2020-21](#)  
[https://www.ncaa.com/news/basketball-men/article/2020-10-20/10-college-basketball-players-primed-breakout-seasons-2020-21?utm\\_campaign=inline-article](https://www.ncaa.com/news/basketball-men/article/2020-10-20/10-college-basketball-players-primed-breakout-seasons-2020-21?utm_campaign=inline-article)
-

**Results for: lakers**

## Our Search Engine Results

### Query expansion

lakers los

- [Davis has 40, LeBron has triple-double as Lakers beat Rockets](#)

<https://www.basketballnews.com/stories/davis-has-40-lebron-has-tripledouble-as-lakers-beat-rockets>

- [James has triple-double, leads Lakers past Knicks in OT](#)

<https://www.basketballnews.com/stories/james-has-tripledouble-leads-lakers-past-knicks-in-ot>

- [John Wall to Los Angeles Clippers on two-year deal](#)

<https://www.basketballnews.com/stories/john-wall-to-los-angeles-clippers-on-twoyear-deal-nba-offseason-free-agency>

---

**Results for: hornets raptors**

## Our Search Engine Results

### Query expansion

hornets raptors toronto

- [Markkanen scores 25 as Jazz beat Hornets, who make only 2 3s](#)

<https://www.basketballnews.com/stories/markkanen-scores-25-as-jazz-beat-hornets-who-make-only-2-3s>

- [Bucks beat Hornets 124-115 for 5th consecutive victory](#)

<https://www.basketballnews.com/stories/bucks-beat-hornets-124115-for-5th-consecutive-victory>

- [Fred VanVleet has 32 as Raptors get 117-111 win over Rockets](#)

<https://www.basketballnews.com/stories/fred-vanvleet-has-32-as-raptors-get-117111-win-over-rockets>

---

**Elaborate on how you have collaborated with the student responsible for the user interface to expose the results of your expanded queries as well**

- I collaborated with Vaishnavi, who worked on the 4th part (User Interface) to display the extended query of my result.

- I get the top 500 ranked relevant documents from the relevance model (PageRank) done by Allama who worked on the 3rd part (Relevance Model) and use it as input to my query expansion module - Associative, Metric, Scalar.
- The expanded query output from my module is then shared with Vaishnavi who uses the relevance model to run the expanded query again and produce the results for the query which are stored as list of dictionaries.
- The expanded result is also exposed in the UI with the relevant html tag.

**Demo queries:**

**Query:** Giannis

**QE:** Giannis Buck

**Query:** Luka

**QE:** Luka Doncic

**Query:** Curry Golden

**QE:** Curry Golden Steph

These queries have been particularly selected to showcase the difference in results between query relevance results and QE results.

For eg: Giannis as a search term would return results about the player while Giannis Buck would return results about the player as well as the recent matches in which he played.