

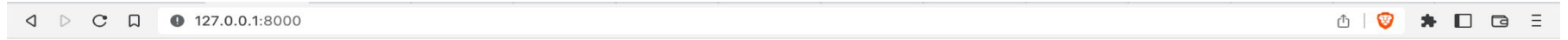
# Retrieval Engine by Team 5 for SSD Term Project, Monsoon '22

Team Members :-

- Rahul Padhy (2022201003)
- Ujjwal Prakash (2022202009)
- Gunank Singh Jakhar (2022201057)
- Venkata G Srikan Pulipaka (2022204004)

Guide - Dr. Charu Sharma

# Opening page - index.html



[Search](#) [Images](#) [News](#) [Videos](#) [Movies](#)



INTERNATIONAL INSTITUTE OF  
INFORMATION TECHNOLOGY  
HYDERABAD


T5 Search

## Opening page - index.html (continued...)

- Consists of a logo (IIITH) and a search bar in the middle.
- Acts as default for global text-based search.
- Contains links for other (deep) search-based pages in the top right corner, which upon clicking redirects to the appropriate page.
- Searching for any term brings search results and redirects to the search page.
- The footer contains links to some commonly visited websites
  - Right part contains link to the source code hosted on Github.
  - Left part contains links to LinkedIn, Github, Instagram, Twitter and Reddit.

# Global text-based search - search.html

127.0.0.1:8000/search



INTERNATIONAL INSTITUTE OF  
INFORMATION TECHNOLOGY  
HYDERABAD

T5 Search

Web Results

### What Is a Formula Equation?

[https://www.reference.com/world-view/formula-equation-a36486fa6ab80ab7?utm\\_content=params%3Ao%3D740005%26ad%3DdirN%26qo%3DserpIndex&uid=02c74edc-0586-4742-9401-80a3b77ffe5f](https://www.reference.com/world-view/formula-equation-a36486fa6ab80ab7?utm_content=params%3Ao%3D740005%26ad%3DdirN%26qo%3DserpIndex&uid=02c74edc-0586-4742-9401-80a3b77ffe5f)

A formula equation is a visual representation of a reaction using chemical formulas. A chemical formula is an expression that states the number and types of atoms that make up any given molecule using the symbols for the elements and sub-sc...

### What Is the Formula for Acceleration?

[https://www.reference.com/science-technology/formula-acceleration-6f1fd84105328f01?utm\\_content=params%3Ao%3D740005%26ad%3DdirN%26qo%3DserpIndex&uid=02c74edc-0586-4742-9401-80a3b77ffe5f](https://www.reference.com/science-technology/formula-acceleration-6f1fd84105328f01?utm_content=params%3Ao%3D740005%26ad%3DdirN%26qo%3DserpIndex&uid=02c74edc-0586-4742-9401-80a3b77ffe5f)

The formula for acceleration is given as  $a = (v_2 - v_1) / (t_2 - t_1)$ , where “a” denotes the acceleration, “v2” indicates the final velocity, “v1” represents the initial velocity and “t2 – t1” is the time interval between the final and initial...

### F1 Schedule 2022 - Official Calendar of Grand Prix Races

<https://www.formula1.com/en/racing/2022.html>

Check out the Formula 1 calendar for 2022, including the inaugural Miami ...

LinkedInGithubInstagramTwitterReddit

Source

# Global text based search - TFIDF

- All category search based on keywords.
- Search results ranked based on tf-idf algorithm.
- Web pages scraped and data cleanup performed to extract the text content of pages on which we make index for tf-idf.
- Also, removal of stopwords, stemming and lemmatization performed before applying tf-idf model.

# Image search

- Relevant images based on keywords provided by the user
- Scraping done from google images.



INTERNATIONAL INSTITUTE OF  
INFORMATION TECHNOLOGY

HYDERABAD

T5 Image Search

Image Results



# News Search

127.0.0.1:8000/newsSearch



INTERNATIONAL INSTITUTE OF  
INFORMATION TECHNOLOGY

HYDERABAD

Country:  Topics:

T5 News

News Results

## [Kenny Pickett, Steelers squander lead but close out Colts - ESPN](#)

The same issues that plagued the Steelers throughout the season cropped up throughout the game -- red zone efficiency, giving up big plays -- but this time, the Steelers held the Colts on fourth-and-three to force a turnover on downs with 30 seconds remaining to seal the win. As the offense stagnated, the defense, which held the Colts to zero yards in the first quarter, got worn down and lost the lead -- and nearly gave up two touchdowns, if not for a Colts fumble near the goal line. Troubling trend: After amassing 232 yards and 16 first downs in the first half for a 16-3 lead, the Steelers ran just eight offensive plays in the third quarter and gained just 18 total yards. Only the Rams (0-7) have more losses without a win when trailing after three quarters this season. The Steelers' lead evaporated after halftime as the defense gave up 152 third-quarter yards to Matt Ryan and the Colts, and the Steelers failed to move the ball.

## [A lingering question as USA preps for Iran: What's up with Gio Reyna? - FOX Sports](#)

Iran vs. United States Preview: Will the USMNT make it out of the Group Stage? Read more from the World Cup: Top stories from FOX Sports: Laken Litman covers college football, college basketball and soccer for FOX Sports. As the U.S. faces World Cup elimination on Tuesday in its third and final group stage match against Iran at Al Thumama Stadium (2 p.m. ET on FOX and the FOX Sports App), a question lingers: What's the deal with Reyna? USMNT's Gio Reyna on his personal journey to the 2022 FIFA World Cup. The Americans have to win Tuesday to advance to the knockout stage of the tournament." [ Forget tactics and lineups: For USMNT, it's about having right mentality ] The 6-foot-1 Reyna is a dangerous attacking player when healthy. He acknowledged his team could have been better in the penalty box, but "it's up to the rest of the group to give them quality service so that they can finish off some opportunities." As far as Reyna goes, Berhalter had this to say: "I think in terms of alternatives at the striker position, we are comfortable with the three that we have.

## [Did Cristiano Ronaldo Head The Ball For Portugal's First Goal vs Uruguay? Watch Video - NDTV Sports](#)

## News Search (continued...)

- News data is being obtained from Google News.
- News can be seen on the basis of country-wise and topic-wise search selection.
- Ranking has been done on the basis of sum of metascores and ratings.
- Article name and embedded link is shown at the top, alongwith individual article summaries.
  - Summaries have been extracted using `spacy.en_core_web_sm`.



# Video Search



INTERNATIONAL INSTITUTE OF  
INFORMATION TECHNOLOGY

HYDERABAD

fifa world cup

T5 Video Search

Video Results

**FIFA Collect: Feed Your Fandom**

**FIFApplusCollect**

**YouTube**

**2 months ago**

**11.3M views**

**WIN a trip for 2 to FIFA World Cup Qatar 2022™**

**FIFApplusCollect**

**Facebook**

**3 weeks ago**

**10.1M views**

**Predicting World Cup 2022: Ping Pong Style 🏓🏆**

**GiveMeSport**

# Video Search

- Scraping has been done from Bing videos, which contains videos from multiple sources such as YouTube, FaceBook, dailymotion, Instagram, etc.
- Ranking has been done on the basis of the number of views for a particular video.

# Movie Search

127.0.0.1:8000/movieSearch



INFORMATION TECHNOLOGY

HYDERABAD

Genres:

T5 Movie Search

Movie Results

**The Banshees of Inisherin**

(2022)

R

['Comedy', 'Drama']

114 min

**Once Upon a Time in Hollywood**

(2019)

R

['Comedy', 'Drama']

161 min

**Knives Out**

(2019)

PG-13

['Comedy', 'Crime', 'Drama']

130 min

# Movie Search

- Scraping is done on the basis of option / genre provided by the user as selection input.
- Scraping has been done from imdb.
- Ranking has been done on the basis of imdb ratings and metascores.

# Limitations

- Latency issue is prevalent in search results and news search results, due to scraping.
- Text processing and summarization take a bit of time to execute since the entire content of web page needs to be processed on the backend.

THANK YOU!!!