

# IR PROJECT PROPOSAL

**TOPIC: Sports History and Records Archives**



**AUTHORS:**

**VANSHAJ SHARMA(MT23103)**

**PULKIT RIHANI (MT23066)**

**SIDDHANT JHA (MT23097)**

**BHARAT NAGDEV (MT23029)**

**LOKESH SAINI (MT23120)**

**RITESH RAJPUT (MT23075)**

# **Introduction**

## **Background/Motivation**

System that leverages information retrieval to provide comprehensive archives of sports history, records, and memorable moments, allowing fans to explore and relive iconic sporting events, that would include images , video clips and brief history about the events and where it took place.

Evolution of sports seems to be an incredible journey over the past years especially in India. In today's growing Era, sports history, records and memorable moments hold a vital position in human society and shapes human history, So far there is no chatbot or a centralized platform to access this wealth of information. Holding such information including sports History, records and its memorable moments preserves cultural heritage and great inspiration for future generations, here is where LLM and Machine Learning based the Sports History and Records Archives Chabot comes into play, processing query optimization using the Machine Learning Technique and integrating Large Language Models for efficient Data Retrieval and providing users the best insights for the best memorable moments the sports journey and its history furnishing with related images and video clips. Our main purpose is to make an efficient and optimized information retrieval system for sports events, history, records and its memorable moments.

## **Existing work:**

Over past few year following researches has been done:

<https://www.thesportsdb.com/>

[https://www.academia.edu/77319254/Context\\_Based\\_Information\\_Retrieval\\_of\\_AthleticSport\\_Management\\_System\\_ASMS](https://www.academia.edu/77319254/Context_Based_Information_Retrieval_of_AthleticSport_Management_System_ASMS)

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5572265>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7363064>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1527838>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8622592>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10043234>

In the existing work , researchers have tried various machine learning techniques to implement the sport retrieval model but till now no one has implemented machine learning for query updation/optimization and Large language model for efficient query retrieval and natural language processing. Our information retrieval system tried to club the benefits of both the Large language model and machine learning to optimize the search results based on user preferences.

## **Methodology**

- **Integrating Large Language Model for High-Performance Computing :**

<https://dl.acm.org/doi/10.1145/3624062.3624172>

- **Toward harnessing user feedback for machine learning**

<https://dl.acm.org/doi/pdf/10.1145/1216295.1216316>

- **Actionable Data Insights for Machine Learning**

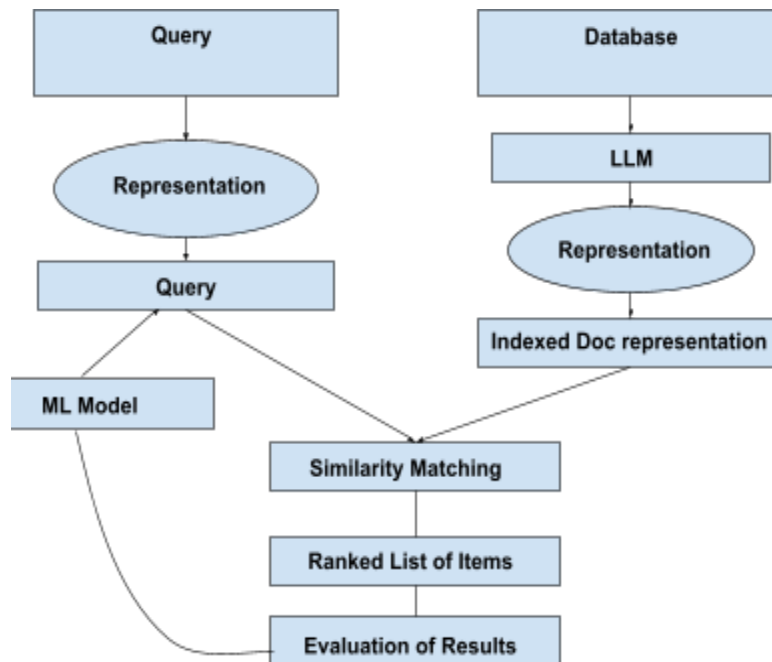
<https://dl.acm.org/doi/pdf/10.1145/3578356.3592581>

In this we will use Machine learning model for query modification based on user feedback and Large Language Model for performing various

**Flow will be :-**

- Text Query from user
- Changing representation as per internal representation
- Giving Query to fine tuned LLM using a large database built using web scraping on data available on the web related to sports events.
- Converting results from LLM to internal representation
- Similarity matching
- Ranking results based on knowledge graph , user cache and user search history
- Based on users direct and indirect feedback evaluation will be done

- Using Machine learning model we will be updating query
- Again fetching results from our information retrieval model based on updated query
- Showing results to user again



**Results** will be in **text , video and images** format

## Evaluation Parameters : Usefulness of Method

**Citation :-** <https://analyticsindiamag.com/metrics-for-reinforcement-learning/> ,  
<https://aisera.com/blog/llm-evaluation/>

Evaluating of Works includes two scenario

### 1. Query Optimization using Machine Learning

- ❖ Dispersion across Time (DT)
- ❖ Short-term Risk across Time (SRT)
- ❖ Long-term Risk across Time (LRT)
- ❖ Dispersion across Runs (DR)
- ❖ Risk across Runs (RR)

- ❖ Dispersion across Fixed-Policy Rollouts (DF)
- ❖ Risk across Fixed-Policy Rollouts (RF)

## 2. Data Retrieval Using Large Language Model

- ❖ Reliability
- ❖ Efficiency
- ❖ Bias Detection
- ❖ User Trust
- ❖ Fine-Tuning

## **Contributions:**

All teammates will be contributing over all the parts of information retrieval system by learning and doing.