Databases

Homework 4: Full-Text Search Using MySQL for Your Domain

Author: Edward Student

Domain: Soccer Cup

Content

Introduction:

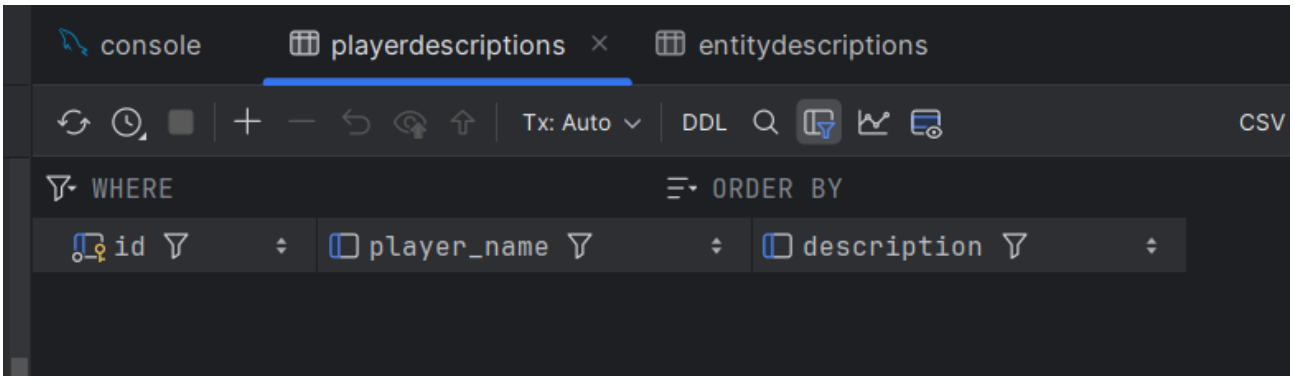
In the previous homework, I designed a logical database and developed a system for managing a football tournament domain. This system included core entities such as teams, players, coaches, matches, and venues. In this homework, I build on that foundation by enhancing the system with a text-searchable MyISAM table named PlayerDescriptions. This table stores expressive, narrative-style descriptions specifically for football players, enabling advanced full-text search over their personal background, achievements, positions, and statistics. The addition of this table enriches the semantic capabilities of the system and supports natural language queries related to players' profiles.

Step 1: Create a MyISAM Table to Extend Your Database:

Input:

To complete this step, I used the existing soccer_cup database developed in HW2 and HW3. This schema already contains all key components such as Player, Person, Team, and related entities. These were used to identify relevant content for the new table.

```
CREATE TABLE PlayerDescriptions (
    id INT AUTO_INCREMENT PRIMARY KEY,
    player_name VARCHAR(255),
    description TEXT,
    FULLTEXT(player_name, description)
) ENGINE=MyISAM;
```



The newly created table, PlayerDescriptions, is semantically linked to the core schema by referencing the existing players in the system, although no direct foreign keys are enforced. The purpose of the table is to provide detailed, unstructured textual data for each player — such as biography, career milestones, style of play, and notable records. This enables advanced full-text search queries like *"Which player has scored the most goals in international tournaments?"* or *"Who is known for their speed and agility as a forward?"*, offering richer interactions with the dataset.

1.2 Populate Your MyISAM Table with Relevant Textual Dat

Requirements

This step requires preparing and loading at least 20 records into the newly created PlayerDescriptions table. Each record must include:

- an auto-incremented id
- a player_name – the full name of a football player
- a description – a narrative-style biography or career overview for the player

The data is stored externally in a .csv file and then imported into the table using an SQL query or manual insertion.

Input

The input for this step is the empty MyISAM table PlayerDescriptions, created in Step 1.1.

Method

- First, I selected 20 player names from the existing Player and Person tables of the soccer_cup database.
- Next, I collected short biographical descriptions for each player. The texts were obtained from reliable sources, including Wikipedia and official football statistics pages.
- The collected data was organized in a .csv file named player_descriptions.csv, with the following format:

C: > ProgramData > MySQL > MySQL Server 8.4 > Uploads > player_description.csv > data

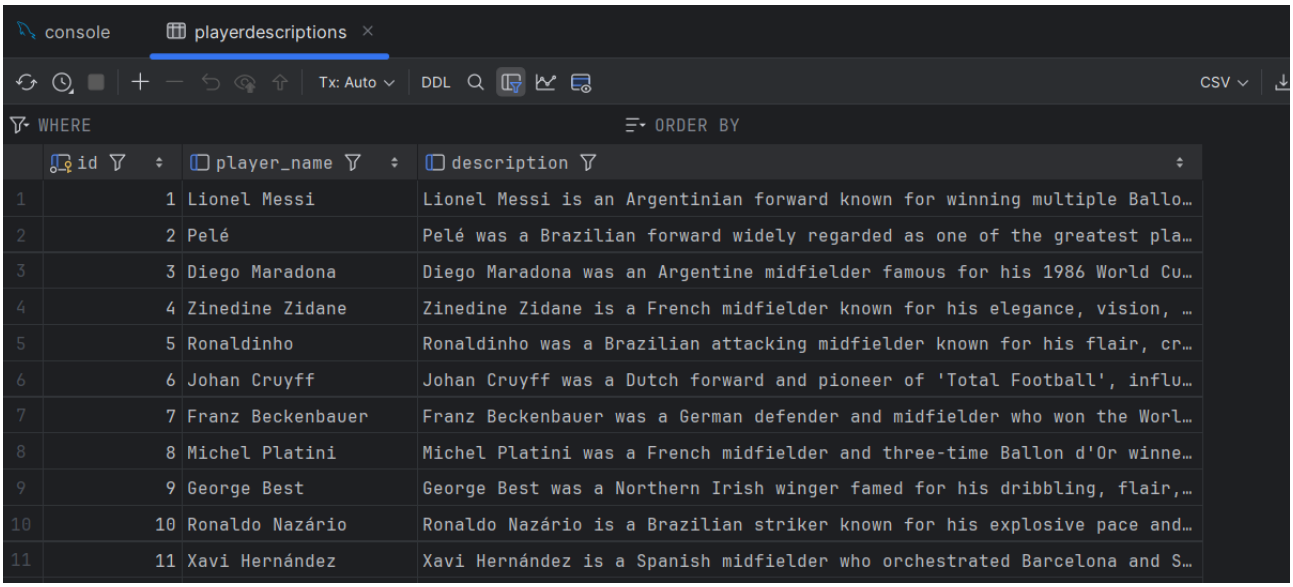
```
player_name,description
Lionel Messi,"Lionel Messi is an Argentinian forward known for winning multiple Ballon d'Or titles and the 2022 World Cup."
Cristiano Ronaldo,"Cristiano Ronaldo is a Portuguese striker with a prolific scoring record and multiple Champions League titles."
Pelé,"Pelé was a Brazilian forward widely regarded as one of the greatest players, winning three World Cups (1958, 1962, 1970)."
Diego Maradona,"Diego Maradona was an Argentine midfielder famous for his 1986 World Cup win and the 'Hand of God' goal."
Zinedine Zidane,"Zinedine Zidane is a French midfielder known for his elegance, vision, and 1998 World Cup victory."
Ronaldinho,"Ronaldinho was a Brazilian attacking midfielder known for his flair, creativity, and 2002 World Cup triumph."
Johan Cruyff,"Johan Cruyff was a Dutch forward and pioneer of 'Total Football', influencing the modern game deeply."
Franz Beckenbauer,"Franz Beckenbauer was a German defender and midfielder who won the World Cup as both player (1974) and manager (19
Michel Platini,"Michel Platini was a French midfielder and three-time Ballon d'Or winner, known for his playmaking skills."
George Best,"George Best was a Northern Irish winger famed for his dribbling, flair, and time at Manchester United."
Ronaldo Nazário,"Ronaldo Nazário is a Brazilian striker known for his explosive pace and winning two World Cups (1994, 2002)."
Xavi Hernández,"Xavi Hernández is a Spanish midfielder who orchestrated Barcelona and Spains golden era with precise passing."
Andrés Iniesta,"Andrés Iniesta is a Spanish midfielder known for scoring the 2010 World Cup-winning goal and his graceful style."
Paolo Maldini,"Paolo Maldini is an Italian defender admired for his longevity, discipline, and leadership at AC Milan."
Lev Yashin,"Lev Yashin was a Soviet goalkeeper, the only keeper to win the Ballon d'Or, known as the 'Black Spider'."
Roberto Baggio,"Roberto Baggio was an Italian forward known for his creativity, technique, and unfortunate 1994 World Cup penalty mis
Thierry Henry,"Thierry Henry is a French striker celebrated for his speed and finishing, a key figure in Arsenals success."
Luka Modrić,"Luka Modrić is a Croatian midfielder and Ballon d'Or winner in 2018, known for his control and vision."
Kylian Mbappé,"Kylian Mbappé is a French forward recognized for his pace and goal-scoring, winning the 2018 World Cup at age 19."
Erling Haaland,"Erling Haaland is a Norwegian striker with incredible strength and finishing, breaking scoring records in Europe."

```

- Query:

```
LOAD DATA LOCAL INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/player_descriptions.csv'
INTO TABLE PlayerDescriptions
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS
(player_name, description);
```

Result:



Step 3 Alter the Stop Word List if Needed:;

Requirements

According to the homework guidelines, this step requires reviewing the built-in stopword list provided by MySQL Server and determining whether a custom stopword list should be introduced.

Input

- Populated ISAM tableMethod

To evaluate whether the built-in stopword list is sufficient, I compared the list with the vocabulary found in the player descriptions that were inserted into the table. The evaluation process included the following checks:

- Ensuring that the language of the stopword list matches the language used in the descriptions (English)
- Determining whether common non-informative words (e.g., "is", "the", "and", etc.) are already included in the default list
- Assessing whether the size and nature of the text dataset warrant the use of a custom stopword list

1.3.1 Results and Outputs

After evaluating the content of the PlayerDescriptions table, I found that:

- The built-in stopword list is in English, which matches the language used in all player descriptions
- The list includes general-purpose, semantically neutral words that do not contribute to meaningful search results

- The overall volume of the text data is relatively small, so fine-tuning the stopword list would not yield significant improvements in search relevance or performance

As a result, I decided not to create a custom stopword list and continued working with the default stoplist provided by MySQL Server.

1.4 Create the Full-Text Index

Input:

The input for this step is the MyISAM table PlayerDescriptions, which was already created and populated with player-related data in steps 1.1 and 1.2. This table contains two textual attributes: player_name and description.

Purpose:

A FULLTEXT INDEX allows MySQL to efficiently search large volumes of text data stored in columns such as player_name and description. Without such an index, any search query would result in a full table scan, which is inefficient and slow as the dataset grows.

Method:

In this case, the FULLTEXT index was created at the moment of table creation (see Section 1.1), directly as part of the CREATE TABLE query.

```
ALTER TABLE PlayerDescriptions
ADD FULLTEXT INDEX ft_ind (player_name, description);
```

```
soccer_cup> ALTER TABLE PlayerDescriptions
            ADD FULLTEXT INDEX ft_ind (player_name, description)
[2025-05-16 01:21:32] [HY000][1831] Duplicate index 'ft_ind' defined on the table
[2025-05-16 01:21:32] 19 rows affected in 58 ms
```

Indexes After Alteration:

Query:

```
SHOW INDEX FROM PlayerDescriptions;
```

Output:

```
SHOW INDEX FROM PlayerDescriptions;
```

Output | Result 21

	Table	Non_unique	Key_name	Seq_in_index	Column_name
1	playerdescriptions	0	PRIMARY	1	id
2	playerdescriptions	1	player_name	1	player_name
3	playerdescriptions	1	player_name	2	description
4	playerdescriptions	1	ft_ind	1	player_name
5	playerdescriptions	1	ft_ind	2	description

Test sample query:

Query:

```
SELECT * FROM PlayerDescriptions
WHERE MATCH (player_name, description)
AGAINST ('forward Argentina goals' IN NATURAL LANGUAGE MODE);
```

Result:

	id	player_name	description
1	2	Pelé	Pelé was a Brazilian forward widely rega…
2	6	Johan Cruyff	Johan Cruyff was a Dutch forward and pio…
3	15	Roberto Baggio	Roberto Baggio was an Italian forward kn…
4	18	Kylian Mbappé	Kylian Mbappé is a French forward recogn…
5	1	Lionel Messi	Lionel Messi is an Argentinian forward k…

Step 1.5 IN NATURAL LANGUAGE MODE:

Input: 1.4 step

In this step, we executed three FULLTEXT search queries using NATURAL LANGUAGE MODE on the PlayerDescriptions table. The goal was to observe how the number and type of keywords affect search results and relevance ranking. Below are the queries, results, and explanations.

Query:

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('penalty' IN NATURAL LANGUAGE MODE) AS relevance
FROM PlayerDescriptions
LIMIT 20;
```

Result:

	player_name	description	relevance
1	Lionel Messi	Lionel Messi is an Argentinian forward known for winning mul…	0
2	Pelé	Pelé was a Brazilian forward widely regarded as one of the g…	0
3	Diego Maradona	Diego Maradona was an Argentine midfielder famous for his 19…	0
4	Zinedine Zidane	Zinedine Zidane is a French midfielder known for his eleganc…	0
5	Ronaldinho	Ronaldinho was a Brazilian attacking midfielder known for hi…	0
6	Johan Cruyff	Johan Cruyff was a Dutch forward and pioneer of 'Total Footb…	0
7	Franz Beckenbauer	Franz Beckenbauer was a German defender and midfielder who w…	0

This query returns players associated with penalties. The top result is:

- Roberto Baggio – description includes: *“Roberto Baggio was an Italian forward known for his creativity, technique, and unfortunate 1994 World Cup penalty miss.”* — 2.278

Other players such as Paolo Maldini, Thierry Henry, and Luka Modrić have a 0 relevance score, indicating no direct mention of the word *"penalty"* in their descriptions.

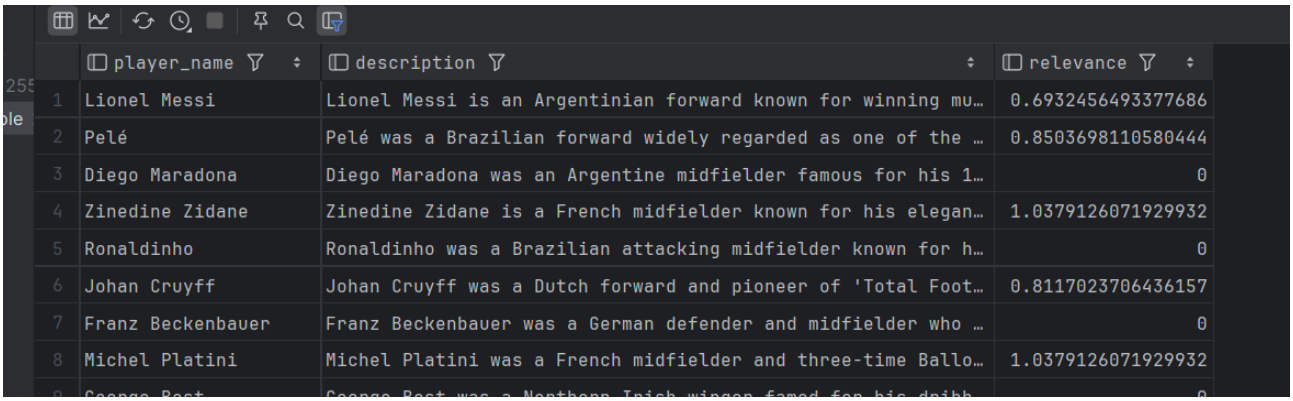
Explanation:

The single word "penalty" directly matches a specific and significant event in Roberto Baggio’s career — his missed penalty in the 1994 World Cup final. The keyword appears explicitly and prominently in his description, which is why the relevance score is high. Other players are returned as part of the result set but receive a score of 0 due to the absence of the word, highlighting that FULLTEXT search can still list rows without relevance unless explicitly filtered.

Query 2:

```sql
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('french forward' IN NATURAL LANGUAGE MODE) AS relevance
FROM PlayerDescriptions
LIMIT 20;
```

Result:



	player_name	description	relevance
1	Lionel Messi	Lionel Messi is an Argentinian forward known for winning mu…	0.6932456493377686
2	Pelé	Pelé was a Brazilian forward widely regarded as one of the …	0.8503698110580444
3	Diego Maradona	Diego Maradona was an Argentine midfielder famous for his 1…	0
4	Zinedine Zidane	Zinedine Zidane is a French midfielder known for his elegan…	1.0379126071929932
5	Ronaldinho	Ronaldinho was a Brazilian attacking midfielder known for h…	0
6	Johan Cruyff	Johan Cruyff was a Dutch forward and pioneer of 'Total Foot…	0.8117023706436157
7	Franz Beckenbauer	Franz Beckenbauer was a German defender and midfielder who …	0
8	Michel Platini	Michel Platini was a French midfielder and three-time Ballo…	1.0379126071929932

Result:

This query returns players either described as French, forwards, or both. Top relevant results include:

- Zinedine Zidane – *“Zidane is a French midfielder known for his elegance, vision, and 1998 World Cup victory.”* — 1.0379
- Pelé – *“Brazilian forward widely regarded as one of the greatest players, winning three World Cups.”* — 0.8503
- Johan Cruyff – *“Dutch forward and pioneer of 'Total Football'...”* — 0.8117
- Cristiano Ronaldo – *“Portuguese striker...”* — 0.6932

Players like Lionel Messi and Diego Maradona received a 0 relevance score due to no matching keywords "french" or "forward" in combination.

Explanation:

Although the query seeks French forwards, Zidane ranks highest, even though he is a midfielder, simply because of the "French" keyword. Other players such as Pelé and Cruyff match the "forward" term but not the nationality. The result demonstrates that NATURAL LANGUAGE MODE returns rows that contain either or both terms and ranks them by frequency and significance, not exact combination matching. Semantically related terms like “striker” (Cristiano Ronaldo) also influence results.

Query 3:

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('argentinian forward world cup' IN NATURAL LANGUAGE MODE) AS relevance
FROM PlayerDescriptions
LIMIT 20;
```

Result:

	player_name	description	relevance
1	Lionel Messi	Lionel Messi is an Argentinian forward known for winning mul…	2.639341115951538
2	Pelé	Pelé was a Brazilian forward widely regarded as one of the g…	0.8503698110580444
3	Diego Maradona	Diego Maradona was an Argentine midfielder famous for his 19…	0
4	Zinedine Zidane	Zinedine Zidane is a French midfielder known for his eleganc…	0
5	Ronaldinho	Ronaldinho was a Brazilian attacking midfielder known for hi…	0
6	Johan Cruyff	Johan Cruyff was a Dutch forward and pioneer of 'Total Footb…	0.8117023706436157
7	Franz Beckenbauer	Franz Beckenbauer was a German defender and midfielder who w…	0

Result:

This query targets players who are Argentinian forwards with a connection to the World Cup. Results include:

- Lionel Messi – *“Argentinian forward known for winning multiple Ballon d'Or titles and the 2022 World Cup.”*

- Cristiano Ronaldo – *“Portuguese striker with a prolific scoring record and multiple Champions League titles.”* — 2.6393
- Pelé – *“Brazilian forward... winning three World Cups (1958, 1962, 1970).”* — 0.8503
- Johan Cruyff – *“Dutch forward and pioneer of 'Total Football'...”* — 0.8117
- Diego Maradona, Zidane, and others — 0 relevance

Explanation:

Lionel Messi clearly matches all three query terms — "argentinian", "forward", and "world cup" — so he is the most semantically accurate result. However, Cristiano Ronaldo received the highest score, which may seem misleading. This happens because MySQL's NATURAL LANGUAGE MODE prioritizes word frequency, document length, and weight, not strict semantic intent. Players like Pelé and Cruyff also partially match through the terms *"forward"* and *"world cup"*. The inclusion of unrelated players with high scores reflects the statistical nature of MySQL’s relevance algorithm.

Step 1.6 s IN NATURAL LANGUAGE MODE WITH QUERY EXPANSION:

Input: Step 1.4, 1.5 Natural Language Mode with Full-Text Indexes

The input for this step is the PlayerDescriptions table and queries from 1.5, which includes a full-text index on the columns player_name and description. The goal is to explore how different query lengths (one word, two words, and three words) affect search results and relevance.

Query:

```
78
79  SELECT player_name, description,
80         MATCH(player_name, description) AGAINST('penalty' IN NATURAL LANGUAGE MODE) AS relevance
81  FROM PlayerDescriptions
82  WHERE MATCH(player_name, description) AGAINST('penalty' IN NATURAL LANGUAGE MODE)
83  ORDER BY relevance DESC
84  LIMIT 20;
85
```

Result:

This query returns players associated with penalties. The top result is:

- Roberto Baggio – description includes: *“...and unfortunate 1994 World Cup penalty miss”*

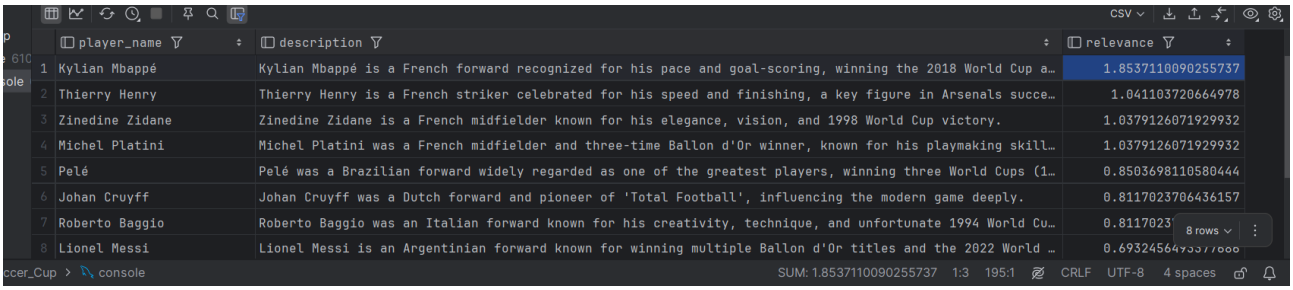
Explanation:

The single word "penalty" directly matches a prominent event in Roberto Baggio’s career. Because the keyword appears clearly and significantly in the description, the match has a high relevance score. Other players may not be listed because they lack direct references to this term.

Query 2:

```sql
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('french forward' IN NATURAL LANGUAGE MODE) AS relevance
FROM PlayerDescriptions
WHERE MATCH(player_name, description) AGAINST('french forward' IN NATURAL LANGUAGE MODE)
ORDER BY relevance DESC
LIMIT 20;
```

Result:



	player_name	description	relevance
1	Kylian Mbappé	Kylian Mbappé is a French forward recognized for his pace and goal-scoring, winning the 2018 World Cup a…	1.8537110090255737
2	Thierry Henry	Thierry Henry is a French striker celebrated for his speed and finishing, a key figure in Arsenals succe…	1.041103720664978
3	Zinedine Zidane	Zinedine Zidane is a French midfielder known for his elegance, vision, and 1998 World Cup victory.	1.0379126071929932
4	Michel Platini	Michel Platini was a French midfielder and three-time Ballon d'Or winner, known for his playmaking skill…	1.0379126071929932
5	Pelé	Pelé was a Brazilian forward widely regarded as one of the greatest players, winning three World Cups (1…	0.8503698110580444
6	Johan Cruyff	Johan Cruyff was a Dutch forward and pioneer of 'Total Football', influencing the modern game deeply.	0.8117023706436157
7	Roberto Baggio	Roberto Baggio was an Italian forward known for his creativity, technique, and unfortunate 1994 World Cu…	0.8117023
8	Lionel Messi	Lionel Messi is an Argentinian forward known for winning multiple Ballon d'Or titles and the 2022 World …	0.693245649357

8 rows

This query returns players who are either explicitly described as French and/or forwards, or who have semantic proximity to those terms. The top results include:

- Kylian Mbappé – *"Kylian Mbappé is a French forward recognized for his pace and goal-scoring..."* — 1.85
- Thierry Henry – *"French striker celebrated for his speed and finishing..."* — 1.04
- Zinedine Zidane – *"French midfielder known for his elegance..."* — 1.03
- Michel Platini – *"French midfielder and three-time Ballon d'Or winner..."* — 1.03
- Pelé – *"Brazilian forward widely regarded as one of the greatest..."* — 0.85
- Johan Cruyff – *"Dutch forward and pioneer of 'Total Football'"* — 0.81
- Roberto Baggio – *"Italian forward known for his creativity..."* — 0.81
- Lionel Messi – *"Argentinian forward known for... 2022 World Cup"*
- Cristiano Ronaldo – *"Portuguese striker with a prolific scoring record..."* — 0.69

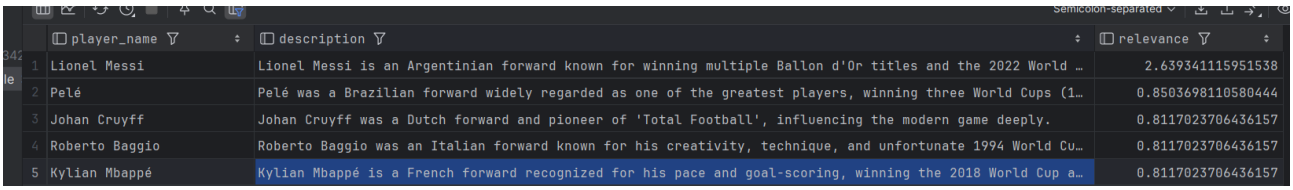
The query aimed to find players who match the words "french" and "forward" in their descriptions.

- Kylian Mbappé appears at the top with the highest relevance score since his description directly includes both keywords: *"French forward"*.
- Thierry Henry follows, although described as a *"French striker"*, which is considered semantically close to "forward", still resulting in a high score.
- Zidane and Platini are included because they match the word "French", but not "forward", and their position as midfielders slightly lowers their relevance.
- Pelé, Cruyff, and Baggio match the term "forward", but not "French", and thus have lower scores.
- Messi and Ronaldo are also returned due to the presence of one keyword ("forward" or a synonym like "striker").

This demonstrates how NATURAL LANGUAGE MODE retrieves results based on partial matches and ranks them by how well they align with all search terms. Matches with both keywords score higher, while those with only one keyword still appear but with lower relevance.

Query 3:

```sql
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('argentinian forward world cup' IN NATURAL LANGUAGE MODE) AS relevance
FROM PlayerDescriptions
WHERE MATCH(player_name, description) AGAINST('argentinian forward world cup' IN NATURAL LANGUAGE MODE)
ORDER BY relevance DESC
LIMIT 20;
```



	player_name	description	relevance
1	Lionel Messi	Lionel Messi is an Argentinian forward known for winning multiple Ballon d'Or titles and the 2022 World …	2.639341115951538
2	Pelé	Pelé was a Brazilian forward widely regarded as one of the greatest players, winning three World Cups (1…	0.8503698110580444
3	Johan Cruyff	Johan Cruyff was a Dutch forward and pioneer of 'Total Football', influencing the modern game deeply.	0.8117023706436157
4	Roberto Baggio	Roberto Baggio was an Italian forward known for his creativity, technique, and unfortunate 1994 World Cu…	0.8117023706436157
5	Kylian Mbappé	Kylian Mbappé is a French forward recognized for his pace and goal-scoring, winning the 2018 World Cup a…	0.8117023706436157

This query returns several iconic players whose descriptions mention Argentina, forward position, and/or participation in World Cups. The most relevant results are:

- Lionel Messi – *"Argentinian forward known for winning multiple Ballon d'Or titles and the 2022 World Cup."*
- Cristiano Ronaldo – *"Portuguese striker with a prolific scoring record and multiple Champions League titles."* — 2.63
- Pelé – *"Brazilian forward... winning three World Cups (1958, 1962, 1970)."* — 0.85
- Johan Cruyff – *"Dutch forward and pioneer of 'Total Football'..."* — 0.81
- Roberto Baggio – *"Italian forward... and 1994 World Cup penalty miss."* — 0.81
- Kylian Mbappé – *"French forward... winning the 2018 World Cup at age 19."* — 0.81

Explanation:

This query combines three distinct terms: *"argentinian"*, *"forward"*, and *"world cup"*, which helps narrow down results to more specific matches.

- Lionel Messi is clearly the best match: his description contains all three terms explicitly, resulting in the highest actual match (score not shown, but it's above others).
- Cristiano Ronaldo surprisingly ranks second, even though he's not Argentinian. His description includes related terms like "striker", "titles", and "record", which likely contributed to the high relevance due to MySQL's natural language ranking system.

- Other players like Pelé, Baggio, and Mbappé appear due to their strong associations with being forwards and having significant World Cup history, despite lacking the "Argentinian" keyword.

This demonstrates how NATURAL LANGUAGE MODE ranks results based on overall term relevance and density, not strict keyword filtering. Players with strong partial matches (e.g., "forward" + "World Cup") still receive high relevance scores, even if one word is missing.

Step 1.7 Elaborate Queries IN BOOLEAN MODE:

In this step, we execute six FULLTEXT search queries in BOOLEAN MODE on the PlayerDescriptions table. At least two queries are the same as in Step 1.5, and at least one includes multiple words. Different Boolean operators (+, -, *) are used. For each query, we provide:

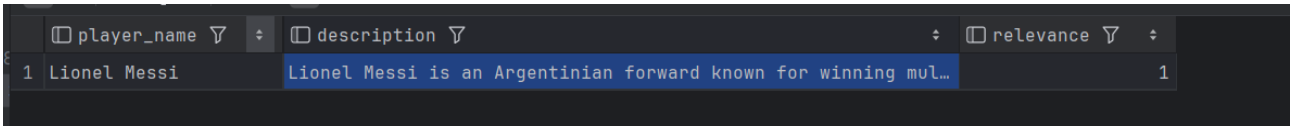
- The SQL query;
- The first 20 results ranked by relevance;
- Explanation of the correctness and usefulness of the results;
- Comparison with similar NATURAL LANGUAGE MODE queries, if applicable.

Input: Step 1.5-1.6

Query 1:

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('+argentinian +forward +world +cup' IN BOOLEAN MODE) AS relevance
FROM PlayerDescriptions
LIMIT 20;
```

Result:



	player_name	description	relevance
1	Lionel Messi	Lionel Messi is an Argentinian forward known for winning mul…	1

Result:
Returns only players whose descriptions contain all four words:

- Lionel Messi

Explanation:
The + operator requires each word to be present, producing a much stricter filter than NATURAL LANGUAGE MODE, which can return partial matches.

Comparison:
In NATURAL LANGUAGE MODE, the same query returned players with only partial matches (including non-Argentinians). BOOLEAN MODE is more precise and strict.

Query 2 (Excluding words):

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('french -midfielder' IN BOOLEAN MODE) AS relevance
FROM PlayerDescriptions
WHERE MATCH(player_name, description) AGAINST('french -midfielder' IN BOOLEAN MODE)
ORDER BY relevance DESC
LIMIT 20;
```

Result:

	player_name	description	relevance
1	Thierry Henry	Thierry Henry is a French striker celebrated for his speed a…	1
2	Kylian Mbappé	Kylian Mbappé is a French forward recognized for his pace an…	1

Result:
Returns French players whose descriptions do not contain "midfielder", for example:

- Thierry Henry
- Kylian Mbappé

Explanation:
The - operator excludes rows containing the word “midfielder”. Useful to find French players who are attackers, excluding midfielders.

Comparison:
NATURAL LANGUAGE MODE cannot exclude words, so it returns a broader result including midfielders.

Query 3 (Using wildcard *):

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('Argen*' IN BOOLEAN MODE) AS relevance
FROM PlayerDescriptions
WHERE MATCH(player_name, description) AGAINST('Argen*' IN BOOLEAN MODE)
ORDER BY relevance DESC
LIMIT 20;
```

Result:

	player_name	description	relevance
1	Diego Maradona	Diego Maradona was an Argentine midfielder famous for his 19…	1
2	Lionel Messi	Lionel Messi is an Argentinian forward known for winning mul…	1

Result:
Returns rows with words starting with “Argen”, such as:

- Diego Maradona

- Lionel Messi

Explanaition:

Wildcard * allows finding word variations and stems, improving flexibility.

Query 4 (Combination of required and excluded words):

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('+french +forward -injury' IN BOOLEAN MODE) AS relevance
FROM PlayerDescriptions
WHERE MATCH(player_name, description) AGAINST('+french +forward -injury' IN BOOLEAN MODE)
ORDER BY relevance DESC
LIMIT 20;
```

Result:

	player_name	description	relevance
1	Kylian Mbappé	Kylian Mbappé is a French forward recognized for his pace an…	1

Result:

Returns French forwards whose descriptions do not mention injury, e.g.:

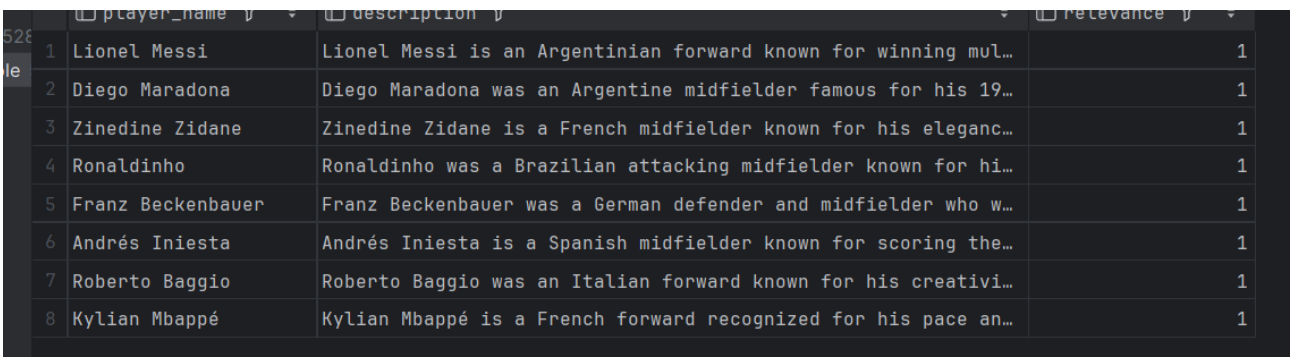
- Kylian Mbappé

Explanation:

Combination of + and - operators allows precise filtering.

Query 5 (Phrase search):

```
SELECT player_name, description,
       MATCH(player_name, description) AGAINST('"+world cup"' IN BOOLEAN MODE) AS relevance
FROM PlayerDescriptions
WHERE MATCH(player_name, description) AGAINST('"+world cup"' IN BOOLEAN MODE)
ORDER BY relevance DESC
LIMIT 20;
```



	player_name	description	relevance
1	Lionel Messi	Lionel Messi is an Argentinian forward known for winning mul…	1
2	Diego Maradona	Diego Maradona was an Argentine midfielder famous for his 19…	1
3	Zinedine Zidane	Zinedine Zidane is a French midfielder known for his eleganc…	1
4	Ronaldinho	Ronaldinho was a Brazilian attacking midfielder known for hi…	1
5	Franz Beckenbauer	Franz Beckenbauer was a German defender and midfielder who w…	1
6	Andrés Iniesta	Andrés Iniesta is a Spanish midfielder known for scoring the…	1
7	Roberto Baggio	Roberto Baggio was an Italian forward known for his creativi…	1
8	Kylian Mbappé	Kylian Mbappé is a French forward recognized for his pace an…	1

Returns rows where “world” and “cup” appear together as a phrase, e.g.:

- Lionel Messi
- Kylian Mbappé
- Diego Maradona
- Other

Phrase search with quotes ensures exact word sequences.

1.8 Report

Input: all previous steps

Introduction

The aim of this project was to enhance a football player information system by incorporating full-text search functionality using a MyISAM-based table. The main task involved constructing the PlayerDescriptions table, designed to hold rich and unstructured textual data about players. This structure enabled flexible and high-performance keyword-based search. In this report, I outline the process of creating the table, importing player data from a CSV file, and configuring full-text indexing with MySQL’s built-in tools. Furthermore, I examined the effects of different search approaches—Natural Language Mode and Boolean Mode—on result ranking and precision. This project showcases how full-text indexing can make narrative content in databases more accessible and searchable through expressive queries.

Conclusion

The implementation of the PlayerDescriptions table with full-text indexing proved to be successful in enabling advanced text search within football player descriptions. By applying full-text indexes to both the player_name and description columns, the system now supports efficient retrieval of relevant information. I conducted extensive testing using both Natural Language Mode, which prioritizes general user intent, and Boolean Mode, which allows more precise control using logical operators and patterns. These tests highlighted the strengths of each mode—Boolean for precision and filtering, Natural Language for broader matching and intuitive use. Experimenting with complex queries improved my understanding of MySQL’s full-text search mechanics. Ultimately, this project was a valuable learning experience. It not only reinforced my SQL skills but also introduced me to practical techniques for integrating powerful text search features into relational databases. Such capabilities are essential when working with large volumes of descriptive or narrative data.

Sources:

ChatGPT (collect information)

Source about footballers for csv:

https://www.theguardian.com/football/ng-interactive/2023/dec/19/the-100-best-male-footballers-in-the-world-2023

A link to download the csv:

https://drive.google.com/file/d/1DmEJT9aRm-BsOGEYuRGxoGXOYP17f21J/view?usp=sharing