**NLQ to SQL**

A system that takes in a natural language question and a database, and then generates an SQL query to retrieve data from the database based on the question. The first model, the Sentence Transformer, is used to identify the table name in the database that is most relevant to the question. The second model, the Text-to-Text Transfer Transformer (T5), is then used to generate an SQL query based on the identified table and the natural language question. Finally, the generated SQL query is executed on the database to retrieve the data.

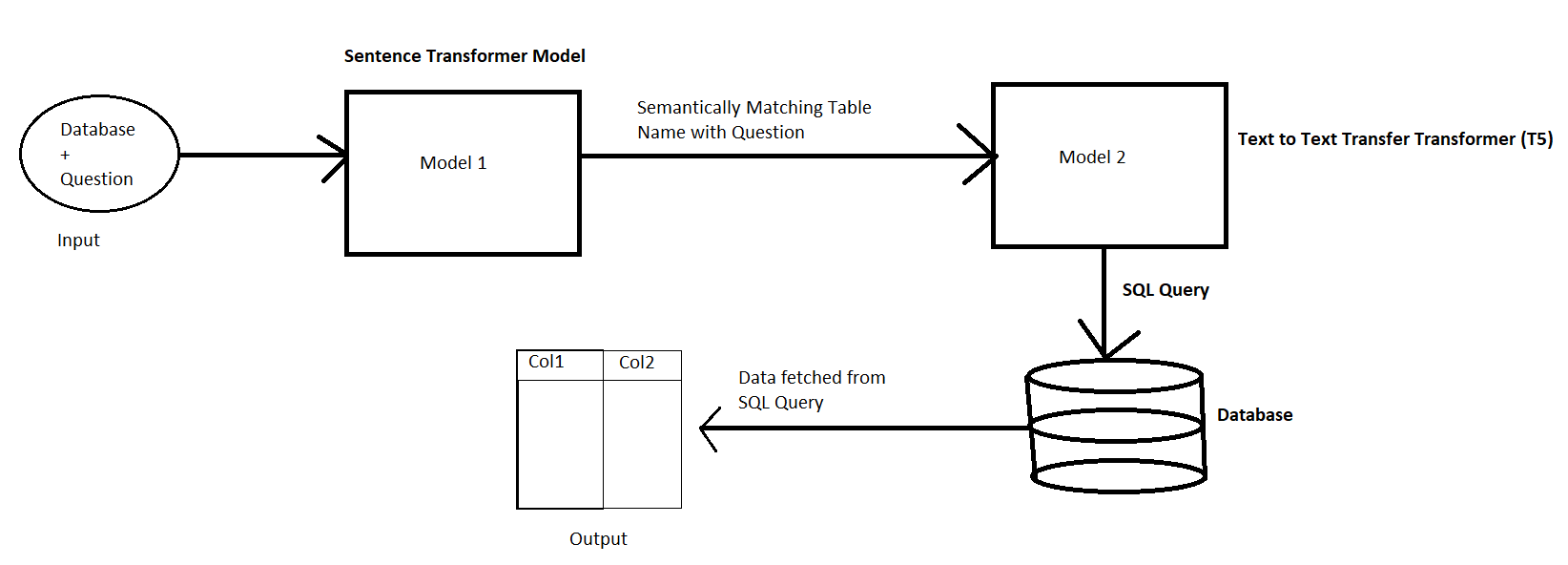


Fig 1: Architecture

Input: Database and a Natural Language Question (English)

Model 1: Sentence Transformer Model

Output from Model 1: Semantically Matching Table Name with Question

Model 2: Text-to-Text Transfer Transformer(T5)

Output from Model 2: SQL Query

Database: Execute SQL Query on given Database and get Data

**Model 1: Sentence Transformer Model**

Trained on 3 data sets

1. Trivia\_qa
2. Snli
3. Multi\_nli

**Model 2: Text-to-Text Transfer Transformer(T5)** (1million Parameters)

Trained on 2 data sets

1. CoSQL SQL-grounded dialogue state tracking dataset
2. Spider text-to-SQL dataset

**Description**

1. The user inputs a natural language question and a database into the system.
2. The Sentence Transformer model takes in the natural language question and processes it to identify the table name in the database that is most relevant to the question. For example, if the question is "What is the population of New York?", the Sentence Transformer model might identify the "cities" table as the relevant table.
3. The Text-to-Text Transfer Transformer (T5) model takes in the natural language question and the identified table name and generates an SQL query to retrieve data from the database based on the question. For example, given the question "What is the population of New York?" and the identified "cities" table, the T5 model might generate the following SQL query: "SELECT population FROM cities WHERE city='New York';"
4. The generated SQL query is then executed on the database, and the data is retrieved and returned to the user.

**Examples:**

Enter question: What is the total sales for each customer?

select sum(Sales), "Customer Name" from train group by "Customer Name"

Enter question: How many orders were placed in each city?

select "City", count(\*) from train group by "City"

Enter question: What is the total sales for each product?

select sum(Sales), "Product Name" from train group by "Product Name"

Enter question: How many orders were placed in each region?

select "Region", count(\*) from train group by "Region"

Enter question: How many orders were placed on each date?

select "Order Date", count(\*) from train group by "Order Date"

Enter question: What is the average sales per order for each category?

select "Category", avg(Sales) from train group by "Category"

Enter question: What are the top 3 cities with the highest total sales?

select "City" from train group by "City" order by sum("sales") desc limit 3

Enter question: What is the average sales value for each shipping mode?

select avg(Sales), "Ship Mode" from train group by "Ship Mode"

Enter question: What are the top 5 products with the highest total sales?

select "Product Name" from train group by "Product Name" order by sum("sales") desc limit 5

Enter question: What is the average sales value for each customer segment?

select "Segment", avg(Sales) from train group by "Segment"

Enter question: For each city, list the total sales generated, broken down by postal code.

select "City", sum("Sales") from train group by "City"

Enter question: In which city were the most orders shipped to?

select "City" from train group by "City" order by count(\*) desc limit 1

Enter question: Which customer segment had the highest average sales value?

select "Segment" from train group by "Segment" order by avg(Sales) desc limit 1

Enter question: What is the most common Ship Mode used?

select "ship mode" from train group by "ship mode" order by count ( \* ) desc limit 1

Enter question: How many unique Customer IDs are there?

select count ( distinct "customer id" ) from train

Enter question: What is the average Sales amount for orders shipped to the region "East"?

select avg(Sales) from train where "region" = "East"