

Text-to-SQL: Leveraging LLMs for Query Translation

Final Project Presentation – Group 42

Harsh Sanas, Harshal Vaidya, Siddharth Khachane, Sai Sudha Panigrahi,
Vatsal Vatsyayan

CSCI 544: Applied Natural Language Processing
University of Southern California



Motivation

- To enable non-technical users to interact with databases effortlessly using natural language, by leveraging LLMs to automate accurate SQL generation.

Problem Statement

- Convert natural language questions into accurate and executable SQL queries
- Ensure generated SQL aligns with both structure and semantics of underlying database schemas
- Design a generalized, schema-aware prompt format usable across different databases
- Build a lightweight and modular system for real-world deployment

Results – WikiSQL

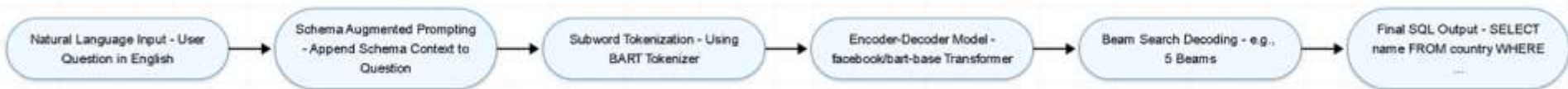
Model	Greedy Accuracy (%)	Beam Accuracy (%)
T5-small	51.5	51.6
T5-base	64.29	64.58
BART-base	66.06	66.06
CodeT5-small	66.95	67.31
CodeT5-base	69.86	70.16

- Trained T5-small/base, BART-base, and CodeT5-small/base on **WikiSQL**
- Compared **Greedy** vs **Beam** Decoding (Beam search gave slight improvements in accuracy)
- Best result: **70.16%** with **CodeT5-base + beam=4**

Results – Spider

- Model: BART-base
- Accuracy: 82%
- Chosen for production integration
- Balanced syntax + semantics + real DB reliability

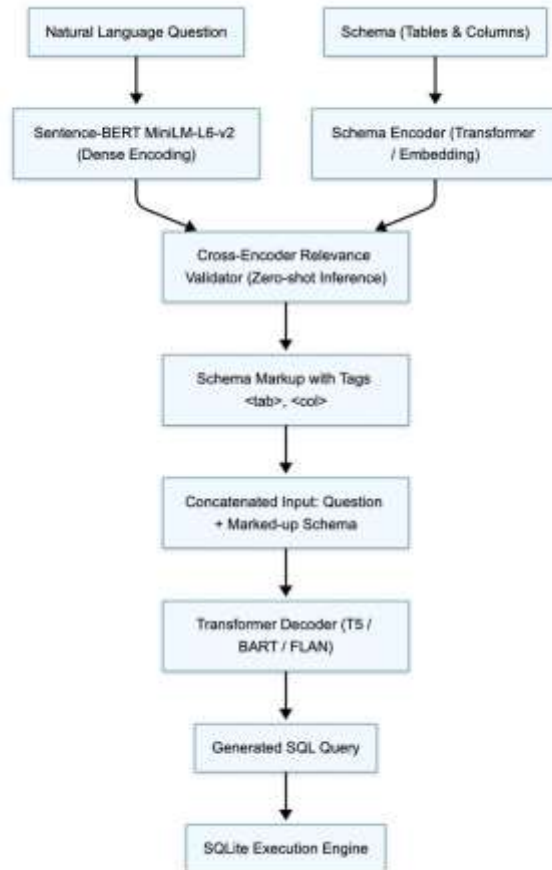
Model Architecture



Future Objectives

- Apply validation-based reward tuning to reduce hallucinated columns. (Reinforcement Learning).
- Incorporate constrained decoding for table-column correctness.
- Improve beam search configurations using cross-validation.

Pipeline



DEMO

Group 42 Text to SQL

What are the title and issues of the books?

Submit

Database: book_2

This schema is relevant to your query.

Schema Information

Table: publication

Column Name	Data Type
Publication_ID	number
Book_ID	number
Publisher	text
Publication_Date	text
Price	number

Table: book

Column Name	Data Type
Book_ID	number
Title	text
Issues	number
Writer	text

Foreign Keys

publication.Book_ID → book.Book_ID

Generate SQL

Generated SQL

```
select title, issues from book;
```

Execute Query

Query Results

Found 10 results

Title	Issues
The Black Lamb	6
Bloody Mary	4
Bloody Mary : Lady Liberty	4
BrainBank	6
Cyberella	12
Dead Corps	4
The Dome: Ground Zero	1
Gemini Blood	9
Michael Moorcock's Multiverse	12
Sheva's War	5

Thank You !