

# Text-to-SQL Semantic Parsing

Group 4  
Nguyen Tan Viet  
Ngo Quoc Bao

February 28, 2022

## Definition of problem

Given a natural language  $Q$  and the schema  $S = \langle \mathcal{T}, \mathcal{C} \rangle$  for a relational database, the parser needs to generate the corresponding SQL query  $Y$ . The schema consists of tables  $\mathcal{T} = \{t_1, \dots, t_N\}$  and fields  $\mathcal{C} = \{c_{11}, \dots, c_{1|\mathcal{T}_1|}, \dots, c_{n1}, \dots, c_{N|\mathcal{T}_N|}\}$ . Each table  $t_i$  and each field  $c_{ij}$  has a textual name. Some fields are primary keys, used for uniquely indexing each data record, and some are foreign keys, used to reference a primary key in a different table. In addition, each field has a data type,  $\tau \in \{\text{number}, \text{text}, \text{time}, \text{boolean}, \text{etc}\}$

# Literal review

# New point in research

