A Comprehensive Exploration on WikiSQL with Table-Aware Word Contextualization

Wonseok Hwang **Jinyeong Yim** **Seunghyun Park** **Minjoon Seo**
Clova AI, NAVER Corp.
{wonseok.hwang, jinyeong.yim, seung.park, minjoon.seo}@navercorp.com

Abstract

We present SQLOVA, the first Natural-language-to-SQL (NL2SQL) model to achieve human performance in WikiSQL dataset. We revisit and discuss diverse popular methods in NL2SQL literature, take a full advantage of BERT (Devlin et al., 2018) through an effective table contextualization method, and coherently combine them, outperforming the previous state of the art by 8.2% and 2.5% in logical form and execution accuracy, respectively. We particularly note that BERT with a seq2seq decoder leads to a poor performance in the task, indicating the importance of a careful design when using such large pretrained models. We also provide a comprehensive analysis on the dataset and our model, which can be helpful for designing future NL2SQL datsets and models. We especially show that our model's performance is near the upper bound in WikiSQL, where we observe that a large portion of the evaluation errors are due to wrong annotations, and our model is already exceeding human performance by 1.3% in execution accuracy.

1 Introduction

NL2SQL is a popular form of semantic parsing tasks that asks for translating a natural language (NL) utterance to a machine-executable SQL query. As one of the first large-scale (80k) human-verified semantic parsing datasets, WikiSQL (Zhong et al., 2017) has attracted much attention in the community and enabled a significant progress through task-specific end-to-end neural models (Xu et al., 2017). On the other side of the NLP community, we have also observed a rapid advancement in contextualized word representations (Peters et al., 2018; Devlin et al., 2018), which have proved to be extremely effective for most language tasks that deal with unstructured text data. However, it has not been clear yet whether the word contextualization is also similarly effective when structured data such as tables in WikiSQL are involved.

In this paper, we discuss our approach on WikiSQL that coherently brings previous NL2SQL liteature and large pretrained models together. Our model, SQLOVA, consists of two layers, encoding layer that obtains table-aware word contextualization and NL2SQL layer that generates the SQL query from the contextualized representations. We show that SQLOVA outperforms the previous best model achieving 83.6% logical form accuracy and 89.6% execution accuracy on WikiSQL test set, outperforming the previous best model by 8.2% and 2.5%, respectively. It is important to note that, while BERT plays a significant role, merely attaching a seq2seq model on the top of BERT leads to a poor performance, indicating the importance of properly and carefully utilizing BERT when dealing with structured data.

We furthermore argue that these scores are near the upper bound in WikiSQL, where we observe that most of the evaluation errors are caused by either wrong annotations by humans or the lack of given information. In fact, according to our crowdsourced statistics on an approximately 10% sampled set of WikiSQL dataset, our model's score exceeds human performance at least by 1.3% in execution accuracy.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

Table:

Player	Country	Points	Winnings ($)
Steve Stricker	United States	9000	1260000
K.J. Choi	South Korea	**5400**	756000
Rory Sabbatini	South Africa	3400	4760000
Mark Calcavecchia	United States	2067	289333
Ernie Els	South Africa	2067	289333

Question: What is the points of South Korea player?
SQL: SELECT Points WHERE Country = South Korea
Answer: 5400

Figure 1: Example of WikiSQL semantic parsing task. For given questions and table headers, the model generates corresponding SQL query and retrieves the answer from the table.

In short, our key contributions are:

- We propose a carefully designed architecture that brings the best of previous NL2SQL approaches and large pretrained language models together. Our model clearly outperforms the previous best model and the human performance in WikiSQL.
- We provide a diverse and detailed analysis on the dataset and our model. These examinations will further help future research on NL2SQL data creation and model development.

The rest of the paper is organized as follows. We first describe our model in Section 3. Then we report the quantitative results of our model in comparison to previous baselines in Section 4. Lastly, we discuss qualitative analysis on both the dataset and our model in Section 5. [1]

2 Related Work

WikiSQL is a large semantic parsing dataset consisting of 80,654 natural language utterances and corresponding SQL annotations on 24,241 tables extracted from Wikipedia (Zhong et al., 2017). The task is to build the model that generates SQL query for given natural language question on single table and table headers without using contents of the table. Some examples, using the table from WikiSQL, are shown in Figure 1.

The large size of the dataset has enabled adopting deep neural techniques for the task and drew much attention in the community recently. Although early studies on neural semantic parsers have started without syntax specific constraints on output space (Dong and Lapata, 2016; Jia and Liang, 2016; Iyer et al., 2017), many state-of-the-art results on WikiSQL have achieved by constraining the output space with the SQL syntax. The initial model proposed by (Zhong et al., 2017) independently generates the two components of the target SQL query, `select`-clause and `where`-clause, which outperforms the vanilla sequence-to-sequence baseline model proposed by the same authors. SQLNet (Xu et al., 2017) further simplifies the generation task by introducing a sequence-to-set model in which only `where` condition value is generated by the sequence-to-sequence model. TypeSQL (Yu et al., 2018) also employs a sequence-to-set structure but with an additional “type" information of natural language tokens.

Coarse2Fine (Dong and Lapata, 2018) first generates rough intermediate output, and then refines the results by decoding full `where`-clauses. Also, the table-aware contextual representation of the question is generated with bi-LSTM with attention mechanism which increases logical form accuracy by 3.1%. Our approach differs in that many layers of self-attentions (Vaswani et al., 2017; Devlin et al., 2018) are employed with a single concatenated input of question and table headers for stronger contextualization of the question.

Pointer-SQL (Wang et al., 2017) proposes a sequence-to-sequence model that uses an attention-based copying mechanism and a value-based loss function. Annotated Seq2seq (Wang et al., 2018b) utilizes a sequence-to-sequence model after automatic annotation of input natural language. MQAN (McCann et al., 2018) suggests a multitask question answering network that jointly learns multiple natural language processing tasks using various attention mechanisms. Execution guided decoding is

[1] The source code and human evaluation data is available from `https://github.com/naver/sqlova`.

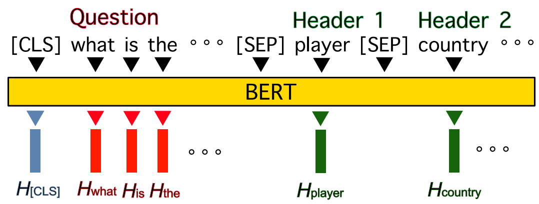


Figure 2: (A) The scheme of input encoding process by table-aware BERT. Final output vectors are represented by colored bars: light blue for `[CLS]` output, red for question words, and green for tokens from table headers.

suggested in (Wang et al., 2018a), in which non-executable (partial) SQL queries candidates are removed from output candidates during decoding step. IncSQL (Shi et al., 2018) proposes a sequence-to-action parsing approach that uses incremental slot filling mechanism with feasible actions from a pre-defined inventory.

3 Model

Our model, SQLOVA, consists of two layers: encoding layer that obtains table- and context-aware question word representations (Section 3.1), and NL2SQL layer that generates the SQL query from the encoded representations (Section 3.2).

3.1 Table-aware Encoding Layer

We extend BERT (Devlin et al., 2018) for encoding the natural language query together with the headers of the entire table. We use `[SEP]`, a special token in BERT, to separate between the query and the headers. That is, each query input $T_{n,1} \dots T_{n,L}$ (L is the number of query words) is encoded as

$$\texttt{[CLS]},\ T_{n,1},\ \cdots\ T_{n,L},\ \texttt{[SEP]},\ T_{h_1,1},\ T_{h_1,2},\ \cdots,\ \texttt{[SEP]},\ \cdots,\ \texttt{[SEP]},\ T_{h_{N_h},1},\ \cdots,\ T_{h_{N_h},M_{N_h}},\texttt{[SEP]}$$

where $T_{h_j,k}$ is the k-th token of the j-th table header, M_j is the total number of tokens of the j-th table headers, and N_h is the total number of table headers. Another input to BERT is the segment id, which is either 0 or 1. We use 0 for the question tokens and 1 for the header tokens. Other configurations largely follow (Devlin et al., 2018). The output from the final two layers of BERT are concatenated and used in NL2SQL LAYER (Section 3.2).

3.2 NL2SQL Layer

In this section, we describe the details of NL2SQL LAYER (Figure 3) on top of the table-aware encoding layer.

In a typical sequence generation model, the output is not explicitly constrained by any syntax, which is highly suboptimal for formal language generation. Hence, following (Xu et al., 2017), NL2SQL LAYER uses syntax-guided sketch, where the generation model consists of six modules, namely `select-column`, `select-aggregation`, `where-number`, `where-column`, `where-operator`, and `where-value` (Figure 3). Also, following (Xu et al., 2017), column-attention is frequently used to contextualize question.

In all sub-modules, the output of table-aware encoding layer (Section 3.1) is further contextualized by two layers of bidirectional LSTM layers with 100 dimension. We denote the LSTM output of the n-th token of the question with E_n. Header tokens are encoded separately and the output of final token of each header from LSTM layer is used. D_c is used to denotes the encoding of header c. The role of each sub-module is described below.

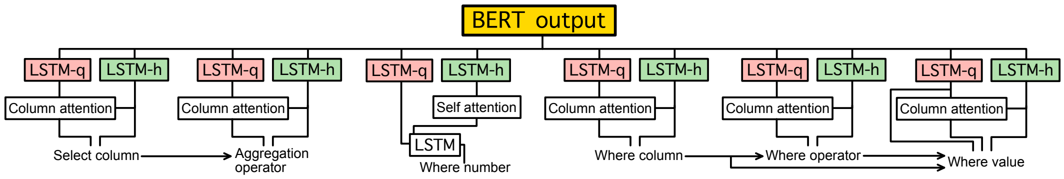


Figure 3: The illustration of NL2SQL LAYER (Section 3.2). The outputs from table-aware encoding layer are encoded again with `LSTM-q` (question encoder) and `LSTM-h` (header encoder).

`select-column` finds `select` column from given natural language utterance by contextualizing question through column-attention mechanism (Xu et al., 2017).

$$\begin{aligned} s(n|c) &= D_c^T \mathcal{W} E_n \\ p(n|c) &= \text{softmax}(s(n|c)) \\ C_c &= \sum_n p(n|c) E_n \\ s_{sc}(c) &= \mathcal{W} \tanh([\mathcal{W} D_c; \mathcal{W} C_c]) \\ p_{sc}(c) &= \text{softmax}(s_{sc}(c)) \end{aligned} \tag{1}$$

Here, $\mathcal{W}$ stands for affine transformation, C_n is context vector of question for given column, $[\cdot;\cdot]$ denotes the concatenation of two vectors, and $p_{sc}(c)$ indicates the probability of generating column c. To make the equation uncluttered, same $\mathcal{W}$ is used to denote any affine transformation in our paper although all of them denote different transformations.

`select-aggregation` finds aggregation operator `agg` for given column c among six possible choices (`NONE`, `MAX`, `MIN`, `COUNT`, `SUM`, and `AVG`). Its probability is obtained by

$$p_{sa}(\texttt{agg}|c) = \text{softmax}(\mathcal{W} \tanh \mathcal{W} C_c) \tag{2}$$

where C_c is the context vector of the question obtained by the same way in `select-column`.

`where-number` finds the number of `where` condition by contextualizing column (C) via self-attention and contextualizing question (C_Q) conditioned on C.

$$\begin{aligned} p(c) &= \text{softmax}(\mathcal{W} D_c) \\ C &= \sum_c p(c) D_c \\ h &= \mathcal{W} C \\ c &= \mathcal{W} C \\ p'(n) &= \text{softmax}(\mathcal{W}\ \text{bi-LSTM}(E_n, h, c)) \\ C_Q &= \sum_n p'(n) E_n \\ s_{wn} &= \mathcal{W} \tanh \mathcal{W} C_Q \end{aligned} \tag{3}$$

Here, h and c are initial “hidden" and “cell" inputs to LSTM encoder. The probability of observing k number of `where` condition is found from k-th element of vector softmax(s_{wn}). This submodule is same with that of SQLNet (Xu et al., 2017) and shown here for comprehensive reading.

`where-column` obtains the probability of observing column c ($p_{wc}(c)$) through column-attention,

$$\begin{aligned} s_{wc}(c) &= \mathcal{W} \tanh([\mathcal{W} D_c; \mathcal{W} C_c]) \\ p_{wc}(c) &= \text{sigmoid}(s_{wc}(c)) \end{aligned} \tag{4}$$

where C_c is the context vector of the question obtained by the same way as in `select-column`. The probability of generating each column is obtained separately from `sigmoid` function and top k columns are selected. k is found from `where-number` sub-module.

`where-operator` finds `where` operator `op` ($\in \{=, >, <\}$) for given column c through column-attention.

$$\begin{aligned} s_{wo}(\mathtt{op}|c) &= \mathcal{W} \tanh \mathcal{W}([\mathcal{W}D_c; \mathcal{W}C_c]) \\ p_{wo}(\mathtt{op}|c) &= \text{softmax} s_{wo}(\mathtt{op}|c) \end{aligned} \tag{5}$$

where C_c is the context vector of the question obtained by the same way in `select-column`.

`where-value` finds `where` condition by locating start- and end-tokens from question for given column c and operator `op`.

$$\begin{aligned} \mathtt{vec} &= [E_n; \mathcal{W}C_c; \mathcal{W}D_c; \mathcal{W}V_{\mathtt{op}}] \\ s_{wv}(n|c, \mathtt{op}) &= \mathcal{W} \tanh \mathcal{W}\, \mathtt{vec} \end{aligned} \tag{6}$$

Here, $V\mathtt{op}$ stands for one-hot vector of `op` ($\in \{=, >, <\}$). The probability of n-th token of question being start-index for given c-th column and `op` is obtained by feeding 1st element of s_{wv} vector to softmax function whereas that of end-index is obtained by using 2nd element of s_{wv} by the same way.

To sum up, our NL2SQL LAYER is motivated by SQLNet (Xu et al., 2017) but have following key differences. Unlike SQLNet, NL2SQL LAYER does not share parameters. Also, instead of using pointer network for inferring the `where` condition values, we train for inferring the start and the end positions of the utterance, following (Dong and Lapata, 2018). Furthermore, the inference of the start and the end tokens in `where-value` module depends on both selected `where-column` and `where-operators` while the inference relies on `where-columns` only in (Xu et al., 2017). Lastly, when combining two vectors corresponding to the question and the headers, concatenation instead of addition is used.

Execution-Guided Decoding (EG) During the decoding (SQL query generation) stage, non-executable (partial) SQL queries can be excluded from the output candidates for more accurate results, following the strategy suggested by (Wang et al., 2018a; Yin and Neubig, 2018). In `select` clause, (`select` column, aggregation operator) pairs are excluded when the string-type columns are paired with numerical aggregation operators such as `MAX`, `MIN`, `SUM`, or `AVG`. The pair with highest joint probability is selected from remaining pairs. In `where` clause decoding, the executability of each (`where` column, operator, value) pair is tested by checking the answer returned by the partial SQL query $\mathtt{select\ agg(col}_s\mathtt{)\ where\ col}_w\mathtt{\ op\ val}$. Here, $\mathtt{col}_s$ is the predicted `select` column, `agg` is the predicted aggregation operator, $\mathtt{col}_w$ is one of the `where` column candidates, `op` is `where` operator, and `val` stands for the `where` condition value. The queries with empty returns are also excluded from the candidates. The final output of `where` clause is determined by selecting the output maximizing the joint probability estimated from the output of `where-number`, `where-column`, `where-operator`, and `where-value` modules.

4 Experiments

During training, BERT-based table-aware encoding layer (`BERT-Large-Uncased`[2]) are loaded and fine-tuned with ADAM optimizer with the learning rate of 10^{-5}, whereas NL2SQL LAYER is trained with the learning rate of 10^{-3}. In both cases, the decay rates of ADAM optimizer are set to $\beta_1 = 0.9, \beta_2 = 0.999$. Batch size is set to 32. To find word vectors, natural language utterance is first tokenized by using Standford CoreNLP (Manning et al., 2014). Each token is further tokenized (into sub-word level) by WordPiece tokenizer (Devlin et al., 2018; Wu et al., 2016). The headers of the tables and SQL vocabulary are tokenized by WordPiece tokenizer directly. The PyTorch version of BERT code[3] is used for word embedding and some part of the code in NL2SQL LAYER is influenced by the original SQLNet source code[4]. All experiments were performed on WikiSQL ver. 1.1 [5].

[2] https://github.com/google-research/bert
[3] https://github.com/huggingface/pytorch-pretrained-BERT
[4] https://github.com/xiaojunxu/SQLNet
[5] https://github.com/salesforce/WikiSQL

Table 1: Comparison of various models. Logical from accuracy (LF) and execution accuracy (X) on dev and test set of WikiSQL. "EG" stands for "execution-guided".

Model	Dev LF (%)	Dev X (%)	Test LF (%)	Test X (%)
Baseline (Zhong et al., 2017)	23.3	37.0	23.4	35.9
Seq2SQL (Zhong et al., 2017)	49.5	60.8	48.3	59.4
SQLNet (Xu et al., 2017)	63.2	69.8	61.3	68.0
PT-MAML (Huang et al., 2018)	63.1	68.3	62.8	68.0
TypeSQL (Yu et al., 2018)	68.0	74.5	66.7	73.5
Coarse2Fine (Dong and Lapata, 2018)	72.5	79.0	71.7	78.5
MQAN (McCann et al., 2018)	76.1	82.0	75.4	81.4
Annotated Seq2seq (Wang et al., 2018b) [1]	72.1	82.1	72.1	82.2
IncSQL (Shi et al., 2018) [1]	49.9	84.0	49.9	83.7
BERT-TO-SEQUENCE (ours)	57.3	-	56.4	-
BERT-TO-TRANSFORMER (ours)	70.5	-	-	-
SQLOVA (ours)	**81.6** (**+5.5**)	**87.2** (**+3.2**)	**80.7** (**+5.3**)	**86.2** (**+2.5**)
PointSQL+EG (Wang et al., 2018a) [1,2]	67.5	78.4	67.9	78.3
Coarse2Fine+EG (Wang et al., 2018a) [1,2]	76.0	84.0	75.4	83.8
IncSQL+EG (Shi et al., 2018) [1,2]	51.3	87.2	51.1	87.1
SQLOVA+EG (ours) [2]	**84.2** (**+8.2**)	**90.2** (**+3.0**)	**83.6** (**+8.2**)	**89.6** (**+2.5**)
Human performance [3]	-	-	-	88.3

[1] Source code is not opened.
[2] Execution guided decoding is employed.
[3] Measured over 1,551 randomly chosen samples from WikiSQL test set (Section 5).

4.1 Accuracy Measurement

The logical form (LF) and the execution accuracy (X) on dev set (consisting of 8,421 queries) and test set (consisting of 15,878 queries) of WikiSQL of several models are shown in Table 1. The execution accuracy is measured by evaluating the answer returned by 'executing' the query on the SQL database. The order of `where` conditions is ignored in measuring logical form accuracy in our models. The top rows in Table 1 show models without execution guidance (EG), and the bottom rows show models augmented with EG. SQLOVA outperforms previous baselines by a large margin, achieving [+5.3% LF] and [+2.5% X] for non-EG and achieving [+8.2% LF] and [+2.5% X] for EG.

To understand the performance of SQLOVA in detail, the logical form accuracy of each sub-module was obtained and shown in Table 2. All sub-modules show $\gtrsim$ 95% in accuracy except `select-aggregation` module whose low accuracy partially results from the error in the ground-truth of WikiSQL (Section-5).[6]

Table 2: The logical from accuracy of each sub-module over WikiSQL dev set. `s-col, s-agg, w-num, w-col, w-op` and `w-val` stand for `select-column, select-aggregation, where-number, where-column, where-operstor`, and `where-value` respectively.

Model	`s-col`	`s-agg`	`w-num`	`w-col`	`w-op`	`w-val`
SQLOVA, Dev	97.3	90.5	98.7	94.7	97.5	95.9
SQLOVA, Test	96.8	90.6	98.5	94.3	97.3	95.4

Choosing not to answer for low-confidence predictions is another important measure of performance. We use the output probability of a generated SQL query from SQLOVA as the confidence score and predict that the question is *unanswerable* when the score is low. The result shows that SQLOVA effectively assigns a low probability to wrong predictions, yielding a high precision of 95%+ with a recall rate of 80%. The precision-recall curve and its area under curve are shown in Figure A2.

[6] In addition to SQLOVA we also provide two BERT-based models which also outperforms previous baselines by large margin, in Appendix A.1.

4.2 Ablation Study

To understand the importance of each part of SQLOVA, we evaluate ablations in Table 3. The results show that word contextualization (without fine-tuning) contributes to overall logical form accuracy by 4.1% (dev) and 3.9% (test) (compare third and fifth rows of the table) which is similar to the observation by (Dong and Lapata, 2018) where the 3.1% increases observed with table-aware LSTM encoder. Consistently, replacing BERT by ELMo (Peters et al., 2018) shows similar results (fourth row of the table). But unlike GloVe, where fine-tuning increases only a few percents in accuracy (Xu et al., 2017), fine-tuning of BERT increases the accuracies by 11.7% (dev) and 12.2% (test) (compare first and third rows in the table) which may be attributed to the use of many layers of self-attention (Vaswani et al., 2017). Use of `BERT-Base` decreases the accuracy by 1.3% on both dev and test set compared to `BERT-Large` cases. We also developed BERT-TO-SEQUENCE where the encoder part of vanilla sequence-to-sequence model with attention (Jia and Liang, 2016) is replaced by BERT. The model achieves 57.3% and 56.4% logical form accuracies in dev and test sets respectively (Table 1) highlighting the importance of using proper decoding layers. To further validate the conclusion, we replaced LSTM decoder in into Transformer (BERT-TO-TRANSFORMER), the model achieved 70.5 logical form accuracy in dev set again achieving 11.1% lower score compared to SQLOVA. The detailed description of the model is presented in Appendix A.1.3.

Table 3: The results of ablation study. Logical from accuracy (LF) and execution accuracy (X) on dev and test sets of WikiSQL are shown.

Model	Dev LF (%)	Dev X (%)	Test LF (%)	Test X (%)
SQLOVA	81.6	87.2	80.7	86.2
(-) BERT-Large (+) BERT-Base	80.3	85.8	79.4	85.2
(-) Fine-tuning	69.9	77.0	68.5	75.6
(-) BERT-Large (+) ELMo (fine-tuned)	71.3	77.7	69.6	76.0
(-) BERT-Large (+) GloVe	65.8	72.9	64.6	71.7

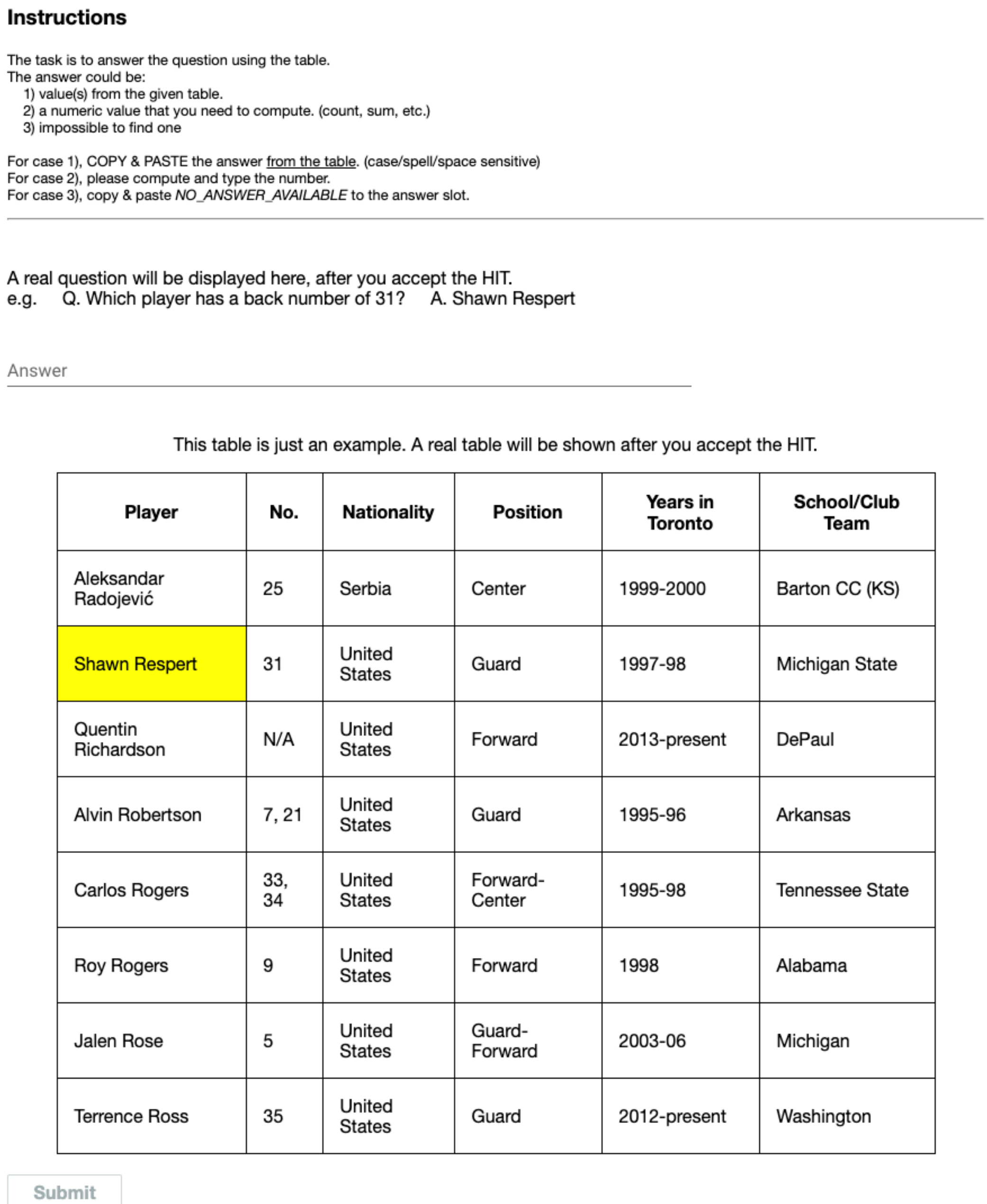
5 Analysis

5.1 Error Analysis

There are 1,533 mismatches in logical form between the ground-truth (GT) and the predictions from SQLOVA in WikiSQL dev set. Among the mismatches, 100 samples were randomly selected, analyzed, and classified into two categories: (1) 26 "unanswerable" cases of which it is not possible to generate correct SQL query for given information (question and table schema), and (2) 74 "answerable" cases.

Unanswerable cases were further categorized into the following four types.

- Type I: the headers of tables do not contain the necessary information. For example, a question "`What was the score between Marseille and Manchester United on the second leg of the Champions League Round of 16?`" and its corresponding table headers {'`Team`', '`Contest and round`', '`Opponent`', '`1st leg score`', '`2nd leg score`', '`Aggregate score`'} in QID-1986 (Table 7) do not contain information about which header should be selected for condition values '`Manchester United`' and '`Marseille`'.
- Type II: There exist multiple valid SQL queries per question (QID-783, 2175, 4229 in Table 7). For example, the GT SQL query of QID-783 has `count` aggregation operator and any header can be used for `select` column.
- Type III: the generation of nested SQL query is required. For example, correct SQL query for QID-332 (Table 7) is "`SELECT count(incumbent) WHERE District=(SELECT District WHERE Incumbent=Alvin Bush)`".
- Type IV: questions are ambiguous. For example, the answer to the question "`What is the number of the player who went to Southern University?`" in QID-156 (Table 7) can vary depending on the interpretation of "`the number of the player`".



Instructions

The task is to answer the question using the table.
The answer could be:
1) value(s) from the given table.
2) a numeric value that you need to compute. (count, sum, etc.)
3) impossible to find one

For case 1), COPY & PASTE the answer from the table. (case/spell/space sensitive)
For case 2), please compute and type the number.
For case 3), copy & paste *NO_ANSWER_AVAILABLE* to the answer slot.

A real question will be displayed here, after you accept the HIT.
e.g. Q. Which player has a back number of 31? A. Shawn Respert

Answer

This table is just an example. A real table will be shown after you accept the HIT.

Player	No.	Nationality	Position	Years in Toronto	School/Club Team
Aleksandar Radojević	25	Serbia	Center	1999-2000	Barton CC (KS)
Shawn Respert	31	United States	Guard	1997-98	Michigan State
Quentin Richardson	N/A	United States	Forward	2013-present	DePaul
Alvin Robertson	7, 21	United States	Guard	1995-96	Arkansas
Carlos Rogers	33, 34	United States	Forward-Center	1995-98	Tennessee State
Roy Rogers	9	United States	Forward	1998	Alabama
Jalen Rose	5	United States	Guard-Forward	2003-06	Michigan
Terrence Ross	35	United States	Guard	2012-present	Washington

Submit

Figure 4: The instruction and example given to crowdworkers during human performance evaluation on the WikiSQL dataset

The categorization of 26 samples is summarized in Table 6 in Appendix A.4..

Further analysis over the remaining 74 answerable examples reveals that there are 49 GT errors in logical forms. 45 out of 49 examples contain GT errors in aggregation operators (e.g. QID-7062), two have GT errors in `select` columns (e.g. QID-841, 5611), and remaining two contain GT errors in `where` clause (e.g. QID-2925, 7725). Interestingly, among 49 examples, 41 logical forms are correctly predicted by SQLOVA, indicating that the actual performances of the models in Table 1 are underestimated. This also may imply that most of examples in WikiSQL have correct GT for training. The results are summarized in Table 5, and all 100 examples are presented in Table 7 in Appendix A.5.

As the questions in WikiSQL are created by paraphrasing queries generated automatically from the templates without considering the table contents, the meanings of the questions could change, especially when the quantitative answer is required, possibly leading to GT errors. For example, QID-3370 in Table 7 is related to an "`year`" and the GT SQL query includes unnecessary `COUNT` aggregation operators.

Overall, the error analysis above may imply that near-90% accuracy of SQLOVA could be near the upper bound in WikiSQL task the "answerable" and non-erroneous questions when the contents of tables are not available.

5.2 Measuring Human Performance

The human performance on WikiSQL dataset has not been measured so far despite its popularity. Here, we provide the approximate human performance by collecting answers from 246 different crowdworkers through Amazon Mechanical Turk over 1,551 randomly sampled examples from the WikiSQL test set (which has 15,878 examples in total). The crowdworkers were selected with following three constraints: (1) 95% or higher task acceptance rate; (2) 1000 or higher HITs; (3) residents of the United States.

During the evaluation, crowdworkers were asked either to find value(s) or to compute a value using the given questions and corresponding tables following the instruction provided (Figure 4). Note that the task requires general capability of understandings English text and finding values from a table without a need for the generation of SQL queries. This effectively mimics the measurement of execution accuracy in WikiSQL. We find that the accuracy of crowdworkers on the randomly sampled test data is 88.3%, as shown in Table 1 while the execution accuracy of SQLOVA over 1,551 samples are 86.8% (w/o EG) and 91.0% (w/ EG). [7]

We manually checked and analyzed all answers from the crowd. Errors made by crowdworkers are similar to that of the model such as a mismatch of `select` columns or `where` columns. One notable mistake by only humans (that our model does not make) is confusion on the ambiguity of natural language. For example, when a question is asking a column value with more than two conditions, crowdworkers show the tendency to consider a single condition only because multiple conditions were written with "and" which is often considered as the meaning of "or" in real life.

6 Conclusion

In this paper, we propose the first NL2SQL model to achieve a super-human accuracy in WikiSQL. We demonstrate the effectiveness of a careful architecture design that brings and combines previous approaches in NL2SQL and table-aware word contextualization with large pretrained language model (BERT) together. We propose a BERT-based table-aware encoder and a task-specific module on the top of the encoder, outperforming the previous best model by 8.2% and 2.5% in logical form and execution accuracy, respectively. We hope our detailed explanation and analysis of the model and the dataset provide an insight on how future research on NL2SQL models and datasets can be effectively approached.

Acknowledgments

We thank Clova AI members for their great support, especially Jung-Woo Ha for proof-reading the manuscript, Sungdong Kim and Dongjun Lee for providing help on using BERT, Guwan Kim for insightful comments. We also thank the Hugging Face Team for sharing the PyTorch implementation of BERT.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *NAACL*, abs/1810.04805.

[7] When measuring human performance, errors in ground truth are manually corrected by experts (us).

Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33–43, Berlin, Germany. Association for Computational Linguistics.

Li Dong and Mirella Lapata. 2018. Coarse-to-fine decoding for neural semantic parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 731–742, Melbourne, Australia. Association for Computational Linguistics.

Po-Sen Huang, Chenglong Wang, Rishabh Singh, Wen tau Yih, and Xiaodong He. 2018. Natural language to structured query generation via meta-learning. In *NAACL*.

Srinivasan Iyer, Ioannis Konstas, Alvin Cheung, Jayant Krishnamurthy, and Luke Zettlemoyer. 2017. Learning a neural semantic parser from user feedback. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 963–973, Vancouver, Canada. Association for Computational Linguistics.

Robin Jia and Percy Liang. 2016. Data recombination for neural semantic parsing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12–22, Berlin, Germany. Association for Computational Linguistics.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2018. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

Tianze Shi, Kedar Tatwawadi, Kaushik Chakrabarti, Yi Mao, Oleksandr Polozov, and Weizhu Chen. 2018. Incsql: Training incremental text-to-sql parsers with non-deterministic oracles. *CoRR*, abs/1809.05054.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.

Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. Pointer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2692–2700. Curran Associates, Inc.

Chenglong Wang, Marc Brockschmidt, and Rishabh Singh. 2017. Pointing out SQL queries from text. Technical Report MSR-TR-2017-45, Microsoft.

Chenglong Wang, Po-Sen Huang, Alex Polozov, Marc Brockschmidt, and Rishabh Singh. 2018a. Execution-guided neural program decoding. In *ICML workshop on Neural Abstract Machines and Program Induction v2 (NAMPI)*.

Wenlu Wang, Yingtao Tian, Hongyu Xiong, Haixun Wang, and Wei-Shinn Ku. 2018b. A transfer-learnable natural language interface for databases. *CoRR*, abs/1809.02649.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.

Xiaojun Xu, Chang Liu, and Dawn Song. 2017. Sqlnet: Generating structured queries from natural language without reinforcement learning. *CoRR*, abs/1711.04436.

Pengcheng Yin and Graham Neubig. 2018. TRANX: A transition-based neural abstract syntax parser for semantic parsing and code generation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 7–12, Brussels, Belgium. Association for Computational Linguistics.

Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, and Dragomir Radev. 2018. TypeSQL: Knowledge-based type-aware neural text-to-SQL generation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 588–594, New Orleans, Louisiana. Association for Computational Linguistics.

Victor Zhong, Caiming Xiong, and Richard Socher. 2017. Seq2sql: Generating structured queries from natural language using reinforcement learning. *CoRR*, abs/1709.00103.

A Appendix

A.1 Additional models

A.1.1 SHALLOW-LAYER

Here, we present another task specific layer SHALLOW-LAYER having lower model complexity compared to NL2SQL LAYER. SHALLOW-LAYER does not contain trainable parameters but controls the flow of information during fine-tuning of BERT via loss function. Like NL2SQL LAYER, SHALLOW-LAYER uses syntax-guided sketch, where the generation model consists of six modules, namely `select-column`, `select-aggregation`, `where-number`, `where-column`, `where-operator`, and `where-value` (Figure A1A).

`select-column` module finds the column in `select` clause from given natural language utterance by modeling the probability of choosing i-th header ($p_{sc}(\texttt{col}_i)$) as

$$p_{sc}(\texttt{col}_i) = \text{softmax}(H_{h,i})_0 \tag{A1}$$

where $H_{h,i}$ is the contextualized output vector of first token of i-th header by table-aware BERT encoder, and $(H_{h,i})_0$ indicates zeroth element of the vector $H_{h,I}$. In general, $(V)_\mu$ denotes μ-th element of vector V in this paper. Also, the conditional probability for given question and table-schema $p(\cdot|\text{Q}, \text{table-schema})$ is simply denoted as $p(\cdot)$ to make equation uncluttered.

`select-aggregation` module finds the aggregation operator for the given `select` column. The probability of generating aggregation operator `agg` for given `select` column $\texttt{col}_i$ is described by

$$p_{sa}(\texttt{agg}_\mu|\texttt{col}_i) = \text{softmax}\left((H_{h,i})_\mu\right) \tag{A2}$$

where $\texttt{agg}_1$, $\texttt{agg}_2$, $\texttt{agg}_3$, $\texttt{agg}_4$, $\texttt{agg}_5$, and $\texttt{agg}_6$ are `none`, `max`, `min`, `count`, `sum`, and `avg` respectively.

`where-number` module predicts the number of `where` conditions by modeling the probability of generating μ-number of conditions as

$$p_{wn}(\mu) = \text{softmax}\left((\mathcal{W}H_{\texttt{[CLS]}})_\mu\right) \tag{A3}$$

where $H_{\texttt{[CLS]}}$ is the output vector of `[CLS]` token from table-aware BERT encoder, and $\mathcal{W}$ stands for affine transformation. Throughout the paper, any affine transformation shall be denoted by $\mathcal{W}$ for the clarity.

`where-column` module calculates the probability of generating each columns in `where` clause. The probability of generating $\texttt{col}_i$ is given by

$$p_{wc}(\text{col}_i) = \text{sigmoid}((H_{h,i})_7). \tag{A4}$$

`where-operator` module finds most probable operators for given `where` column among three possible choices ($>, =, <$). The probability of generating operator $\texttt{op}_\mu$ for given `where` column $\texttt{col}_i$ is modeled as

$$p_{wo}(\texttt{op}_\mu|\texttt{col}_i) = \text{softmax}\left((H_{h,i})_\mu\right) \tag{A5}$$

where $\texttt{op}_8$, $\texttt{op}_9$, and $\texttt{op}_{10}$ are >, =, and < respectively.

`where-value` module finds which tokens of a question correspond to condition values for given `where` columns by locating start- and end-tokens. The probability that k-th question token is selected as a start token for given `where` column $\texttt{col}_\mu$ is modeled as

$$p_{wv,st}(k|\texttt{col}_\mu) = \text{softmax}\left((H_{n,k})_\mu\right). \tag{A6}$$

Similarly the probability of k-th question token is selected as an end token is

$$p_{wv,ed}(k|\texttt{col}_\mu) = \text{softmax}\left((H_{n,k})_{\mu+100}\right) \tag{A7}$$

100 is selected to avoid overlap during inference between start- and end-token models. The maximum number of table headers in single table is 44 in WikiSQL task.

A.1.2 DECODER-LAYER

DECODER-LAYER contains LSTM decoders adopted from pointer network (Vinyals et al., 2015; Zhong et al., 2017) (Fig. 3B) with following special features. Instead of generating entire header tokens, we only generate first token of each header and interpret them as entire header tokens during inference stage using `Point-to-SQL` module (Fig. 3B). Similarly, the model generates only the pointers to start- and end- `where`-value tokens omitting intermediate points. Decoding process can be expressed as following equations which use the attention mechanism.

$$\begin{aligned}
D_t &= \text{LSTM}(P_{t-1}, (h_{t-1}, c_{t-1})) \\
h_0 &= (\mathcal{W}H_{(\texttt{[CLS]})})_{0:d} \\
c_0 &= (\mathcal{W}H_{(\texttt{[CLS]})})_{d:2d} \\
s_t(i) &= \mathcal{W}(\mathcal{W}H_i + \mathcal{W}D_t) \\
p_t(i) &= \text{softmax}\; s_t(i).
\end{aligned} \tag{A8}$$

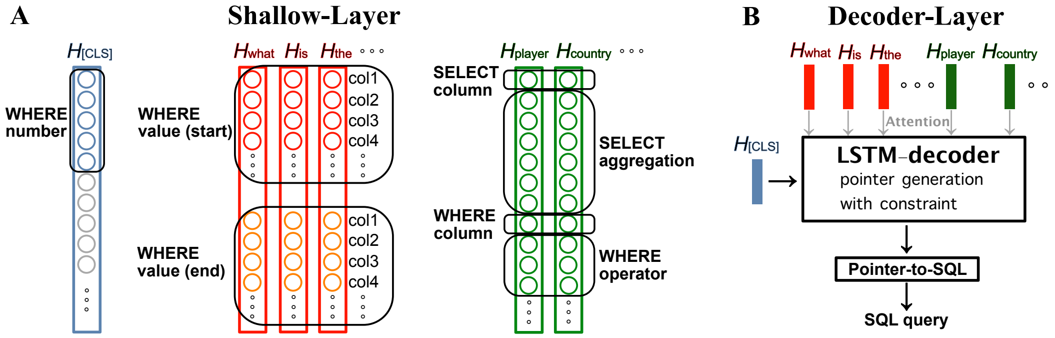


Figure A1: (A) The model scheme of SHALLOW-LAYER. Each circle represents the single element of the output vector from table-aware BERT-encoder. The circles are grouped by black squares according to their roles in SQL query generation. (B) The scheme DECODER-LAYER. `LSTM-docoder` of pointer network (Vinyals et al., 2015) generates the sequence of pointers to augmented inputs which include SQL vocabulary, `start`, `end`, question words, and header tokens. Generated pointer squences are interpreted by `Pointer-to-SQL` module which generates final SQL queries.

P_{t-1} stands for the one-hot vector (pointer) at time $t-1$, h_{t-1} and c_{t-1} are hidden- and cell- vectors of LSTM decoder, d is the hidden dimension of BERT, H_i is the BERT output of i-th token, and $p_t(i)$ is the probability observing i-th token at time t.

A.1.3 BERT-TO-SEQUENCE

BERT-TO-SEQUENCE consists of the table-aware BERT encoder and LSTM decoder which is essentially a sequence-to-sequence model (with attention) (Jia and Liang, 2016) except that the LSTM encoder part is replaced by BERT. The encoding process is same with NL2SQL LAYER. The decoding process is described by following equations.

$$\begin{aligned}
E_t &= \mathrm{emb}_{\mathrm{BERT}}(w_t) \\
\rho_i(t) &= H_i^T \mathcal{W} E_t \\
C_t &= \sum_i H_i \rho_i(t) \\
h_{t+1} &= \mathrm{LSTM}([C_t; E_t], h_t) \\
h_0 &= (\mathcal{W} H_{(\texttt{[CLS]})})_{0:d} \\
c_0 &= (\mathcal{W} H_{(\texttt{[CLS]})})_{d:2d} \\
p(w_{t+1}) &= \mathrm{softmax}(\mathcal{W} \tanh h_{t+1}[C_t; h_{t+1}])
\end{aligned} \tag{A9}$$

where w_t stands for word token (among 30,522 token vocabulary used in BERT) predicted at time t, $\mathrm{emb}_{\mathrm{BERT}}$ is a map that transform the token to embedding vector E_t via word embedding module of BERT, H_i is the output vector from BERT encoder of i-th input token, and $p(w_{t+1})$ is the probability of generating token w_{t+1} at time $t+1$.

A.1.4 The performance of SHALLOW-LAYER and DECODER-LAYER

Compared to previous best results, SHALLOW-LAYER shows +5.5% LF and +3.1% X, DECODER-LAYER shows +4.4% LF and +1.8% X for non-EG case (Table. 4). For EG case, SHALLOW-LAYER shows +6.4% LF and +0.4% X, DECODER-LAYER shows +7.8% LF and +2.5% X (Table. 4).

SHALLOW-LAYER shows [+6.% LF] and [+3.1% X] whereas

Table 4: Logical from accuracy (LF) and execution accuracy (X) on dev and test set of WikiSQL. "EG" stands for "execution-guided".

Model	Dev LF (%)	Dev X (%)	Test LF (%)	Test X (%)
SHALLOW-LAYER (ours)	**81.5** (+5.4)	**87.4** (+3.2)	**80.9** (+5.5)	**86.8** (+3.1)
DECODER-LAYER (ours)	**79.7** (+3.6)	**85.5** (+1.1)	**79.8** (+4.4)	**85.5** (+1.8)
SHALLOW-LAYER-EG (ours)	**82.3** (+6.3)	**88.1** (+0.9)	**81.8** (+6.4)	**87.5** (+0.4)
DECODER-LAYER-EG (ours)	**83.4** (+7.4)	**89.9** (+2.7)	**83.2** (+7.8)	**89.6** (+2.5)

A.2 The Precision-Recall Curve

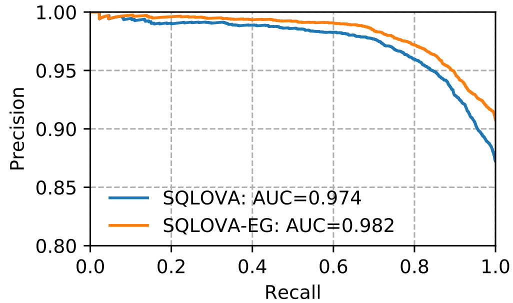


Figure A2: Precision-Recall curve and area under curve (AUC) with SQLova (blue) and SQLova-EG (orange). Precision and recall rates are controlled by varying the threshold value for the confidence score.

A.3 The Contingency Table

Table 5: Contingency table of 74 answerable questions. Corresponding 74 ground truth- (GT) and predicted-SQL queries by SQLova are manually classified to correct and incorrect cases.

		SQL (GT)		
		correct	incorrect	total
SQL (Ours)	correct	0	**41**	41
	incorrect	25	8	33
	total	25	49	74

A.4 The Types of Unanswerable Examples

Table 6: 26 unanswerable examples. "types" denotes the type of unanswerable cases that each question belongs to. "total" means the number of examples in the type.

types	QID	total
Type I	19, 261, 557, 597, 598, 738, 1089, 1122, 1430, 1986, 2891, 3050, 3602, 3925, 5893, 6028, 6533, 7070, 7912, 8041, 8111	21
Type II	783, 2175, 4229	3
Type III	332	1
Type IV	156	1

A.5 100 Examples in the WikiSQL Dataset

Table 7: The dataset examples from WikiSQL dev set used in Section 5. 100 samples were randomly selected from 1,533 mismatches between the ground-truth and the predictions of SQLova. QID denotes an index of the question among 8,421 wikiSQL dev set data. There are three types of queries: natural language queries (NL), ground truth SQL queries (SQL (T)), predicted SQL queries (SQL (P)). Other fields indicate ground truth answer (ANS (T)), predicted answer (ANS (P)), and a type of error (ERROR), respectively. Note that the types of unanswerable cases that the question belongs to are shown in the parentheses after "Question" in the "Error" field.

No.	QID	Type	Description
1	19	NL	How many capital cities does Australia have?

		TBL	“Country (exonym)”, “Capital (exonym)”, “Country (endonym)”, “Capital (endonym)”, “Official or native language(s) (alphabet/script)”
		SQL (T)	`SELECT` count(Capital (endonym)) `FROM` 1-1008653-1 `WHERE` Country (endonym) = Australia
		SQL (P)	`SELECT` count(Capital (exonym)) `FROM` 1-1008653-1 `WHERE` Country (exonym) = australia
		ANS (T)	1
		ANS (P)	1
		ERROR	Qestion (I)
2	55	NL	What are the races that johnny rutherford has won?
		TBL	“Rd”, “Name”, “Pole Position”, “Fastest Lap”, “Winning driver”, “Winning team”, “Report”
		SQL (T)	`SELECT` (Name) `FROM` 1-10706879-3 `WHERE` Winning driver = Johnny Rutherford
		SQL (P)	`SELECT` (Rd) `FROM` 1-10706879-3 `WHERE` Winning driver = johnny rutherford
		ANS (T)	kraco car stereo 200
		ANS (P)	1.0
		ERROR	None
3	156	NL	What is the number of the player who went to Southern University?
		TBL	“Player”, “No.(s)”, “Height in Ft.”, “Position”, “Years for Rockets”, “School/Club Team/Country”
		SQL (T)	`SELECT` (No.(s)) `FROM` 1-11734041-9 `WHERE` School/Club Team/Country = Southern University
		SQL (P)	`SELECT` count(No.(s)) `FROM` 1-11734041-9 `WHERE` School/Club Team/Country = southern university
		ANS (T)	6
		ANS (P)	1
		ERROR	Qestion (IV)
4	212	NL	What is the toll for heavy vehicles with 3/4 axles at Verkeerdevlei toll plaza?
		TBL	“Name”, “Location”, “Light vehicle”, “Heavy vehicle (2 axles)”, “Heavy vehicle (3/4 axles)”, “Heavy vehicle (5+ axles)”
		SQL (T)	`SELECT` (Heavy vehicle (3/4 axles)) `FROM` 1-1211545-2 `WHERE` Name = Verkeerdevlei Toll Plaza
		SQL (P)	`SELECT` (Heavy vehicle (3/4 axles)) `FROM` 1-1211545-2 `WHERE` Heavy vehicle (3/4 axles) = verkeerdevlei toll plaza
		ANS (T)	r117.00
		ANS (P)	None
		ERROR	None
5	250	NL	How many millions of U.S. viewers watched the episode "Buzzkill"?
		TBL	“No. in series”, “No. in season”, “Title”, “Directed by”, “Written by”, “Original air date”, “U.S. viewers (millions)”
		SQL (T)	`SELECT` count(U.S. viewers (millions)) `FROM` 1-12570759-2 `WHERE` Title = "Buzzkill"
		SQL (P)	`SELECT` (U.S. viewers (millions)) `FROM` 1-12570759-2 `WHERE` Title = "buzzkill"
		ANS (T)	1
		ANS (P)	13.13
		ERROR	Ground Truth

6	261	NL	Name the perfect stem for jo
		TBL	“Perfect stem”, “Future stem”, “Imperfect stem”, “Short stem”, “Meaning”
		SQL (T)	`SELECT` count(Perfect stem) `FROM` 1-12784134-24 `WHERE` Short stem = jo
		SQL (P)	`SELECT` (Perfect stem) `FROM` 1-12784134-24 `WHERE` Imperfect stem = jo
		ANS (T)	1
		ANS (P)	None
		ERROR	Qestion (I)
7	332	NL	How many incumbents come from alvin bush’s district?
		TBL	“District”, “Incumbent”, “Party”, “First elected”, “Result”, “Candidates”
		SQL (T)	`SELECT` count(Candidates) `FROM` 1-1341930-38 `WHERE` Incumbent = Alvin Bush
		SQL (P)	`SELECT` count(Incumbent) `FROM` 1-1341930-38 `WHERE` District = alvin bush
		ANS (T)	1
		ANS (P)	0
		ERROR	Qestion (III)
8	475	NL	Name the finished for kerry katona
		TBL	“Celebrity”, “Famous for”, “Entered”, “Exited”, “Finished”
		SQL (T)	`SELECT` count(Finished) `FROM` 1-14345690-4 `WHERE` Celebrity = Kerry Katona
		SQL (P)	`SELECT` (Finished) `FROM` 1-14345690-4 `WHERE` Celebrity = kerry katona
		ANS (T)	1
		ANS (P)	1st
		ERROR	Ground Truth
9	557	NL	Name the english gloss for haŋȟ’áŋna
		TBL	“English gloss”, “Santee-Sisseton”, “Yankton-Yanktonai”, “Northern Lakota”, “Southern Lakota”
		SQL (T)	`SELECT` (English gloss) `FROM` 1-1499774-5 `WHERE` Santee-Sisseton = haŋȟ’áŋna
		SQL (P)	`SELECT` (English gloss) `FROM` 1-1499774-5 `WHERE` Southern Lakota = haŋȟ’áŋna
		ANS (T)	morning
		ANS (P)	None
		ERROR	Qestion (I)
10	597	NL	Name the year for sammo hung for ip man 2
		TBL	“Year”, “Best Film”, “Best Director”, “Best Actor”, “Best Actress”, “Best Supporting Actor”, “Best Supporting Actress”
		SQL (T)	`SELECT` (Year) `FROM` 1-15301258-1 `WHERE` Best Supporting Actor = Sammo Hung for Ip Man 2
		SQL (P)	`SELECT` (Year) `FROM` 1-15301258-1 `WHERE` Best Actor = sammo hung
		ANS (T)	2011 5th
		ANS (P)	None
		ERROR	Qestion (I)
11	598	NL	Name the best supporting actress for sun honglei for mongol
		TBL	“Year”, “Best Film”, “Best Director”, “Best Actor”, “Best Actress”, “Best Supporting Actor”, “Best Supporting Actress”
		SQL (T)	`SELECT` (Best Supporting Actress) `FROM` 1-15301258-1 `WHERE` Best Supporting Actor = Sun Honglei for Mongol

		SQL (P)	SELECT (Best Supporting Actress) FROM 1-15301258-1 WHERE Best Film = mongol AND Best Actor = sun honglei
		ANS (T)	joan chen for the sun also rises
		ANS (P)	None
		ERROR	Qestion (I)
12	625	NL	What is the sexual abuse rate where the conflict is the Burundi Civil War?
		TBL	“Conflict”, “United Nations Mission”, “Sexual abuse 1”, “Murder 2”, “Extortion/Theft 3”
		SQL (T)	SELECT min(Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = Burundi Civil War
		SQL (P)	SELECT (Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = burundi civil war
		ANS (T)	80.0
		ANS (P)	80.0
		ERROR	Ground Truth
13	627	NL	What is the sexual abuse rate where the conflict is the Second Sudanese Civil War?
		TBL	“Conflict”, “United Nations Mission”, “Sexual abuse 1”, “Murder 2”, “Extortion/Theft 3”
		SQL (T)	SELECT min(Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = Second Sudanese Civil War
		SQL (P)	SELECT (Sexual abuse 1) FROM 1-15652027-1 WHERE Conflict = second sudanese civil war
		ANS (T)	400.0
		ANS (P)	400.0
		ERROR	Ground Truth
14	654	NL	What is the total population in the city/town of Arendal?
		TBL	“City/town”, “Municipality”, “County”, “City/town status”, “Population”
		SQL (T)	SELECT count(Population) FROM 1-157826-1 WHERE City/town = Arendal
		SQL (P)	SELECT sum(Population) FROM 1-157826-1 WHERE City/town = arendal
		ANS (T)	1
		ANS (P)	39826.0
		ERROR	Ground Truth
15	738	NL	Name the location for illinois
		TBL	“Date”, “Time”, “ACC Team”, “Big Ten Team”, “Location”, “Television”, “Attendance”, “Winner”, “Challenge Leader”
		SQL (T)	SELECT (Location) FROM 1-1672976-7 WHERE Big Ten Team = Illinois
		SQL (P)	SELECT (Location) FROM 1-1672976-7 WHERE ACC Team = illinois
		ANS (T)	littlejohn coliseum • clemson, sc
		ANS (P)	None
		ERROR	Qestion (I)
16	783	NL	How many times was Plan B 4th place?
		TBL	“Poll Year”, “Winner”, “Second”, “Third”, “Fourth”, “Fifth”, “Sixth”, “Seventh”, “Eighth”, “Ninth”, “Tenth”
		SQL (T)	SELECT count(Winner) FROM 1-17111812-1 WHERE Fourth = Plan B
		SQL (P)	SELECT count(Ninth) FROM 1-17111812-1 WHERE Fourth = plan b
		ANS (T)	1
		ANS (P)	1

		ERROR	Qestion (II)
17	795	NL	If the equation is (10 times 8) + 4, what would be the 2nd throw?
		TBL	“1st throw”, “2nd throw”, “3rd throw”, “Equation”, “Result”
		SQL (T)	`SELECT` max(2nd throw) `FROM` 1-17265535-6 `WHERE` Equation = (10 times 8) + 4
		SQL (P)	`SELECT` (2nd throw) `FROM` 1-17265535-6 `WHERE` Equation = (10 times 8) + 4
		ANS (T)	4.0
		ANS (P)	4.0
		ERROR	Ground Truth
18	841	NL	When there was a bye in the round of 32, what was the result in the round of 16?
		TBL	“Athlete”, “Event”, “Round of 32”, “Round of 16”, “Quarterfinals”, “Semifinals”
		SQL (T)	`SELECT` (Semifinals) `FROM` 1-1745820-5 `WHERE` Round of 32 = Bye
		SQL (P)	`SELECT` (Round of 16) `FROM` 1-1745820-5 `WHERE` Round of 32 = bye
		ANS (T)	did not advance
		ANS (P)	simelane (swz) w (rsc)
		ERROR	Ground Truth
19	912	NL	How many lines have the segment description of red line mos-2 west?
		TBL	“Segment description”, “Date opened”, “Line(s)”, “Endpoints”, “# of new stations”, “Length (miles)”
		SQL (T)	`SELECT` (Line(s)) `FROM` 1-1817879-2 `WHERE` Segment description = Red Line MOS-2 West
		SQL (P)	`SELECT` count(Line(s)) `FROM` 1-1817879-2 `WHERE` Segment description = red line mos-2 west
		ANS (T)	red, purple 1
		ANS (P)	1
		ERROR	Ground Truth
20	1035	NL	Name the number of candidates for # of seats won being 43
		TBL	“Election”, “Leader”, “# of candidates”, “# of seats to be won”, “# of seats won”, “# of total votes”, “% of popular vote”
		SQL (T)	`SELECT` (# of candidates) `FROM` 1-19283982-4 `WHERE` # of seats won = 43
		SQL (P)	`SELECT` count(# of candidates) `FROM` 1-19283982-4 `WHERE` # of seats won = 43
		ANS (T)	295.0
		ANS (P)	1
		ERROR	None
21	1056	NL	When the total score is 740, what is tromso?
		TBL	“Song”, “Porsgrunn”, “Bergen”, “Bodø”, “Stavanger”, “Ålesund”, “Elverum”, “Tromsø”, “Fredrikstad”, “Trondheim”, “Oslo”, “Total”
		SQL (T)	`SELECT` min(Tromsø) `FROM` 1-19439864-2 `WHERE` Total = 740
		SQL (P)	`SELECT` (Tromsø) `FROM` 1-19439864-2 `WHERE` Total = 740
		ANS (T)	70.0
		ANS (P)	70.0
		ERROR	Ground Truth
22	1089	NL	Name the total number of date for l 63-77
		TBL	“Game”, “Date”, “Opponent”, “Score”, “High points”, “High rebounds”, “High assists”, “Location”, “Record”

		SQL (T)	`SELECT` count(Date) `FROM` 1-19789597-5 `WHERE` Score = L 63-77
		SQL (P)	`SELECT` count(Date) `FROM` 1-19789597-5 `WHERE` Record = l 63-77
		ANS (T)	1
		ANS (P)	0
		ERROR	Qestion (I)
23	1122	NL	What was the res for the game against Payam?
		TBL	“Date”, “Team #1”, “Res.”, “Team #2”, “Competition”, “Attendance”, “Remarks”
		SQL (T)	`SELECT` (Res.) `FROM` 1-2015453-1 `WHERE` Team #2 = Payam
		SQL (P)	`SELECT` (Res.) `FROM` 1-2015453-1 `WHERE` Team #1 = payam
		ANS (T)	1–1
		ANS (P)	None
		ERROR	Qestion (I)
24	1199	NL	what is the No in series when Rob wright & Debra j. Fisher & Erica Messer were the writers?
		TBL	“No. in series”, “No. in season”, “Title”, “Directed by”, “Written by”, “Original air date”, “Production code”, “U.S. viewers (millions)”
		SQL (T)	`SELECT` min(No. in series) `FROM` 1-21313327-1 `WHERE` Written by = Rob Wright & Debra J. Fisher & Erica Messer
		SQL (P)	`SELECT` (No. in series) `FROM` 1-21313327-1 `WHERE` Written by = rob wright & debra j. fisher & erica messer
		ANS (T)	149.0
		ANS (P)	149.0
		ERROR	Ground Truth
25	1263	NL	What episode had 10.14 million viewers (U.S.)?
		TBL	“No.”, “#”, “Title”, “Directed by”, “Written by”, “U.S. viewers (million)”, “Original air date”, “Production code”
		SQL (T)	`SELECT` min(#) `FROM` 1-21550897-1 `WHERE` U.S. viewers (million) = 10.14
		SQL (P)	`SELECT` (Title) `FROM` 1-21550897-1 `WHERE` U.S. viewers (million) = 10.14
		ANS (T)	11.0
		ANS (P)	" arrow of time "
		ERROR	Ground Truth
26	1430	NL	Name the surface for philadelphia
		TBL	“Outcome”, “Year”, “Championship”, “Surface”, “Opponent in the final”, “Score in the final”
		SQL (T)	`SELECT` (Surface) `FROM` 1-23235767-4 `WHERE` Championship = Philadelphia
		SQL (P)	`SELECT` (Surface) `FROM` 1-23235767-4 `WHERE` Opponent in the final = philadelphia
		ANS (T)	carpet
		ANS (P)	None
		ERROR	Qestion (I)
27	1449	NL	How many games had been played when the Mavericks had a 46-22 record?
		TBL	“Game”, “Date”, “Team”, “Score”, “High points”, “High rebounds”, “High assists”, “Location Attendance”, “Record”
		SQL (T)	`SELECT` max(Game) `FROM` 1-23284271-9 `WHERE` Record = 46-22
		SQL (P)	`SELECT` (Game) `FROM` 1-23284271-9 `WHERE` Record = 46-22
		ANS (T)	68.0

		ANS (P)	68.0
		ERROR	Ground Truth
28	1591	NL	What was the rating for Brisbane the week that Adelaide had 94000?
		TBL	“WEEK”, “Sydney”, “Melbourne”, “Brisbane”, “Adelaide”, “Perth”, “TOTAL”, “NIGHTLY RANK”
		SQL (T)	`SELECT` min(Brisbane) `FROM` 1-24291077-8 `WHERE` Adelaide = 94000
		SQL (P)	`SELECT` (Brisbane) `FROM` 1-24291077-8 `WHERE` Adelaide = 94000
		ANS (T)	134000.0
		ANS (P)	134000.0
		ERROR	Ground Truth
29	1851	NL	Which year did enrolled Gambier members leave?
		TBL	“Institution”, “Location (all in Ohio)”, “Nickname”, “Founded”, “Type”, “Enrollment”, “Joined”, “Left”, “Current Conference”
		SQL (T)	`SELECT` min(Left) `FROM` 1-261946-3 `WHERE` Location (all in Ohio) = Gambier
		SQL (P)	`SELECT` (Left) `FROM` 1-261946-3 `WHERE` Nickname = gambier
		ANS (T)	1984.0
		ANS (P)	None
		ERROR	Ground Truth
30	1978	NL	How many games were played where the height of the player is 1.92m?
		TBL	“Player”, “Position”, “Starting No.#”, “D.O.B”, “Club”, “Height”, “Weight”, “Games”
		SQL (T)	`SELECT` count(Games) `FROM` 1-26847237-2 `WHERE` Height = 1.92m
		SQL (P)	`SELECT` (Games) `FROM` 1-26847237-2 `WHERE` Height = 1.92m
		ANS (T)	1
		ANS (P)	7.0
		ERROR	Ground Truth
31	1986	NL	What was the score between Marseille and Manchester United on the second leg of the Champions League Round of 16?
		TBL	“Team”, “Contest and round”, “Opponent”, “1st leg score*”, “2nd leg score**”, “Aggregate score”
		SQL (T)	`SELECT` (2nd leg score**) `FROM` 1-26910311-8 `WHERE` Opponent = Marseille
		SQL (P)	`SELECT` (2nd leg score**) `FROM` 1-26910311-8 `WHERE` Team = marseille `AND` Contest and round = champions league round of 16 `AND` Opponent = manchester united
		ANS (T)	2–1 (h)
		ANS (P)	None
		ERROR	Qestion (I)
32	2090	NL	What year was mcmahon stadium founded?
		TBL	“Institution”, “Team”, “City”, “Province”, “Founded”, “Affiliation”, “Enrollment”, “Endowment”, “Football stadium”, “Capacity”
		SQL (T)	`SELECT` max(Founded) `FROM` 1-27599216-6 `WHERE` Football stadium = McMahon Stadium
		SQL (P)	`SELECT` (Founded) `FROM` 1-27599216-6 `WHERE` Football stadium = mcmahon stadium
		ANS (T)	1966.0
		ANS (P)	1966.0
		ERROR	Ground Truth

33	2159	NL	Which game had a score of w 95-85?
		TBL	“Game”, “Date”, “Team”, “Score”, “High points”, “High rebounds”, “High assists”, “Location Attendance”, “Record”
		SQL (T)	`SELECT` min(Game) `FROM` 1-27902171-7 `WHERE` Score = W 95-85
		SQL (P)	`SELECT` (Game) `FROM` 1-27902171-7 `WHERE` Score = w 95-85
		ANS (T)	48.0
		ANS (P)	48.0
		ERROR	Ground Truth
34	2175	NL	How many times has Ma Long won the men’s singles?
		TBL	“Year Location”, “Mens Singles”, “Womens Singles”, “Mens Doubles”, “Womens Doubles”
		SQL (T)	`SELECT` count(Mens Doubles) `FROM` 1-28138035-33 `WHERE` Mens Singles = Ma Long
		SQL (P)	`SELECT` count(Womens Doubles) `FROM` 1-28138035-33 `WHERE` Mens Singles = ma long
		ANS (T)	1
		ANS (P)	1
		ERROR	Qestion (II)
35	2223	NL	What daft pick number is the player coming from Regina Pats (WHL)?
		TBL	“Pick #”, “Player”, “Position”, “Nationality”, “NHL team”, “College/junior/club team”
		SQL (T)	`SELECT` (Pick #) `FROM` 1-2850912-1 `WHERE` College/junior/club team = Regina Pats (WHL)
		SQL (P)	`SELECT` min(Pick #) `FROM` 1-2850912-1 `WHERE` College/junior/club team = regina pats (whl)
		ANS (T)	21.0
		ANS (P)	21.0
		ERROR	None
36	2286	NL	What is the area when the Iga name is Ahoada East?
		TBL	“LGA Name”, “Area (km 2)”, “Census 2006 population”, “Administrative capital”, “Postal Code”
		SQL (T)	`SELECT` max(Area (km 2)) `FROM` 1-28891101-3 `WHERE` LGA Name = Ahoada East
		SQL (P)	`SELECT` (Area (km 2)) `FROM` 1-28891101-3 `WHERE` LGA Name = ahoada east
		ANS (T)	341.0
		ANS (P)	341.0
		ERROR	Ground Truth
37	2311	NL	What is the train number when the time is 10:38?
		TBL	“Sl. No.”, “Train number”, “Train name”, “Origin”, “Destination”, “Time”, “Service”, “Route/Via.”
		SQL (T)	`SELECT` max(Train number) `FROM` 1-29202276-2 `WHERE` Time = 10:38
		SQL (P)	`SELECT` (Train number) `FROM` 1-29202276-2 `WHERE` Time = 10:38
		ANS (T)	16381.0
		ANS (P)	16381.0
		ERROR	Ground Truth
38	2323	NL	What team hired Renato Gaúcho?

		TBL	“Team”, “Outgoing manager”, “Manner of departure”, “Date of vacancy”, “Position in table”, “Replaced by”, “Date of appointment”
		SQL (T)	`SELECT` (Team) `FROM` 1-29414946-3 `WHERE` Replaced by = Renato Gaúcho
		SQL (P)	`SELECT` (Team) `FROM` 1-29414946-3 `WHERE` Outgoing manager = renato gaúcho
		ANS (T)	atlético paranaense
		ANS (P)	grêmio
		ERROR	None
39	2565	NL	What was attendance of the whole season when the average attendance for League Cup was 32,415?
		TBL	“Season”, “Season Total Att.”, “K-League Season Total Att.”, “Regular Season Average Att.”, “League Cup Average Att.”, “FA Cup Total / Average Att.”, “ACL Total / Average Att.”, “Friendly Match Att.”
		SQL (T)	`SELECT` (Season Total Att.) `FROM` 2-1056336-11 `WHERE` League Cup Average Att. = 32,415
		SQL (P)	`SELECT` (Season) `FROM` 2-1056336-11 `WHERE` League Cup Average Att. = 32,415
		ANS (T)	458,605
		ANS (P)	2005
		ERROR	None
40	2812	NL	What is the name of the driver with a rotax max engine, in the rotax heavy class, with arrow as chassis and on the TWR Raceline Seating team?
		TBL	“Team”, “Class”, “Chassis”, “Engine”, “Driver”
		SQL (T)	`SELECT` (Driver) `FROM` 2-15162596-2 `WHERE` Engine = rotax max `AND` Class = rotax heavy `AND` Chassis = arrow `AND` Team = twr raceline seating
		SQL (P)	`SELECT` (Driver) `FROM` 2-15162596-2 `WHERE` Team = twr raceline seating `AND` Class = rotax heavy `AND` Chassis = arrow `AND` Engine = rotax max engine, in the rotax heavy
		ANS (T)	rod clarke
		ANS (P)	None
		ERROR	None
41	2891	NL	If played is 22 and the tries against are 43, what are the points?
		TBL	“Club”, “Played”, “Drawn”, “Lost”, “Points for”, “Points against”, “Tries for”, “Tries against”, “Try bonus”, “Losing bonus”, “Points”
		SQL (T)	`SELECT` (Points for) `FROM` 2-13741576-4 `WHERE` Played = 22 `AND` Tries against = 43
		SQL (P)	`SELECT` (Points) `FROM` 2-13741576-4 `WHERE` Played = 22 `AND` Tries against = 43
		ANS (T)	353
		ANS (P)	46
		ERROR	Qestion (I)
42	2925	NL	What was the first Round with a Pick # greater than 1 and 140 Overall?
		TBL	“Round”, “Pick #”, “Overall”, “Name”, “Position”, “College”
		SQL (T)	`SELECT` min(Round) `FROM` 2-15198842-23 `WHERE` Pick # > 1 `AND` Overall > 140
		SQL (P)	`SELECT` min(Round) `FROM` 2-15198842-23 `WHERE` Pick # > 1 `AND` Overall = 140
		ANS (T)	None
		ANS (P)	6.0

		ERROR	Ground Truth
43	3028	NL	What was the attendance of the game that had an away team of FK Mogren?
		TBL	“Venue”, “Home”, “Guest”, “Score”, “Attendance”
		SQL (T)	SELECT (Attendance) FROM 2-13883437-1 WHERE Guest = fk mogren
		SQL (P)	SELECT (Attendance) FROM 2-13883437-1 WHERE Home = away
		ANS (T)	1.2
		ANS (P)	None
		ERROR	None
44	3050	NL	Which team is in the Southeast with a home at Philips Arena?
		TBL	“Conference”, “Division”, “Team”, “City”, “Home Arena”
		SQL (T)	SELECT (Team) FROM 2-14519555-8 WHERE Division = southeast AND Home Arena = philips arena
		SQL (P)	SELECT (Team) FROM 2-14519555-8 WHERE Conference = southeast AND Home Arena = philips arena
		ANS (T)	atlanta hawks
		ANS (P)	None
		ERROR	Qestion (I)
45	3314	NL	What is the lowest number of bronze a short track athlete with 0 gold medals has?
		TBL	“Athlete”, “Sport”, “Type”, “Olympics”, “Gold”, “Silver”, “Bronze”, “Total”
		SQL (T)	SELECT min(Bronze) FROM 2-13554889-6 WHERE Sport = short track AND Gold = 0
		SQL (P)	SELECT min(Bronze) FROM 2-13554889-6 WHERE Type = short track AND Gold < 0
		ANS (T)	2.0
		ANS (P)	None
		ERROR	None
46	3328	NL	What was the total in 2009 for years of river vessels when 2008 was more than 8,030 and 2007 was more than 1,411,414?
		TBL	“Years”, “2004”, “2005”, “2006”, “2007”, “2008”, “2009”, “2010”, “2011”
		SQL (T)	SELECT count(2009) FROM 2-13823555-1 WHERE 2007 > 1,411,414 AND Years = river vessels AND 2008 > 8,030
		SQL (P)	SELECT sum(2009) FROM 2-13823555-1 WHERE Years = river vessels AND 2007 > 1,411,414 AND 2008 > 8,030
		ANS (T)	1
		ANS (P)	6.0
		ERROR	Ground Truth
47	3370	NL	When did Hans Hartmann drive?
		TBL	“Year”, “Event”, “Venue”, “Driver”, “Result”, “Category”, “Report”
		SQL (T)	SELECT count(Year) FROM 2-14287417-3 WHERE Driver = hans hartmann
		SQL (P)	SELECT (Year) FROM 2-14287417-3 WHERE Driver = hans hartmann
		ANS (T)	1
		ANS (P)	1939.0
		ERROR	Ground Truth
48	3499	NL	Which driver has less than 37 wins and at 14.12%?
		TBL	“Driver”, “Seasons”, “Entries”, “Wins”, “Percentage”

		SQL (T)	SELECT avg(Entries) FROM 2-13599687-6 WHERE Wins < 37 AND Percentage = 14.12%
		SQL (P)	SELECT (Driver) FROM 2-13599687-6 WHERE Wins < 37 AND Percentage = 14.12%
		ANS (T)	177.0
		ANS (P)	niki lauda
		ERROR	Ground Truth
49	3602	NL	In what Year did the German Open have Yoo Sang-Hee as Partner?
		TBL	“Outcome”, “Event”, “Year”, “Venue”, “Partner”
		SQL (T)	SELECT (Year) FROM 2-14895591-2 WHERE Partner = yoo sang-hee AND Venue = german open
		SQL (P)	SELECT (Year) FROM 2-14895591-2 WHERE Event = german open AND Partner = yoo sang-hee
		ANS (T)	1986
		ANS (P)	None
		ERROR	Qestion (I)
50	3650	NL	How many Byes have Against of 1076 and Wins smaller than 13?
		TBL	“Ballarat FL”, “Wins”, “Byes”, “Losses”, “Draws”, “Against”
		SQL (T)	SELECT avg(Byes) FROM 2-1552908-21 WHERE Against = 1076 AND Wins < 13
		SQL (P)	SELECT count(Byes) FROM 2-1552908-21 WHERE Wins < 13 AND Against = 1076
		ANS (T)	None
		ANS (P)	0
		ERROR	Ground Truth
51	3728	NL	How many 2007’s have a 2000 greater than 56,6, 23,2 as 2006, and a 1998 greater than 61,1?
		TBL	“Capital/Region”, “1997”, “1998”, “1999”, “2000”, “2001”, “2002”, “2003”, “2004”, “2005”, “2006”, “2007”
		SQL (T)	SELECT sum(2007) FROM 2-15348345-1 WHERE 2000 > 56,6 AND 2006 = 23,2 AND 1998 > 61,1
		SQL (P)	SELECT count(2007) FROM 2-15348345-1 WHERE 1998 > 61,1 AND 2000 > 56,6 AND 2006 = 23,2
		ANS (T)	None
		ANS (P)	0
		ERROR	None
52	3730	NL	What is the average 2000 that has a 1997 greater than 34,6, a 2006 greater than 38,7, and a 2998 less than 76?
		TBL	“Capital/Region”, “1997”, “1998”, “1999”, “2000”, “2001”, “2002”, “2003”, “2004”, “2005”, “2006”, “2007”
		SQL (T)	SELECT avg(2000) FROM 2-15348345-1 WHERE 1997 > 34,6 AND 2006 > 38,7 AND 1998 < 76
		SQL (P)	SELECT avg(2000) FROM 2-15348345-1 WHERE 1997 > 34,6 AND 1998 > 34,6 AND 2006 > 38,7
		ANS (T)	40.0
		ANS (P)	35.416666666666664
		ERROR	None
53	3770	NL	What is the total number of offensive rebounds for players with under 65 total rebounds, 5 defensive rebounds, and under 7 assists?

		TBL	“Player”, “FG Pct”, “3FGA”, “3FGM”, “3FG Pct”, “FT Pct”, “Off Reb”, “Def Reb”, “Total Reb”, “Asst”
		SQL (T)	SELECT count(Off Reb) FROM 2-15746812-4 WHERE Total Reb $<$ 65 AND Def Reb $=$ 5 AND Asst $<$ 7
		SQL (P)	SELECT count(Asst) FROM 2-15746812-4 WHERE Off Reb $=$ 5 defensive rebounds AND Total Reb $<$ 65 AND Asst $<$ 7
		ANS (T)	0
		ANS (P)	0
		ERROR	Ground Truth
54	3890	NL	What is the rank of the reynard 2ki chassis before 2002?
		TBL	“Year”, “Team”, “Chassis”, “Engine”, “Rank”, “Points”
		SQL (T)	SELECT (Rank) FROM 2-1615758-2 WHERE Year $<$ 2002 AND Chassis $=$ reynard 2ki
		SQL (P)	SELECT sum(Rank) FROM 2-1615758-2 WHERE Year $<$ 2002 AND Chassis $=$ reynard 2ki
		ANS (T)	19th
		ANS (P)	19.0
		ERROR	None
55	3925	NL	What is the value for the item "Tries" when the value of the item "Played" is 18 and the value of the item "Points" is 375?
		TBL	“Club”, “Played”, “Drawn”, “Lost”, “Points for”, “Points against”, “Points difference”, “Tries For”, “Tries Against”
		SQL (T)	SELECT (Tries For) FROM 2-15467476-4 WHERE Played $=$ 18 AND Points against $=$ 375
		SQL (P)	SELECT (Tries Against) FROM 2-15467476-4 WHERE Played $=$ 18 AND Points for $=$ 375
		ANS (T)	54
		ANS (P)	None
		ERROR	Qestion (I)
56	3992	NL	What is the average number of gold medals when the total was 1335 medals, with more than 469 bronzes and more than 14 silvers?
		TBL	“Rank”, “Gold”, “Silver”, “Bronze”, “Total”
		SQL (T)	SELECT avg(Gold) FROM 2-15428689-2 WHERE Silver $>$ 14 AND Total $=$ 1335 AND Bronze $>$ 469
		SQL (P)	SELECT avg(Gold) FROM 2-15428689-2 WHERE Silver $>$ 14 AND Bronze $>$ 469 AND Total $=$ 1335 medals
		ANS (T)	None
		ANS (P)	None
		ERROR	None
57	4229	NL	What venue had an event on 17 November 1963?
		TBL	“Season”, “Date”, “Winner”, “Score [C]”, “Venue”, “Competition round”
		SQL (T)	SELECT (Venue) FROM 2-17299309-4 WHERE Season $=$ 1963 AND Date $=$ 17 november 1963
		SQL (P)	SELECT (Venue) FROM 2-17299309-4 WHERE Date $=$ 17 november 1963
		ANS (T)	estadio nacional
		ANS (P)	estadio nacional
		ERROR	Qestion (II)

58	4381	NL	WHAT IS THE WEEK WITH AN ATTENDANCE OF 75,555?
		TBL	“Week”, “Date”, “Opponent”, “Result”, “TV Time”, “Attendance”
		SQL (T)	`SELECT` sum(Week) `FROM` 2-16764708-1 `WHERE` Attendance = 75,555
		SQL (P)	`SELECT` (Week) `FROM` 2-16764708-1 `WHERE` Attendance = 75,555
		ANS (T)	11.0
		ANS (P)	11.0
		ERROR	Ground Truth
59	4522	NL	How many total golds do teams have when the total medals is less than 1?
		TBL	“Rank”, “Nation”, “Gold”, “Silver”, “Bronze”, “Total”
		SQL (T)	`SELECT` sum(Gold) `FROM` 2-16340209-1 `WHERE` Total < 1
		SQL (P)	`SELECT` count(Gold) `FROM` 2-16340209-1 `WHERE` Total < 1
		ANS (T)	None
		ANS (P)	0
		ERROR	None
60	4561	NL	How much Overall has a Name of bob anderson?
		TBL	“Round”, “Pick”, “Overall”, “Name”, “Position”, “College”
		SQL (T)	`SELECT` count(Overall) `FROM` 2-17100961-17 `WHERE` Name = bob anderson
		SQL (P)	`SELECT` sum(Overall) `FROM` 2-17100961-17 `WHERE` Name = bob anderson
		ANS (T)	1
		ANS (P)	68.0
		ERROR	Ground Truth
61	4785	NL	What is the name of the free transfer fee with a transfer status and an ENG country?
		TBL	“Name”, “Country”, “Status”, “Transfer window”, “Transfer fee”
		SQL (T)	`SELECT` (Name) `FROM` 2-16549823-7 `WHERE` Transfer fee = free `AND` Status = transfer `AND` Country = eng
		SQL (P)	`SELECT` (Name) `FROM` 2-16549823-7 `WHERE` Country = eng `AND` Status = `AND` Transfer fee = free transfer
		ANS (T)	bailey
		ANS (P)	None
		ERROR	None
62	5055	NL	What is the To par of Player Andy North with a Total larger than 153?
		TBL	“Player”, “Country”, “Year(s) won”, “Total”, “To par”
		SQL (T)	`SELECT` count(To par) `FROM` 2-17162255-3 `WHERE` Player = andy north `AND` Total > 153
		SQL (P)	`SELECT` (To par) `FROM` 2-17162255-3 `WHERE` Player = andy north `AND` Total > 153
		ANS (T)	0
		ANS (P)	None
		ERROR	Ground Truth
63	5304	NL	What is the total avg/g of McCrary, Greg?
		TBL	“Name”, “GP-GS”, “Effic”, “Cmp-Att-Int”, “Avg/G”
		SQL (T)	`SELECT` count(Avg/G) `FROM` 2-16981858-6 `WHERE` Name = mccrary, greg
		SQL (P)	`SELECT` sum(Avg/G) `FROM` 2-16981858-6 `WHERE` Name = mccrary, greg
		ANS (T)	1
		ANS (P)	58.9

		ERROR	Ground Truth
64	5456	NL	What year has a Schwante smaller than 2.043, an Eichstädt smaller than 848, and a Bärenklau smaller than 1.262?
		TBL	“Year”, “Bötzow”, “Schwante”, “Vehlefanz”, “Neu-Vehlefanz”, “Marwitz”, “Bärenklau”, “Eichstädt”
		SQL (T)	`SELECT` count(Year) `FROM` 2-11680175-1 `WHERE` Schwante $<$ 2.043 `AND` Eichstädt $<$ 848 `AND` Bärenklau $<$ 1.262
		SQL (P)	`SELECT` sum(Year) `FROM` 2-11680175-1 `WHERE` Schwante $<$ 2.043 `AND` Bärenklau $<$ 1.262 `AND` Eichstädt $<$ 848
		ANS (T)	0
		ANS (P)	None
		ERROR	Ground Truth
65	5611	NL	Who was home at Princes Park?
		TBL	“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”
		SQL (T)	`SELECT` (Home team score) `FROM` 2-10809157-18 `WHERE` Venue = princes park
		SQL (P)	`SELECT` (Home team) `FROM` 2-10809157-18 `WHERE` Venue = princes park
		ANS (T)	9.16 (70)
		ANS (P)	fitzroy
		ERROR	Ground Truth
66	5705	NL	What is the grid for the Minardi Team USA with laps smaller than 90?
		TBL	“Driver”, “Team”, “Laps”, “Time/Retired”, “Grid”, “Points”
		SQL (T)	`SELECT` (Grid) `FROM` 2-10823048-3 `WHERE` Laps $<$ 90 `AND` Team = minardi team usa
		SQL (P)	`SELECT` sum(Grid) `FROM` 2-10823048-3 `WHERE` Team = minardi team usa `AND` Laps $<$ 90
		ANS (T)	12.0
		ANS (P)	12.0
		ERROR	None
67	5707	NL	What is Fitzroy’s Home team Crowd?
		TBL	“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”
		SQL (T)	`SELECT` sum(Crowd) `FROM` 2-10809142-16 `WHERE` Home team = fitzroy
		SQL (P)	`SELECT` (Crowd) `FROM` 2-10809142-16 `WHERE` Home team = fitzroy
		ANS (T)	20.0
		ANS (P)	20,000
		ERROR	Ground Truth
68	5746	NL	How many goals were scored on 21 Junio 2008?
		TBL	“Goal”, “Date”, “Venue”, “Result”, “Competition”
		SQL (T)	`SELECT` count(Goal) `FROM` 2-1192553-1 `WHERE` Date = 21 junio 2008
		SQL (P)	`SELECT` (Goal) `FROM` 2-1192553-1 `WHERE` Date = 21 junio 2008
		ANS (T)	1
		ANS (P)	13.0
		ERROR	Ground Truth
69	5882	NL	What is the average year of the Fantasia Section Award?
		TBL	“Festival”, “Year”, “Result”, “Award”, “Category”

		SQL (T)	SELECT avg(Year) FROM 2-1201864-1 WHERE Award = fantasia section award
		SQL (P)	SELECT avg(Year) FROM 2-1201864-1 WHERE Award = fantasia section
		ANS (T)	1999.0
		ANS (P)	None
		ERROR	None
70	5893	NL	Name the team for launceston
		TBL	“Race Title”, “Circuit”, “City / State”, “Date”, “Winner”, “Team”
		SQL (T)	SELECT (Team) FROM 2-11880375-2 WHERE Race Title = launceston
		SQL (P)	SELECT (Team) FROM 2-11880375-2 WHERE City / State = launceston
		ANS (T)	shell ultra-hi racing
		ANS (P)	None
		ERROR	Qestion (I)
71	6028	NL	What position does the player from arkansas play?
		TBL	“Player”, “Pos.”, “From”, “School/Country”, “Rebs”, “Asts”
		SQL (T)	SELECT (Pos.) FROM 2-11482079-13 WHERE School/Country = arkansas
		SQL (P)	SELECT (Pos.) FROM 2-11482079-13 WHERE From = arkansas
		ANS (T)	c
		ANS (P)	None
		ERROR	Qestion (I)
72	6089	NL	What are the draws when wins are fwewer than 9 and byes fewer than 2?
		TBL	“Tallangatta DFL”, “Wins”, “Byes”, “Losses”, “Draws”, “Against”
		SQL (T)	SELECT count(Draws) FROM 2-11338646-3 WHERE Wins < 9 AND Byes < 2
		SQL (P)	SELECT avg(Draws) FROM 2-11338646-3 WHERE Wins < 9 AND Byes < 2
		ANS (T)	0
		ANS (P)	None
		ERROR	Ground Truth
73	6194	NL	How many attended the game at Arden Street Oval?
		TBL	“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”
		SQL (T)	SELECT avg(Crowd) FROM 2-10806592-7 WHERE Venue = arden street oval
		SQL (P)	SELECT (Crowd) FROM 2-10806592-7 WHERE Venue = arden street oval
		ANS (T)	15.0
		ANS (P)	15,000
		ERROR	Ground Truth
74	6224	NL	On January 29, who had the decision of Mason?
		TBL	“Date”, “Visitor”, “Score”, “Home”, “Decision”, “Attendance”, “Record”
		SQL (T)	SELECT (Visitor) FROM 2-11756731-6 WHERE Decision = mason AND Date = january 29
		SQL (P)	SELECT (Decision) FROM 2-11756731-6 WHERE Date = january 29 AND Decision = mason
		ANS (T)	nashville
		ANS (P)	mason
		ERROR	None
75	6392	NL	What is the grid number with less than 52 laps and a Time/Retired of collision, and a Constructor of arrows - supertec?

		TBL	“Driver”, “Constructor”, “Laps”, “Time/Retired”, “Grid”
		SQL (T)	`SELECT` count(Grid) `FROM` 2-1123405-2 `WHERE` Laps < 52 `AND` Time/Retired = collision `AND` Constructor = arrows - supertec
		SQL (P)	`SELECT` (Grid) `FROM` 2-1123405-2 `WHERE` Constructor = arrows `AND` Laps < 52 `AND` Time/Retired = collision
		ANS (T)	1
		ANS (P)	None
		ERROR	Ground Truth
76	6439	NL	In the match where the home team scored 14.20 (104), how many attendees were in the crowd?
		TBL	“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”
		SQL (T)	`SELECT` sum(Crowd) `FROM` 2-10790397-5 `WHERE` Home team score = 14.20 (104)
		SQL (P)	`SELECT` (Crowd) `FROM` 2-10790397-5 `WHERE` Home team score = 14.20 (104)
		ANS (T)	25.0
		ANS (P)	25,000
		ERROR	Ground Truth
77	6440	NL	In the match where the away team scored 2.7 (19), how many peopel were in the crowd?
		TBL	“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”
		SQL (T)	`SELECT` max(Crowd) `FROM` 2-10790397-5 `WHERE` Away team score = 2.7 (19)
		SQL (P)	`SELECT` (Crowd) `FROM` 2-10790397-5 `WHERE` Away team score = 2.7 (19)
		ANS (T)	15,000
		ANS (P)	15,000
		ERROR	Ground Truth
78	6533	NL	Name the Score united states of tom watson in united state?
		TBL	“Place”, “Player”, “Country”, “Score”, “To par”
		SQL (T)	`SELECT` (Score) `FROM` 2-18113463-4 `WHERE` Country = united states `AND` Player = tom watson
		SQL (P)	`SELECT` (Score) `FROM` 2-18113463-4 `WHERE` Place = united states `AND` Player = tom watson `AND` Country = united states
		ANS (T)	68.0
		ANS (P)	None
		ERROR	Qestion (I)
79	6688	NL	How many ties did he have when he had 1 penalties and more than 20 conversions?
		TBL	“Played”, “Drawn”, “Lost”, “Winning %”, “Tries”, “Conversions”, “Penalties”, “s Drop goal”, “Points total”
		SQL (T)	`SELECT` sum(Drawn) `FROM` 2-1828549-1 `WHERE` Penalties = 1 `AND` Conversions > 20
		SQL (P)	`SELECT` (Drawn) `FROM` 2-1828549-1 `WHERE` Conversions > 20 `AND` Penalties = 1
		ANS (T)	None
		ANS (P)	None
		ERROR	None
80	6729	NL	What was the year that had Anugerah Bintang Popular Berita Harian 23 as competition?

		TBL	“Year”, “Competition”, “Awards”, “Category”, “Result”
		SQL (T)	SELECT count(Year) FROM 2-17838670-5 WHERE Competition = anugerah bintang popular berita harian 23
		SQL (P)	SELECT (Year) FROM 2-17838670-5 WHERE Competition = anugerah bintang popular berita harian 23
		ANS (T)	1
		ANS (P)	2010.0
		ERROR	Ground Truth
81	6779	NL	What week was the opponent the San Diego Chargers?
		TBL	“Week”, “Date”, “Opponent”, “Result”, “Kickoff Time”, “Attendance”
		SQL (T)	SELECT avg(Week) FROM 2-17643221-2 WHERE Opponent = san diego chargers
		SQL (P)	SELECT (Week) FROM 2-17643221-2 WHERE Opponent = san diego chargers
		ANS (T)	1.0
		ANS (P)	1.0
		ERROR	Ground Truth
82	6927	NL	Which Number of electorates (2009) has a Constituency number of 46?
		TBL	“Constituency number”, “Name”, “Reserved for (SC / ST /None)”, “District”, “Number of electorates (2009)”
		SQL (T)	SELECT avg(Number of electorates (2009)) FROM 2-17922541-1 WHERE Constituency number = 46
		SQL (P)	SELECT (Number of electorates (2009)) FROM 2-17922541-1 WHERE Constituency number = 46
		ANS (T)	136.0
		ANS (P)	136,987
		ERROR	Ground Truth
83	7062	NL	What is the MIntage after 2006 of the Ruby-Throated Hummingbird Theme coin?
		TBL	“Year”, “Theme”, “Face Value”, “Weight”, “Diameter”, “Mintage”, “Issue Price”
		SQL (T)	SELECT max(Mintage) FROM 2-17757354-2 WHERE Year > 2006 AND Theme = ruby-throated hummingbird
		SQL (P)	SELECT (Mintage) FROM 2-17757354-2 WHERE Year > 2006 AND Theme = ruby-throated hummingbird
		ANS (T)	25,000
		ANS (P)	25,000
		ERROR	Ground Truth
84	7070	NL	What is the date of the zolder circuit, which had a.z.k./roc-compétition a.z.k./roc-compétition as the winning team?
		TBL	“Round”, “Circuit”, “Date”, “Winning driver”, “Winning team”
		SQL (T)	SELECT (Date) FROM 2-17997366-2 WHERE Winning team = a.z.k./roc-compétition a.z.k./roc-compétition AND Circuit = zolder
		SQL (P)	SELECT (Date) FROM 2-17997366-2 WHERE Circuit = zolder AND Winning team = a.z.k./roc-compétition
		ANS (T)	5 may
		ANS (P)	None
		ERROR	Qestion (I)
85	7106	NL	What was the score of the BCS National Championship game?
		TBL	“Date”, “Bowl Game”, “Big Ten Team”, “Opp. Team”, “Score”

		SQL (T)	SELECT (Score) FROM 2-18102742-1 WHERE Bowl Game = bcs national championship
		SQL (P)	SELECT (Score) FROM 2-18102742-1 WHERE Bowl Game = bcs national championship game
		ANS (T)	38-24
		ANS (P)	None
		ERROR	None
86	7182	NL	What is the finishing time with a 2/1q finish on the Meadowlands track?
		TBL	"Date", "Track", "Race", "Finish", "Fin. Time", "Last 1/4", "Driver", "Trainer"
		SQL (T)	SELECT (Fin. Time) FROM 2-18744745-2 WHERE Finish = 2/1q AND Track = the meadowlands
		SQL (P)	SELECT (Fin. Time) FROM 2-18744745-2 WHERE Track = meadowlands AND Finish = 2/1q
		ANS (T)	1:47.1
		ANS (P)	None
		ERROR	None
87	7290	NL	What is the total poverty (2009) HPI-1 % when the extreme poverty (2011) <1.25 US$ % of 16.9, and the human development (2012) HDI is less than 0.581?
		TBL	"Country", "Human development (2012) HDI", "GDP (PPP) (2012) US$ per capita", "Real GDP growth (2011) %", "Income inequality (2011) Gini", "Poverty (2009) HPI-1 %", "Extreme poverty (2011) <1.25 US$ %", "Literacy (2010) %", "Life expectancy (2011) Years", "Murder (2012) Rate per 100,000", "Peace (2012) GPI"
		SQL (T)	SELECT sum(Poverty (2009) HPI-1 %) FROM 2-18524-3 WHERE Extreme poverty (2011) <1.25 US$ % = 16.9 AND Human development (2012) HDI < 0.581
		SQL (P)	SELECT count(Poverty (2009) HPI-1 %) FROM 2-18524-3 WHERE Human development (2012) HDI < 0.581 AND Extreme poverty (2011) <1.25 US$ % = 16.9
		ANS (T)	None
		ANS (P)	0
		ERROR	None
88	7306	NL	Which Heat has a Nationality of bulgaria, and a Result larger than 55.97?
		TBL	"Rank", "Heat", "Name", "Nationality", "Result"
		SQL (T)	SELECT min(Heat) FROM 2-18579281-5 WHERE Nationality = bulgaria AND Result > 55.97
		SQL (P)	SELECT avg(Heat) FROM 2-18579281-5 WHERE Nationality = bulgaria AND Result > 55.97
		ANS (T)	None
		ANS (P)	None
		ERROR	Ground Truth
89	7425	NL	What is galway county's total?
		TBL	"Rank", "Player", "County", "Tally", "Total", "Opposition"
		SQL (T)	SELECT sum(Total) FROM 2-18936986-3 WHERE County = galway
		SQL (P)	SELECT (Total) FROM 2-18936986-3 WHERE County = galway
		ANS (T)	9.0
		ANS (P)	9.0
		ERROR	Ground Truth
90	7438	NL	With a Col (m) larger than 2012, what is Mount Kazbek's Prominence (m)?

		TBL	“Peak”, “Country”, “Elevation (m)”, “Prominence (m)”, “Col (m)”
		SQL (T)	`SELECT` (Prominence (m)) `FROM` 2-18918776-12 `WHERE` Col (m) $>$ 2012 `AND` Peak $=$ mount kazbek
		SQL (P)	`SELECT` max(Prominence (m)) `FROM` 2-18918776-12 `WHERE` Peak $=$ mount kazbek `AND` Col (m) $>$ 2012
		ANS (T)	2,353
		ANS (P)	2,353
		ERROR	None
91	7479	NL	What’s the position that has a total less than 66.5m, a compulsory of 30.9 and voluntary less than 33.7?
		TBL	“Position”, “Athlete”, “Compulsory”, “Voluntary”, “Total”
		SQL (T)	`SELECT` min(Position) `FROM` 2-18662083-1 `WHERE` Total $<$ 66.5 `AND` Compulsory $=$ 30.9 `AND` Voluntary $<$ 33.7
		SQL (P)	`SELECT` sum(Position) `FROM` 2-18662083-1 `WHERE` Compulsory $=$ 30.9 `AND` Voluntary $<$ 33.7 `AND` Total $<$ 66.5
		ANS (T)	None
		ANS (P)	None
		ERROR	Ground Truth
92	7578	NL	What is the high checkout when Legs Won is smaller than 9, a 180s of 1, and a 3-dart Average larger than 88.36?
		TBL	“Player”, “Played”, “Legs Won”, “Legs Lost”, “100+”, “140+”, “180s”, “High Checkout”, “3-dart Average”
		SQL (T)	`SELECT` sum(High Checkout) `FROM` 2-18621456-22 `WHERE` Legs Won $<$ 9 `AND` 180s $=$ 1 `AND` 3-dart Average $>$ 88.36
		SQL (P)	`SELECT` max(High Checkout) `FROM` 2-18621456-22 `WHERE` Legs Won $<$ 9 `AND` 180s $=$ 1 `AND` 3-dart Average $>$ 88.36
		ANS (T)	None
		ANS (P)	None
		ERROR	Ground Truth
93	7664	NL	What’s Brazil’s lane with a time less than 21.15?
		TBL	“Rank”, “Lane”, “Athlete”, “Nationality”, “Time”, “React”
		SQL (T)	`SELECT` min(Lane) `FROM` 2-18569011-6 `WHERE` Nationality $=$ brazil `AND` Time $<$ 21.15
		SQL (P)	`SELECT` sum(Lane) `FROM` 2-18569011-6 `WHERE` Nationality $=$ brazil `AND` Time $<$ 21.15
		ANS (T)	None
		ANS (P)	None
		ERROR	Ground Truth
94	7682	NL	What’s the total of rank 8 when Silver medals are 0 and gold is more than 1?
		TBL	“Rank”, “Nation”, “Gold”, “Silver”, “Bronze”, “Total”
		SQL (T)	`SELECT` count(Total) `FROM` 2-18807607-2 `WHERE` Silver $=$ 0 `AND` Rank $=$ 8 `AND` Gold $>$ 1
		SQL (P)	`SELECT` sum(Total) `FROM` 2-18807607-2 `WHERE` Rank $=$ 8 when silver medals are 0 `AND` Gold $>$ 1 `AND` Silver $=$ 0
		ANS (T)	0
		ANS (P)	None
		ERROR	None
95	7725	NL	How many cuts made in the tournament he played 13 times?

		TBL	"Tournament", "Wins", "Top-25", "Events", "Cuts made"
		SQL (T)	`SELECT` sum(Cuts made) `FROM` 2-12702607-1 `WHERE` Events > 13
		SQL (P)	`SELECT` (Cuts made) `FROM` 2-12702607-1 `WHERE` Wins = 13
		ANS (T)	None
		ANS (P)	None
		ERROR	Ground Truth
96	7854	NL	What Nominating festival was party of the adjustment film?
		TBL	"Category", "Film", "Director(s)", "Country", "Nominating Festival"
		SQL (T)	`SELECT` (Nominating Festival) `FROM` 2-12152327-6 `WHERE` Film = adjustment
		SQL (P)	`SELECT` (Nominating Festival) `FROM` 2-12152327-6 `WHERE` Film = party of the adjustment
		ANS (T)	prix uip angers
		ANS (P)	None
		ERROR	None
97	7912	NL	When did Gaspare Bona win the Pozzo Circuit?
		TBL	"Name", "Circuit", "Date", "Winning driver", "Winning constructor", "Report"
		SQL (T)	`SELECT` (Date) `FROM` 2-12631771-2 `WHERE` Winning driver = gaspare bona `AND` Name = pozzo circuit
		SQL (P)	`SELECT` (Date) `FROM` 2-12631771-2 `WHERE` Circuit = pozzo `AND` Winning driver = gaspare bona
		ANS (T)	20 march
		ANS (P)	20 march
		ERROR	Qestion (I)
98	7954	NL	What was the attendance when the record was 77-54?
		TBL	"Date", "Opponent", "Score", "Loss", "Attendance", "Record"
		SQL (T)	`SELECT` min(Attendance) `FROM` 2-12207430-6 `WHERE` Record = 77-54
		SQL (P)	`SELECT` (Attendance) `FROM` 2-12207430-6 `WHERE` Record = 77-54
		ANS (T)	30,224
		ANS (P)	30,224
		ERROR	Ground Truth
99	8041	NL	Name the subject of shiyan
		TBL	"Chapter", "Chinese", "Pinyin", "Translation", "Subject"
		SQL (T)	`SELECT` (Subject) `FROM` 2-1216675-1 `WHERE` Pinyin = shiyan
		SQL (P)	`SELECT` (Subject) `FROM` 2-1216675-1 `WHERE` Translation = shiyan
		ANS (T)	verbs, adjectives, adverbs
		ANS (P)	None
		ERROR	Qestion (I)
100	8111	NL	What is the language of the film Rosie?
		TBL	"Year (Ceremony)", "Film title used in nomination", "Original title", "Language(s)", "Result"
		SQL (T)	`SELECT` (Language(s)) `FROM` 2-13330057-1 `WHERE` Original title = rosie
		SQL (P)	`SELECT` (Language(s)) `FROM` 2-13330057-1 `WHERE` Film title used in nomination = rosie
		ANS (T)	dutch
		ANS (P)	dutch