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 literature 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.
Table:

<table>
<thead>
<tr>
<th>Player</th>
<th>Country</th>
<th>Points</th>
<th>Winnings ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steve Stricker</td>
<td>United States</td>
<td>9000</td>
<td>1260000</td>
</tr>
<tr>
<td>K.J. Choi</td>
<td>South Korea</td>
<td>5400</td>
<td>756000</td>
</tr>
<tr>
<td>Rory Sabbatini</td>
<td>South Africa</td>
<td>3400</td>
<td>4760000</td>
</tr>
<tr>
<td>Mark Calcavecchia</td>
<td>United States</td>
<td>2067</td>
<td>289333</td>
</tr>
<tr>
<td>Ernie Els</td>
<td>South Africa</td>
<td>2067</td>
<td>289333</td>
</tr>
</tbody>
</table>

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.

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

---

1The source code and human evaluation data is available from [https://github.com/naver/sqlova](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 $T_{n,1}, \ldots T_{n,L}$ ($L$ is the number of query words) is encoded as

$$[CLS], T_{n,1}, \ldots, T_{n,L}, [SEP], T_{h,1}, T_{h,2}, \ldots, [SEP], \ldots, [SEP], T_{h,N_h,1}, \ldots, T_{h,N_h,M_{N_h}}, [SEP]$$

where $T_{h,k}$ 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].

\[
\begin{align*}
    s(n|c) &= D_n^T W E_n \\
    p(n|c) &= \text{softmax}(s(n|c)) \\
    C_c &= \sum_n p(n|c) E_n \\
    s_{sc}(c) &= W \tanh([WD_c; WC_c]) \\
    p_{sc}(c) &= \text{softmax}(s_{sc}(c))
\end{align*}
\]  

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

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

\[
p_{sa}(\text{agg}|c) = \text{softmax}(W \tanh WC_c)
\]  

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 \( C \).

\[
\begin{align*}
    p(c) &= \text{softmax}(WD_c) \\
    C &= \sum_c p(c) D_c \\
    h &= WC \\
    c &= WC \\
    p'(n) &= \text{softmax}(W \text{ bi-LSTM}(E_n, h, c)) \\
    C_Q &= \sum_n p'(n) E_n \\
    s_{wn} &= W \tanh WC_Q
\end{align*}
\]  

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 \( \text{softmax}(s_{wn}) \). 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,

\[
\begin{align*}
    s_{wc}(c) &= W \tanh([WD_c; WC_c]) \\
    p_{wc}(c) &= \text{sigmoid}(s_{wc}(c))
\end{align*}
\]  

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.

$$s_{wo}(op|c) = W \tanh W([WD_c; WC_c])$$
$$p_{wo}(op|c) = \text{softmax} s_{wo}(op|c)$$  \hspace{1cm} (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 $op$.

$$\text{vec} = [E_n; WC_c; WD_c; WW_{op}]$$
$$s_{wv}(n|c, op) = W \tanh W \text{vec}$$  \hspace{1cm} (6)$$

Here, $V_{op}$ 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 \cite{Xu2017} 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 \cite{Dong2018}. 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 \cite{Xu2017}. 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 \cite{Wang2018a,Yin2018}. 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 $\text{select } \text{agg}(\text{col}_w) \text{ where } \text{col}_w \text{ op val}$. Here, col$_w$ is the predicted select column, agg is the predicted aggregation operator, 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) 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 Stanford CoreNLP \cite{Manning2014}. Each token is further tokenized (into sub-word level) by WordPiece tokenizer \cite{Devlin2018, Wu2016}. The headers of the tables and SQL vocabulary are tokenized by WordPiece tokenizer directly. The PyTorch version of BERT code\footnote{https://github.com/google-research/bert} is used for word embedding and some part of the code in NL2SQL LAYER is influenced by the original SQLNet source code\footnote{https://github.com/salesforce/WikiSQL}. All experiments were performed on WikiSQL \footnote{https://github.com/huggingface/pytorch-pretrained-BERT} ver. 1.1 \footnote{https://github.com/xiaojunxu/SQLNet}.
Table 1: Comparison of various models. Logical form accuracy (LF) and execution accuracy (X) on dev and test set of WikiSQL. “EG” stands for “execution-guided”.

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev LF (%)</th>
<th>Dev X (%)</th>
<th>Test LF (%)</th>
<th>Test X (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (Zhong et al., 2017)</td>
<td>23.3</td>
<td>37.0</td>
<td>23.4</td>
<td>35.9</td>
</tr>
<tr>
<td>Seq2SQL (Zhong et al., 2017)</td>
<td>49.5</td>
<td>60.8</td>
<td>48.3</td>
<td>59.4</td>
</tr>
<tr>
<td>SQLNet (Xu et al., 2017)</td>
<td>63.2</td>
<td>69.8</td>
<td>61.3</td>
<td>68.0</td>
</tr>
<tr>
<td>PT-MAML (Huang et al., 2018)</td>
<td>63.1</td>
<td>68.3</td>
<td>62.8</td>
<td>68.0</td>
</tr>
<tr>
<td>TypeSQL (Yu et al., 2018)</td>
<td>68.0</td>
<td>74.5</td>
<td>66.7</td>
<td>73.5</td>
</tr>
<tr>
<td>Coarse2Fine (Dong and Lapata, 2018)</td>
<td>72.5</td>
<td>79.0</td>
<td>71.7</td>
<td>78.5</td>
</tr>
<tr>
<td>MQAN (McCann et al., 2018)</td>
<td>76.1</td>
<td>82.0</td>
<td>75.4</td>
<td>81.4</td>
</tr>
<tr>
<td>Annotated Seq2Seq (Wang et al., 2018)</td>
<td>72.1</td>
<td>82.1</td>
<td>72.1</td>
<td>82.2</td>
</tr>
<tr>
<td>Coarse2Fine+EG (Wang et al., 2018)</td>
<td>76.0</td>
<td>84.0</td>
<td>75.4</td>
<td>83.8</td>
</tr>
<tr>
<td>IncSQL (Shi et al., 2018)</td>
<td>49.9</td>
<td>84.0</td>
<td>49.9</td>
<td>83.7</td>
</tr>
<tr>
<td>BERT-TO-SEQUENCE (ours)</td>
<td>57.3</td>
<td>-</td>
<td>56.4</td>
<td>-</td>
</tr>
<tr>
<td>BERT-TO-TRANSFORMER (ours)</td>
<td>70.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SQLOVA (ours)</td>
<td>81.6 (+5.5)</td>
<td>87.2 (+3.2)</td>
<td>80.7 (+5.3)</td>
<td>86.2 (+2.5)</td>
</tr>
<tr>
<td>PointSQL+EG (Wang et al., 2018)</td>
<td>67.5</td>
<td>78.4</td>
<td>67.9</td>
<td>78.3</td>
</tr>
<tr>
<td>Coarse2Fine+EG (Wang et al., 2018)</td>
<td>76.0</td>
<td>84.0</td>
<td>75.4</td>
<td>83.8</td>
</tr>
<tr>
<td>IncSQL+EG (Shi et al., 2018)</td>
<td>51.3</td>
<td>87.2</td>
<td>51.1</td>
<td>87.1</td>
</tr>
<tr>
<td>SQLOVA+EG (ours)</td>
<td>84.2 (+8.2)</td>
<td>90.2 (+3.0)</td>
<td>83.6 (+8.2)</td>
<td>89.6 (+2.5)</td>
</tr>
<tr>
<td>Human performance</td>
<td>88.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 ≥ 95% in accuracy except select-aggregation module whose low accuracy partially results from the error in the ground-truth of WikiSQL (Section 5).

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-operator and where-value respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>s-col</th>
<th>s-agg</th>
<th>w-num</th>
<th>w-col</th>
<th>w-op</th>
<th>w-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQLOVA, Dev</td>
<td>97.3</td>
<td>90.5</td>
<td>98.7</td>
<td>94.7</td>
<td>97.5</td>
<td>95.9</td>
</tr>
<tr>
<td>SQLOVA, Test</td>
<td>96.8</td>
<td>90.6</td>
<td>98.5</td>
<td>94.3</td>
<td>97.3</td>
<td>95.4</td>
</tr>
</tbody>
</table>

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 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.

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

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/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.

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

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 SQL OVA, 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 SQL OVA 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 SQL OVA over 1,551 samples are 86.8% (w/o EG) and 91.0% (w/ EG).

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


Footnotes:

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


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.


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 A1).

select-column module finds the column in select clause from given natural language utterance by modeling the probability of choosing i-th header \( p_{sc}(\text{col}_i) \) as
\[
p_{sc}(\text{col}_i) = \text{softmax}(H_{h,i})_0
\]
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 \( \mu \)-th element of vector \( V \) in this paper. Also, the conditional probability for given question and table-schema \( p(\cdot|Q, \text{table-schema}) \) 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 \( \text{agg} \) for given select column \( \text{col}_i \) is described by
\[
p_{sa}(\text{agg}_\mu|\text{col}_i) = \text{softmax}((H_{h,i})_\mu)
\]
where \( \text{agg}_1, \text{agg}_2, \text{agg}_3, \text{agg}_4, \text{agg}_5, \) and \( \text{agg}_6 \) are none, max, min, count, sum, and \( \text{avg} \) respectively.

where-number module predicts the number of where conditions by modeling the probability of generating \( \mu \)-number of conditions as
\[
p_{wn}(\mu) = \text{softmax}((WH_{[CLS]})_\mu)
\]
where \( H_{[CLS]} \) is the output vector of \([CLS]\) token from table-aware BERT encoder, and \( W \) stands for affine transformation. Throughout the paper, any affine transformation shall be denoted by \( W \) for the clarity.

where-column module calculates the probability of generating each columns in where clause. The probability of generating \( \text{col}_i \) is given by
\[
p_{wc}(\text{col}_i) = \text{sigmoid}((H_{h,i})_\gamma).
\]
where-operator module finds most probable operators for given where column among three possible choices \( (>,-,\leq) \). The probability of generating operator \( \text{op}_\mu \) for given where column \( \text{col}_i \) is modeled as
\[
p_{wo}(\text{op}_\mu|\text{col}_i) = \text{softmax}((H_{h,i})_\mu)
\]
where \( \text{op}_1, \text{op}_2, \) and \( \text{op}_3 \) 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 column \( \text{col}_i \) is modeled as
\[
p_{wv,st}(k|\text{col}_i) = \text{softmax}((H_{n,k})_\mu).
\]
Similarly the probability of \( k \)-th question token is selected as an end token is
\[
p_{wv,ed}(k|\text{col}_i) = \text{softmax}((H_{n,k+100})_\mu).
\]
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 \cite{vinyals2015pointer} \cite{zhong2017topical} (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 attention mechanism.

\[
D_t = \text{LSTM}(P_{t-1}, (h_{t-1}, c_{t-1}))
\]
\[
h_0 = (WH_{[CLS]})_{0:d}
\]
\[
c_0 = (WH_{[CLS]})_{d:2d}
\]
\[
s_t(i) = W(WH_i + WD_t)
\]
\[
p_t(i) = \text{softmax}(s_t(i)).
\]
There is text in the image that needs to be transcribed into a plain text representation. The text is part of a document discussing the performance of different models in generating SQL queries. It includes a table comparing the performance of shallow and deeper models with and without EG. The text explains the model schemes and provides insights into the decoding process and the performance metrics. The table shows the logical from accuracy (LF) and execution accuracy (X) on dev and test sets for WikiSQL. The models include Shallow-Layer and Decoder-Layer with and without EG. The performance metrics are given for both development and test sets.
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.

<table>
<thead>
<tr>
<th>SQL (GT)</th>
<th>correct</th>
<th>incorrect</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL (Ours)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>correct</td>
<td>0</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>incorrect</td>
<td>25</td>
<td>8</td>
<td>33</td>
</tr>
<tr>
<td>total</td>
<td>25</td>
<td>49</td>
<td>74</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>types</th>
<th>QID</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>19, 261, 557, 597, 598, 738, 1089, 1122, 1430, 1986, 2891, 3050, 3602, 3925, 5893, 6028, 6533</td>
<td>21</td>
</tr>
<tr>
<td>Type II</td>
<td>783, 2175, 4239</td>
<td>3</td>
</tr>
<tr>
<td>Type III</td>
<td>332</td>
<td>1</td>
</tr>
<tr>
<td>Type IV</td>
<td>156</td>
<td>1</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>No.</th>
<th>QID</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>NL</td>
<td>How many capital cities does Australia have?</td>
</tr>
</tbody>
</table>
2 55 NL What are the races that Johnny Rutherford has won?
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 Question (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** Question (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** Question (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 haŋ’áŋ na

**TBL** “English gloss”, “Santee-Sisseton”, “Yankton-Yanktonai”, “Northern Lakota”, “Southern Lakota”

**SQL (T)** SELECT (English gloss) FROM 1-1499774-5 WHERE Santee-Sisseton = haŋ’áŋ na

**SQL (P)** SELECT (English gloss) FROM 1-1499774-5 WHERE Southern Lakota = haŋ’áŋ na

**ANS (T)** morning

**ANS (P)** None

**ERROR** Question (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** Question (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

**ANS (T)** None

**ANS (P)** None

**ERROR** Question (I)
12 625 NL What is the sexual abuse rate where the conflict is the Burundi Civil War?

TBL

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

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

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 Question (I)

16 783 NL How many times was Plan B 4th place?

TBL

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
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

18 841 NL When there was a bye in the round of 32, what was the result in the round of 16?


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)

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

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

21 1056 NL When the total score is 740, what is tromso?


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

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 = '63-77' |
| SQL (P) | SELECT count(Date) FROM 1-19789597-5 WHERE Record = '163-77' |
| ANS (T) | 1 |
| ANS (P) | 0 |

**ERROR** Question (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** Question (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** Question (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
28 1591 NL What was the rating for Brisbane the week that Adelaide had 94000?


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

29 1851 NL Which year did enrolled Gambier members leave?


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

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

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

32 2090 NL What year was McMahon stadium founded?


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
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

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

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 min(Pick #) FROM 1-2850912-1 WHERE College/junior/club team = Regina Pats (WHL)
SQL (P) SELECT (Pick #) FROM 1-2850912-1 WHERE College/junior/club team = regina pats (whl)

ANS (T) 21.0
ANS (P) 21.0

ERROR

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

37 2311 NL What is the train number when the time is 10:38?


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

38 2323 NL What team hired Renato Gaúcho?
### 39 2565 NL

**Question:** What was attendance of the whole season when the average attendance for League Cup was 32,415?

**Table:**

<table>
<thead>
<tr>
<th>Season,</th>
<th>Season Total Att.,</th>
<th>K-League Season Total Att.,</th>
<th>Regular Season Average Att.,</th>
<th>League Cup Average Att.,</th>
<th>FA Cup Total / Average Att.,</th>
<th>ACL Total / Average Att.,</th>
<th>Friendly Match Att.</th>
</tr>
</thead>
</table>

**SQL (T)**

```sql
SELECT (Season Total Att.)
FROM 2-1056336-11
WHERE League Cup Average Att. = 32,415
```

**SQL (P)**

```sql
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

**Question:** 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?

**Table:**

<table>
<thead>
<tr>
<th>Team,</th>
<th>Class,</th>
<th>Chassis,</th>
<th>Engine,</th>
<th>Driver</th>
</tr>
</thead>
</table>

**SQL (T)**

```sql
SELECT (Driver)
FROM 2-15162596-2
WHERE Engine = rotax max AND Class = rotax heavy AND Chassis = arrow AND Team = twr raceline seating
```

**SQL (P)**

```sql
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

**Question:** If played is 22 and the tries against are 43, what are the points?

**Table:**

<table>
<thead>
<tr>
<th>Club,</th>
<th>Played,</th>
<th>Drawn,</th>
<th>Lost,</th>
<th>Points for,</th>
<th>Points against,</th>
<th>Tries for,</th>
<th>Tries against,</th>
<th>Points</th>
</tr>
</thead>
</table>

**SQL (T)**

```sql
SELECT (Points for)
FROM 2-13741576-4
WHERE Played = 22 AND Tries against = 43
```

**SQL (P)**

```sql
SELECT (Points)
FROM 2-13741576-4
WHERE Played = 22 AND Tries against = 43
```

**ANS (T)**

353

**ANS (P)**

46

**ERROR**

None

---

### 42 2925 NL

**Question:** What was the first Round with a Pick # greater than 1 and 140 Overall?

**Table:**

<table>
<thead>
<tr>
<th>Round,</th>
<th>Pick #,</th>
<th>Overall,</th>
<th>Name,</th>
<th>Position,</th>
<th>College</th>
</tr>
</thead>
</table>

**SQL (T)**

```sql
SELECT min(Round)
FROM 2-15198842-23
WHERE Pick # > 1 AND Overall > 140
```

**SQL (P)**

```sql
SELECT min(Round)
FROM 2-15198842-23
WHERE Pick # > 1 AND Overall = 140
```

**ANS (T)**

None

**ANS (P)**

6.0

---

22
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?
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 None

47 3370 NL When did Hans Hartmann drive?
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 None

48 3499 NL Which driver has less than 37 wins and at 14.12%?

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?


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, a 2006 greater than 38,7, and a 1998 less than 76?


SQL (T)  SELECT sum(2007) FROM 2-15348345-1 WHERE 2000 > 56,6 AND 2006 = 23,2 AND 1998 > 61,1


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?


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 < 76

ANS (T)  40.0

ANS (P)  35.41666666666664

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?
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?

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?

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 Question (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?

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?

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 Question (II)
### Question 58

**What is the week with an attendance of 75,555?**

<table>
<thead>
<tr>
<th>TBL</th>
<th>“Week”, “Date”, “Opponent”, “Result”, “TV Time”, “Attendance”</th>
</tr>
</thead>
</table>

| 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

### Question 59

**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

### Question 60

**How much Overall has a Name of bob anderson?**

| TBL | “Round”, “Pick”, “Overall”, “Name”, “Position”, “College” |

| SQL (T) | `SELECT sum(Overall) FROM 2-17100961-17 WHERE Name = 'andy north' AND Total > 153` |
| SQL (P) | `SELECT count(Overall) FROM 2-17100961-17 WHERE Name = 'andy north' AND Total > 153` |

| ANS (T) | 1 |
| ANS (P) | 68.0 |

**Error:** None

### Question 61

**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 sum(Avg/G) FROM 2-16981858-6 WHERE Name = 'mccrary, greg'` |

| ANS (T) | None |
| ANS (P) | None |

**Error:** None

### Question 62

**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

### Question 63

**What is the total avg/g of McCrary, Greg?**


| SQL (T) | `SELECT sum(Avg/G) FROM 2-16981858-6 WHERE Name = 'mccrary, greg'` |
| SQL (P) | `SELECT count(Avg/G) FROM 2-16981858-6 WHERE Name = 'mccrary, greg'` |

| ANS (T) | 1 |
| ANS (P) | 58.9 |
64 5456 NL What year has a Schwante smaller than 2.043, an Eichstädt smaller than 848, and a Bärenklau smaller than 1.262?


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

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

66 5705 NL What is the grid for the Minardi Team USA with laps smaller than 90?


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

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

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

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  Question (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  Question (I)

72  6089  NL  What are the draws when wins are fewer than 9 and byes fewer than 2?
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?
<table>
<thead>
<tr>
<th>NT</th>
<th>ID</th>
<th>NL</th>
<th>TBL</th>
<th>SQL (T)</th>
<th>SQL (P)</th>
<th>ANS (T)</th>
<th>ANS (P)</th>
<th>ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>6439</td>
<td>In the match where the home team scored 14.20 (104), how many attendees were in the crowd?</td>
<td>“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”</td>
<td>SELECT sum(Crowd) FROM 2-10790397-5 WHERE Home team score = 14.20 (104)</td>
<td>SELECT (Crowd) FROM 2-10790397-5 WHERE Home team score = 14.20 (104)</td>
<td>25.0</td>
<td>25,000</td>
<td>None</td>
</tr>
<tr>
<td>77</td>
<td>6440</td>
<td>In the match where the away team scored 2.7 (19), how many people were in the crowd?</td>
<td>“Home team”, “Home team score”, “Away team”, “Away team score”, “Venue”, “Crowd”, “Date”</td>
<td>SELECT max(Crowd) FROM 2-10790397-5 WHERE Away team score = 2.7 (19)</td>
<td>SELECT (Crowd) FROM 2-10790397-5 WHERE Away team score = 2.7 (19)</td>
<td>15,000</td>
<td>15,000</td>
<td>None</td>
</tr>
<tr>
<td>78</td>
<td>6533</td>
<td>Name the Score united states of tom watson in united state?</td>
<td>“Place”, “Player”, “Country”, “Score”, “To par”</td>
<td>SELECT (Score) FROM 2-18113463-4 WHERE Country = united states AND Player = tom watson</td>
<td>SELECT (Score) FROM 2-18113463-4 WHERE Place = united states AND Player = tom watson AND Country = united states</td>
<td>68.0</td>
<td>None</td>
<td>Qestion (I)</td>
</tr>
<tr>
<td>79</td>
<td>6688</td>
<td>How many ties did he have when he had 1 penalties and more than 20 conversions?</td>
<td>“Played”, “Drawn”, “Lost”, “Winning %”, “Tries”, “Conversions”, “Penalties”, “s Drop goal”, “Points total”</td>
<td>SELECT sum(Drawn) FROM 2-1828549-1 WHERE Penalties = 1 AND Conversions &gt; 20</td>
<td>SELECT (Drawn) FROM 2-1828549-1 WHERE Conversions &gt; 20 AND Penalties = 1</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>80</td>
<td>6729</td>
<td>What was the year that had Anugerah Bintang Popular Berita Harian 23 as competition?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>None</td>
</tr>
</tbody>
</table>
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?
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 Question (I)

85 7106 NL What was the score of the BCS National Championship game?
TBL “Date”, “Bowl Game”, “Big Ten Team”, “Opp. Team”, “Score”
86 7182  What is the finishing time with a 2/1q finish on the Meadowlands track?
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  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?
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  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  None

89 7425  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  With a Col (m) larger than 2012, what is Mount Kazbek’s Prominence (m)?
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?
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 Qestion (I)

When did Gaspare Bona win the Pozzo Circuit?

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)

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

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)

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