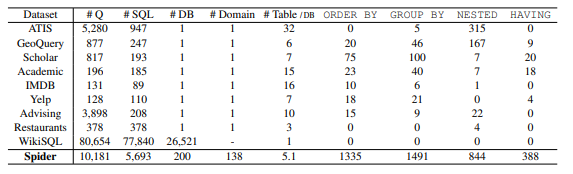
Datasets SQL



# WikiSQL

Github: <https://github.com/salesforce/WikiSQL>

Paper: <https://arxiv.org/abs/1709.00103>

Descripción: dataset de preguntas sobre una base de datos acompañadas de una consulta SQL, una tabla con datos y la respuesta a la consulta sobre la tabla. Los autores proponen un algoritmo de Reinforcement Learning que aprende ejecutando consultas que genera sobre una base de datos.

**Ver soluiciones en leaderboard**

Tamaño: 80654 ejemplos, 24241 tablas.

Código incluido:

* Anotación
* Evaluación

Metricas de evaluación:

* Exact match sobre la query
* Match sobre los resultados

# Spider

Github: <https://github.com/taoyds/spider>

Paper: <https://arxiv.org/pdf/1809.08887.pdf>

Descripción: dataset de preguntas sobre una base de datos acompañadas de una consulta SQL, una tabla con datos y la respuesta a la consulta sobre la tabla. Los autores afirman que los datos y las consultas tienen un nivel de complejidad superior al de WikiSQL.

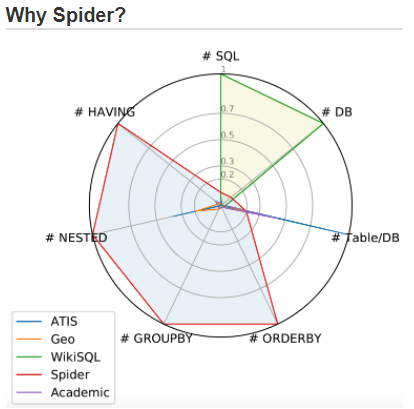
Tamaño: 200 bases de datos, 10.181 preguntas y 5.693 respuestas.

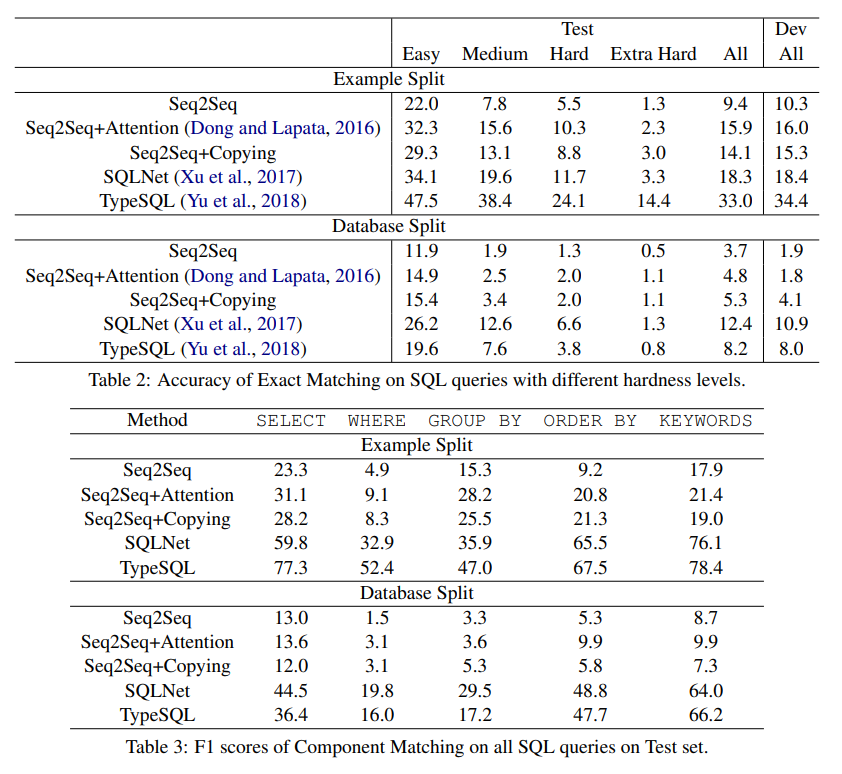
Código incluido:

* Entrenamiento de modelos
* Modelos entrenados
* Evaluación
* Pre-proceso de consultas SQL

Métricas de evaluación:

* Component matching
* Exact matching
* Execution accuracy





# Estrategias

* Uso de otros modelos (SQLNet, TypeSQL, …)
* Evaluar capacidad SQL de modelos actuales con WikiSQL y Spider (Codex)
* Entrenar y evaluar capacidad SQL de nuevos modelos con WikiSQL y Spider