

# Dataset character

Conv: Conversational

Paper	NLQ	Table	SQL	Feature	Knowledge Source
WTQ (2015)	Single	Single		<ul style="list-style-type: none"> <li>20% questions are not answerable by pure SQL;</li> <li>All/ Simple/ Complex;</li> <li>2,108 semi-structured Wikipedia tables, 22,033 examples;</li> </ul>	Wikipedia tables
WikiSQL (2017)	Single	Single		<ul style="list-style-type: none"> <li>Simple QA (cell filtering and aggregation, table no foreign key);</li> <li>24,241 tables, 80,654 examples;</li> </ul>	24,241 HTML tables from Wikipedia
SQA (2017)	Conv	Single		<ul style="list-style-type: none"> <li>Annotators decompose questions originally from WTQ to a sequence of conversational questions;</li> <li>982 tables, 6,066 examples;</li> </ul>	WTQ
Spider (2018)	Single	Multi	✓	<ul style="list-style-type: none"> <li>8,659 train and 1,034 dev example, covers 200 complex databases across 138 domains;</li> </ul>	Scholar, WikiSQL, GeoQuery, etc.
TabFact (2020)	Claim	Single		<ul style="list-style-type: none"> <li>16k tables as the evidence for 118k human-annotated natural language statements (entailed/ refuted);</li> <li>All/ Simple/ Complex;</li> </ul>	Wikipedia
SQUALL (2020)	Single	Single	✓	<ul style="list-style-type: none"> <li>Enriches WTQ with manually created SQL equivalents plus alignments between SQL and NLQ fragments;</li> <li>11,468 {table, Q&amp;A, logical form, alignment};</li> </ul>	WTQ
Spider-Syn (2021)	Single	Multi	✓	<ul style="list-style-type: none"> <li>Modify NL questions in Spider using synonym substitution;</li> <li>7000 train and 1034 dev samples;</li> </ul>	Spider
Spider-DK (2021)	Single	Multi	✓	<ul style="list-style-type: none"> <li>5 types of domain knowledge;</li> <li>535 NL-SQL pairs from Spider dev, where 265 pairs are modified to incorporate the domain knowledge;</li> </ul>	Spider
FeTaQA (2021)	Single	Single		<ul style="list-style-type: none"> <li>10K {table, question, free-form answer, supporting table cells} pairs;</li> <li>Long-form answers, require complex reasoning and integration info;</li> </ul>	Wikipedia
SEM-TAB-FACTS (2021)	Single	Single		<ul style="list-style-type: none"> <li>981 manually generated tables and 1980 auto-generated tables, over 180K statements and over 16M evidence annotations;</li> <li>Supported/ Refuted/ Unknown;</li> </ul>	IBM Watson Discovery
FinQA 🏦 (2021)	Single	Single		<ul style="list-style-type: none"> <li>8,281 {Q&amp;A pairs with program, supporting fact} (table&lt;20 rows);</li> </ul>	FinTabNet

TAT-QA 🦿 (2021)	Single	Single		<ul style="list-style-type: none"> <li>• 16,552 samples with hybrid context of both tables and textual content from real-world financial reports;</li> <li>• Over 40% require numerical reasoning skills (+, -, *, /, count, comparison, and compositional operations);</li> <li>• Table: 3-30 rows and 3-6 columns</li> </ul>	500 financial reports from <a href="https://www.annualreports.com/">https://www.annualreports.com/</a>
RobuT (2023)	Single	Single		<ul style="list-style-type: none"> <li>• 143,477 human-annotated perturbed examples;</li> <li>• 10 perturbation types on four different levels (i.e., table header, table content, natural language question, and mix);</li> </ul>	WTQ, WIKISQL-WEAK, SQA
BIRD (2023)	Single	Multi	✅	<ul style="list-style-type: none"> <li>• Large-scale databases</li> <li>• Dirty database contents, external knowledge between NL questions and database contents, and SQL efficiency</li> <li>• Large number of domains</li> </ul>	32% from Kaggle, 48% from CTU Prague Relational Learning Repo, 20% built following DuSQL