Auto-complete NLQ and Convert to SQL

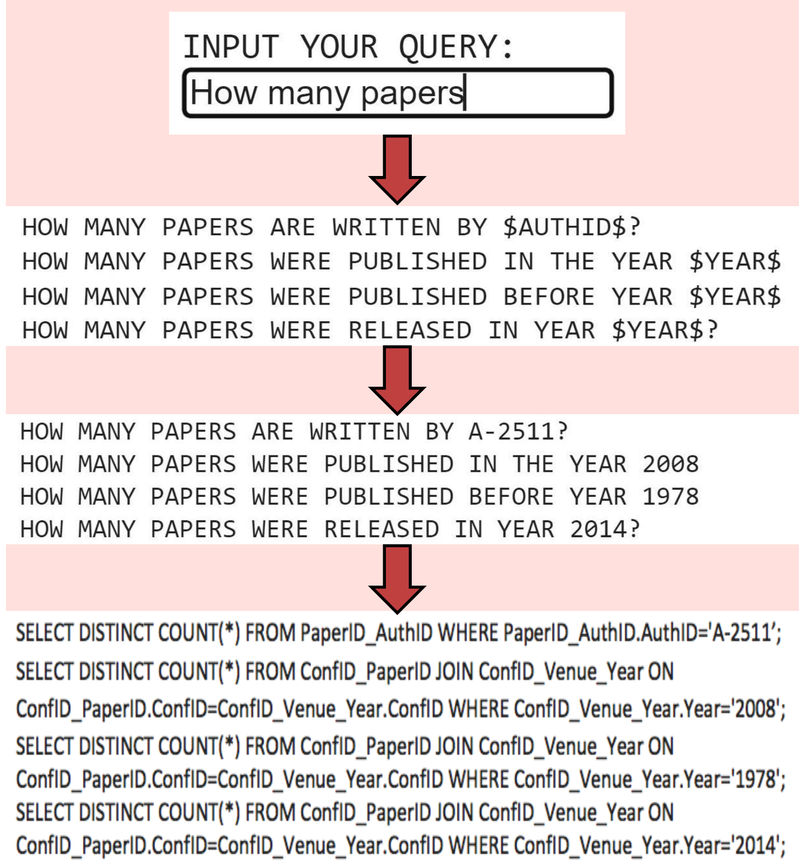
FIGHT ON ! - Sneha Kedia, Supraja Krovvidi, Priyanshi Agrawal, Laxmi Garde, Priya Bannur

USC Viterbi School of Engineering

Description

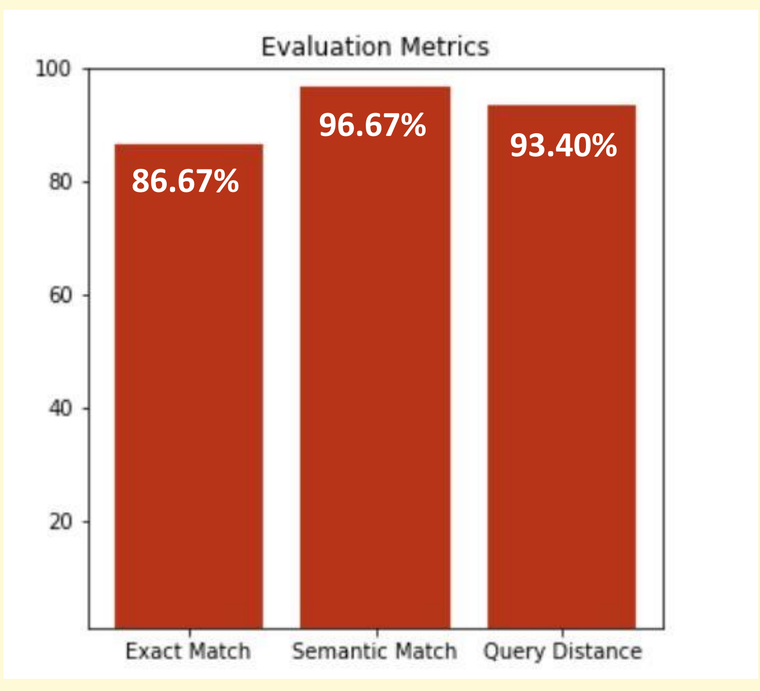
A system to predict SQL queries from incomplete NLQs by using auto-completion and BERT based conversion module for RDBMS.

Example



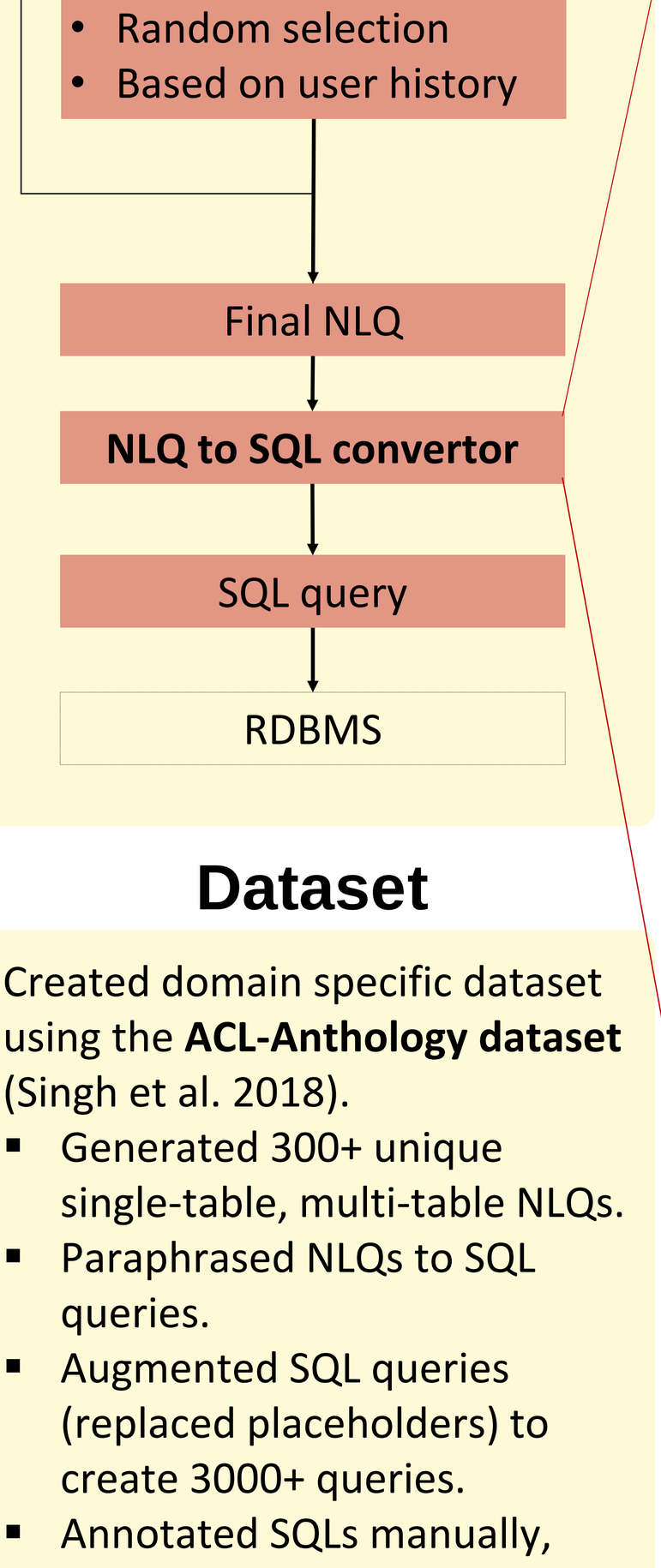
Results

- **Exact Match:** % of test set in which predicted SQL query is same as actual.
- **Semantic Match:** % of test set in which output of predicted and actual SQL query is same, though the queries might be different.
- **Query Distance:** Closeness of match based on no. of edges between the given columns in BFS tree. (QD = 1.132)



- **Zero Result Rate** = 0.29%

Method



Dataset

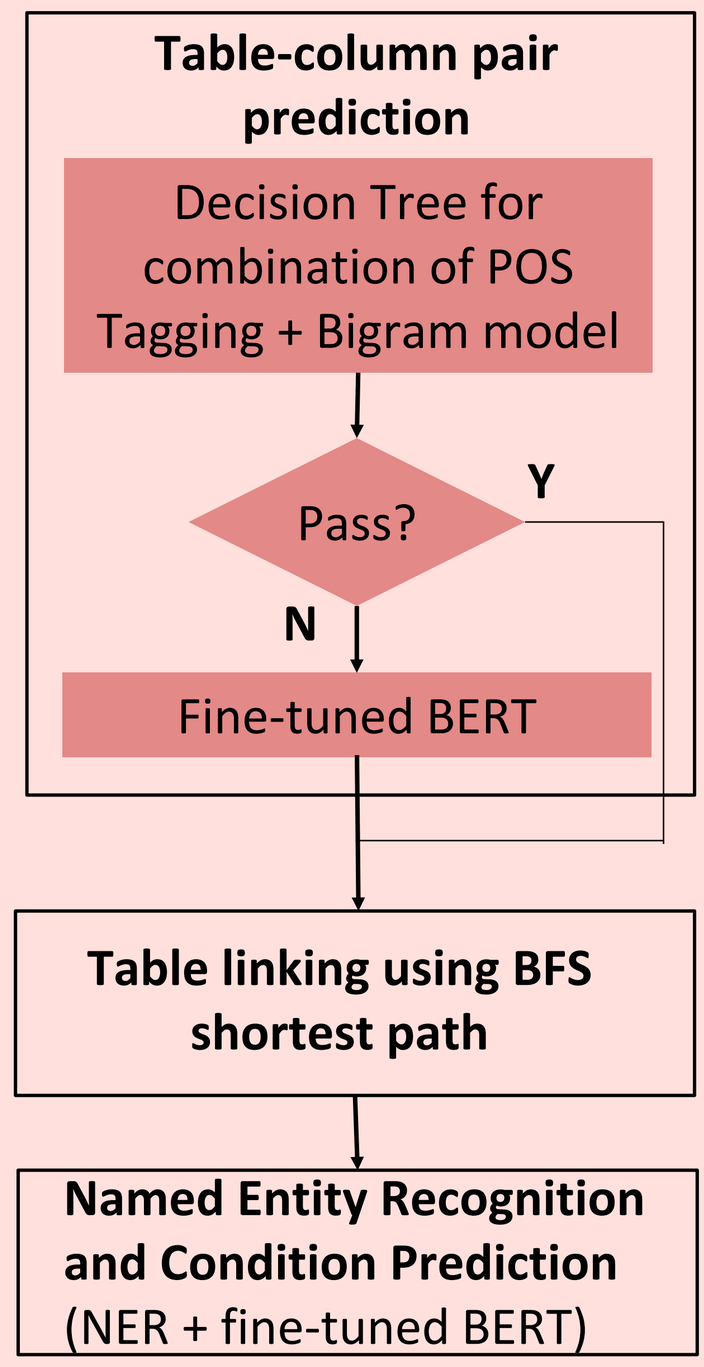
Created domain specific dataset using the **ACL-Anthology dataset** (Singh et al. 2018).

- Generated 300+ unique single-table, multi-table NLQs.
- Paraphrased NLQs to SQL queries.
- Augmented SQL queries (replaced placeholders) to create 3000+ queries.
- Annotated SQLs manually, used a semi-automated script for checking SQL table references.

Discussion

- Baseline NLQ-SQL model: SyntaxSQLNet, a syntax tree model for cross-domain tasks.
- Unlike baseline model, we encoded the relational schema (natural numbers mapping) and embedded it as a part of model.
- Flexibility for implementation on databases with varied schemas and can be fine-tuned for accurate results.

NLQ to SQL convertor



Conclusion

In comparison to existing architectures, our approach is customized for a specific database and hence giving good accuracy. Our model cannot predict complex queries with clauses like AND, HAVING, ORDER By and nested queries beyond 5 table linkages.