Introduction

- SQL is the standard query language widely used in order to communicate with relational databases and retrieve expected results.
- But in order to use SQL, user needs to be trained and must know the database schema to write the query.
- As the database user base is shifting towards non-experts, designing user-friendly query interfaces will be a more important goal in database community.

Goal

- The goal of this project is to make querying relational database really easy for users so that you don’t have to spend any resources on training.
- User will be provided with a simple query interface where user can type in their question in plain English and get the results back.

Approach

- User connects to the database through UI and all the database information is retrieved and stored.
- User enters query in plain English through UI which is converted into a parse tree using Stanford Parser.
- For e.g. user enters ‘return name for all patients whose doctors salary is greater than 2000 and name is Ankit’
- Each word in a sentence will be a leaf node in the tree.
- Each leaf node’s immediate parent node is the tagging node. Tagging nodes indicates what the word is for e.g. NNS indicates table name or attribute name, NNP indicates value of an attribute etc.
- Then each node in the parse tree is compared with each node in schema graph using WUP similarity function.
- Once all the nodes are mapped to corresponding attributes/tables updated tree is passed to the SQL generator.
- SQL generator has set of pre defined rules that converts the tree into SQL Query using those rules.

Architecture

- User enters query in plain English through UI which is converted into a parse tree using Stanford Parser.
- For e.g. user enters ‘return name for all patients whose doctors salary is greater than 2000 and name is Ankit’
- Each word in a sentence will be a leaf node in the tree.
- Each leaf node’s immediate parent node is the tagging node. Tagging nodes indicates what the word is for e.g. NNS indicates table name or attribute name, NNP indicates value of an attribute etc.
- Then each node in the parse tree is compared with each node in schema graph using WUP similarity function.
- Once all the nodes are mapped to corresponding attributes/tables updated tree is passed to the SQL generator.
- SQL generator has set of pre defined rules that converts the tree into SQL Query using those rules.

References

Fei Li, H. V. Jagadish, Constructing an interactive natural language interface for relational databases, Proceedings of the VLDB Endowment, v.8 n.1, p.73-84, September 2014.