A natural language interface with DBPal by TU Darmstadt

training time

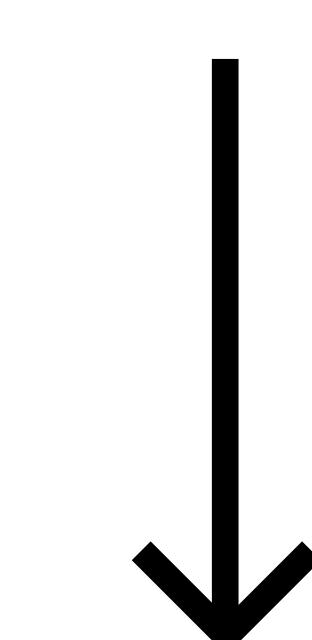
interference time

DBPal is a framework that translates natural language queries into SQL queries for any database schema without manual effort. In this way, billions of SQL/NL pairs can be generated as training data for arbitrary deepl learning models. Following the principle of weak supervision, templates that map concrete SQL queries to simple NL statements are used to automatically generate SQL/NL pairs. These are then augmented to cover a greater variety of NL statements.

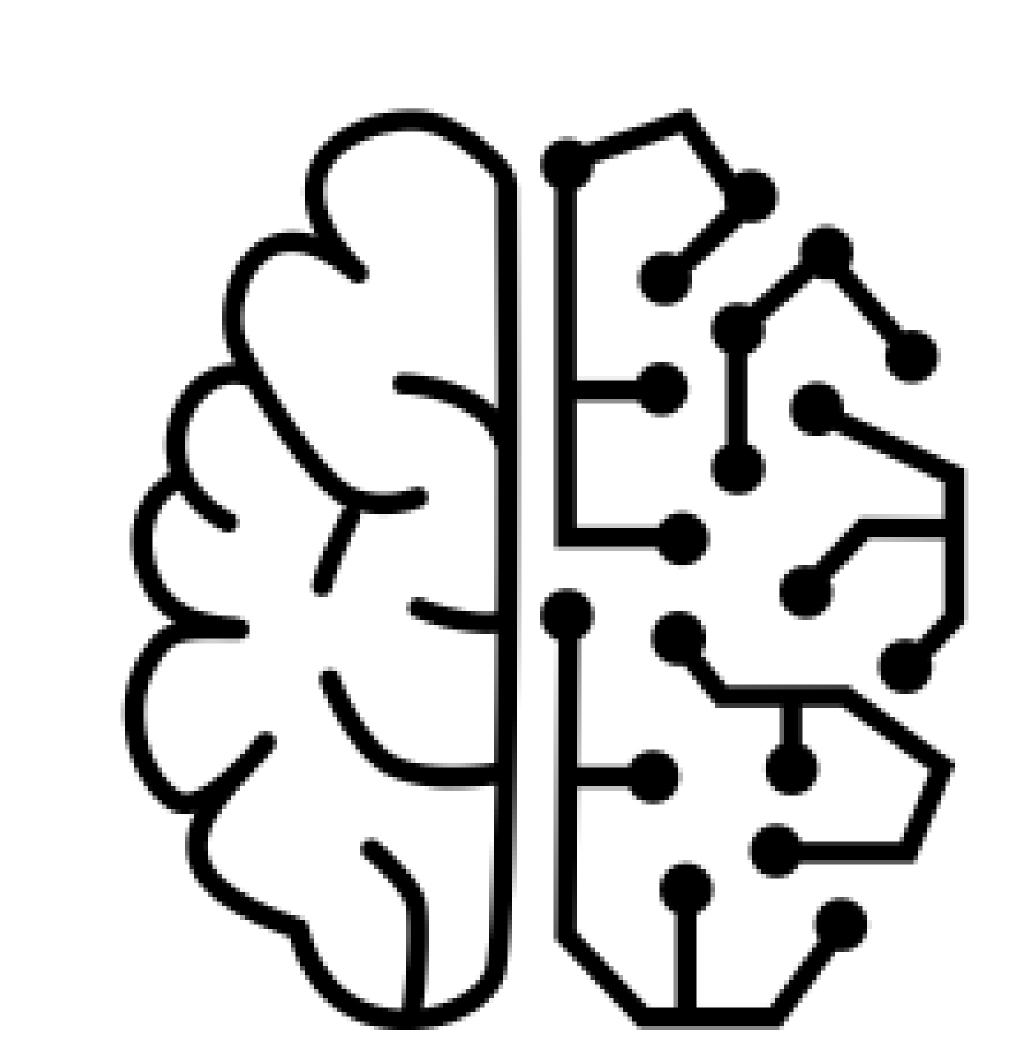
D B Pal

Natural Language query

"Show me the names of all patients with fewer."



Deep Learning Models



Any standard deep learning model for language translation can be trained with the training data

DB Schema

table patients

name	diagnose
Alice	fewer
Carsten	fewer
Ben	flu

corresponding template:

SELECT < Attribute >
FROM < Table >
WHERE < Filter >

"Show me the <Attribute>s of <Table>s with <Filter>"

SQL/NL pairs

The template is instantiated with the information about the database schema to create a few concrete queries that a user could execute on this database.

instantiated template:

SELECT name
FROM patients
WHERE diagnose=fewer

"Show me the names of patients with diagnose fewer"

Augmented SQL/NL pairs

Augmentation increases the number and variety of SQL/NL pairs by paraphrasing and noising. Paraphrasing is done using existing models (e.g. from Wikipedia) or by language pivoting (translation of NL statements to another NL and back). Noising removes single words to simulate missing information in a user input.

"Show me the names of patients with diagnose fewer"

SQL Query

"SELECT name
FROM patients
WHERE diagnose=fewer"