A natural language interface with DBPal by TU Darmstadt

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 deep 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.

### Abstract

Natural language interfaces offer the possibility to make interactive data analysis simple and concise for everyone, even for non SQL experts. One of the biggest challenges in natural language query processing is to understand incomplete, paraphrased or ambiguous statements. In contrast to less robust rule-based approaches, deep learning models can successfully overcome these challenges. But they require training data that is not available to the same extent for natural language to SQL translation as for other language translations. Even with crowd sourcing, it is not possible to collect sufficient training data because few people have SQL skills and new data would need to be collected for each database schema. With DBPal the TU Darmstadt presented a solution for this problem. The framework, which is based on weak supervision, automatically synthesizes a large amount of SQL/NL pairs for a given arbitrary database schema. This poster shows how DBPal generates training data that can then be used to run language translation models in order to enable interactive data analysis via a natural language interface.

Katharina Blaß (Bachelor)
Hasso-Plattner-Institut, Potsdam, Germany
E-Mail: katharina.blass@student.hpi.de

Referring Lecture: "Towards Interactive Data Analytics"
Icon source: Machine Learning by Angela from the Noun Project