

Technical Perspective: How Do Humans and Data Systems Establish a Common Query Language?

H. V. Jagadish
University of Michigan
jag@umich.edu

We all structure information in our brains: without structure, we would not be able to deal with the huge quantities of highly heterogeneous information we process. However, each of us structures this information slightly differently, often leading to misunderstandings or requiring additional rounds of dialog to clarify. Database schema are also designed by humans. The structure imposed on the information by the schema reflects the human designers' perspective on the world, even if mediated through formal design techniques or computer software. Therefore, the structured data querying task can be viewed as having a schema mapping problem at its core: mapping between the "schema" the human has in her brain and the schema used to organize the database.

In short, database querying would be much easier if only human users could know exactly how the database was structured (and also what sort of data it contained). Indeed, there is a significant body of work on data exploration. The basic idea here is that the human learns about what is in the database and how it is structured, leading her to the data items of interest. In the process, the user typically specifies a sequence of (exploratory) queries, each based in part on the knowledge about the database that the user has gained thus far. The system attempts to facilitate this exploration by the user. See, for example, [1], for a review of such work.

A completely distinct stream of work deals with approximately specified queries. The user provides a query intent, whether in a query language with wild cards, in natural language, or by example; the system then works hard to understand (and execute) the intended query. In the process, the user's incomplete specification is completed, any errors in it are corrected, and so on. In some proposals, the system may even engage in dialog with the user to clarify

user intent. However, the assumption is that the user intent is fixed during this process.

Seeing these two very distinct bodies of work on making databases easier to query, one should take a step back and see that in both cases, the system and the user have a shared objective of the system providing the user with the desired information. In the former, the user works to learn about the database; in the latter, the system works to learn about user intent. This paper bridges the two approaches and asks why both couldn't learn about each other.

Of course, this sort of two-sided learning is easier said than done. Figuring out how to do this effectively is the heart of the technical content in this paper. The authors model it as a cooperative two-player game, where both the user and the system are trying to achieve the same objective, which is to satisfy the user's information need. In each round of the game, the user may specify the need differently to help the system get to the right answer, and the system may interpret the user query differently, based on its growing understanding of the user's need. The strategy to play this game is learned through reinforcement learning.

This paper will open a whole new line of research, in which the user query statement is not kept fixed even if their information need remains unchanged.

1. REFERENCES

- [1] Matteo Lissandrini, Davide Mottin, Themis Palpanas, and Yannis Velegrakis, 2018. Data Exploration Using Example-Based Methods. *In Synthesis Lectures on Data Management*, Morgan and Claypool. DOI=<https://doi.org/10.2200/S00881ED1V01Y201810DTM053>