

From queries to data examples and back

Balder ten Cate
Universiteit van Amsterdam

In this proposed talk, we discuss some recent results around the use of data examples for illustrating and/or obtaining database queries.

A data example (for a query Q) is a database instance I together with information about the behavior of Q on this instance. In its simplest form, it is simply a pair $(I, Q(I))$. We could call such data examples *input-output examples*. Another type of data examples that have been studied are *positive and negative examples*, which exhibit partial information about the answers to the query in the given instance.

Data examples are useful for illustrating the behavior of a query. Conversely, given one or more data examples, one may wish to find a query that “fits” the data examples. In the talk I will review some recent results and ongoing work about various algorithmic problems that arise in this setting:

- *given a query, is there a small collection of data examples that uniquely characterize this query? If so, generate such a collection of data examples.*
- *given a collection of data examples, is there a query that fits the data examples? If so, find such a query. Also: is the query in question uniquely characterized by the data examples?*
- *given black-box access to (a compiled version of) a query Q , can we reverse-engineer Q (that is, algorithmically identify Q with certainty) by running it on a number of instances?*

We will focus on two classes of queries, namely *conjunctive queries* and *acyclic connected conjunctive queries*. The latter are an important subclass of CQs because they correspond precisely to the well-studied description logic concept language \mathcal{ELI} . For these classes, we will discuss answers to the above questions that have been obtained, in the form of algorithms, matching complexity bounds, and structural characterizations.

The talk is based on recent and ongoing work with Victor Dalmau and others.