Technical Perspective: Juggling Functions Inside a Database

Dan Olteanu
University of Oxford
dan.olteanu@cs.ox.ac.uk

The paper entitled "Juggling Functions Inside a Database" gives a brief overview of FAQ, a framework for computational problems expressed as **F**unctional **A**ggregate **Q**ueries. This work falls into my bucket of select database research contributions that go significantly beyond the state of the art along several dimensions. First, it provides an elegant and declarative formalism for a host of ubiquituous computational problems across Computer Science and at the right level of abstraction that exposes structural properties of the problem instances and allows for fine-grained complexity analysis. Second, it is technically deep, proposing an algorithmic solution that achieves lower than or the same complexity as specialized approaches in their respective domain. Third, it is implemented in a commercial database system with scores of real-world applications. Fourth, it is currently applied to in-database analytics and I expect more applications will manifest themselves in the near future.

By unifying many problems under the same formalism, FAQ bears the promise of accelerating research: Scalable data management solutions developed by our community for aggregates over joins, e.g., incremental view maintenance, index data structures, or distributed processing, may become general-purpose solutions for problems outside databases.

I will next expand on some of its contributions.

Unified approach to a host of computational problems. FAQ captures problems in relational databases, logic, matrix and tensor computation, probabilistic graphical models, constraint satisfaction, and signal processing. For instance, FAQ expressions represent queries with joins and aggregates in relational databases, matrix chain computation, maximum a posteriori and marginal distribution queries in probabilistic graphical models.

Technical contribution beyond state of the art. FAQ exploits recent groundbreaking developments on worst-case optimal join algorithms, started by researchers including the FAQ authors, and asymptotically tight bounds on join processing time and result size. This leads to lower complexities for long-standing problems, including counting quantified conjunctive queries in logic and inference in probabilistic graphical models.

FAQ also recovers database techniques, such as pushing aggregates past joins and computing aggregates over factorized joins.

Commercial deployment. FAQ can be easily plugged into existing relational database systems as a standalone library, since it mainly performs query rewriting; however, to attain its low complexity for queries with cycles, an optimal multi-way join algorithm would be also needed. It is in fact already deployed in the LogicBlox engine. FAQ expressions are translated into optimized programs consisting of strata of Datalog-like rules expressing joins over extensional and intensional predicates closed by aggregates with free variables. Although the paper does not report on performance of the FAQ implementation within LogicBlox, it is conceivable that its relative performance over existing solutions follows the reported complexity gap.

Bright future ahead. In-database analytics are a new application of FAQ, where optimization problems are pushed inside the database. The motivation for this application is twofold. First, since data usually resides inside the database, bringing the analytics closer to the data saves export/import time at the interface between database systems and statistical packages. Second, large chunks of machine learning code can be phrased as FAQ expressions!

In conclusion, this paper reports on a well-rounded work of both theoretical and practical relevance. It represents a significant improvement over the state of the art. While a database problem at its core, it can effectively accelerate research across Computer Science. It is an excellent lesson of elegance and technical mastery and I hope you will enjoy learning from it as much as I did.