

The semijoin algebra

Dirk Leinders,

PhD student under supervision of Jan Van den Bussche

Hasselt University and Transnational University of Limburg

The semijoin algebra is the fragment of the classical relational algebra obtained by replacing the join operator by the semijoin operator. Whereas the relational algebra is equivalent to the relational calculus, i.e., first-order logic, we show that when we only consider equi-semijoins, the semijoin algebra is equivalent in expressive power to the guarded fragment of first-order logic. This fragment was introduced by Andréka, van Benthem and Némethi in the late 1990's and has been studied extensively. As a result, the semijoin algebra inherits some nice properties of the guarded fragment, such as the finite model property.

On the other hand, also the semijoin operator is well known in the field of databases. Semijoins are often used as part of a query pre-processing phase where dangling tuples are eliminated. Here, an interesting property is that the size of a relation resulting from a semijoin is always linear in the size of the input. We will show that the semijoin algebra exactly captures the class of relational algebra expressions that always produce intermediate results of linear size. As an application, we show that any query plan for relational division, built up with relational algebra operators, must produce intermediate results of quadratic size.

This talk presents joint work with Jan Van den Bussche, Jerzy Tyszkiewicz, and Maarten Marx.