Lesson 13: Bounds on join queries

Theme: Bounds on the output of join queries.

The topic of this lesson is paper [1]. If you access it through the university network, you can download the article SIAM website: `http://epubs.siam.org/doi/abs/10.1137/110859440`. We will discuss Sections 1–4, and the main result is Theorem 1. If we have time, we will also briefly touch on Section 5.

A note on the notations. Recall that join queries(conjunctive queries) are queries of the form:

$$\varphi(\bar{z}) \quad := \quad R_1(\bar{x}_1) \ \wedge \ \cdots \ \wedge \ R_n(\bar{x}_n)$$

where $\bar{z}$ and each $\bar{x}_i$ are vectors of variables and $\{\bar{z}\} = \{\bar{x}_n\} \cup \cdots \cup \{\bar{x}_n\}$.

In [1] the authors use the following notations. A schema σ is a set of relation names $\{R_1, \ldots, R_n\}$. Each relation $R_i \in \sigma$ is equipped with a set of attributes $A_i = \{a_1, \ldots, a_m\}$. We write $R_i(A_i)$ to mean that A_i is the set of attributes of R_i.

For a set of attributes A, an A-tuple is a mapping $t : A \to \mathbb{N}$, where for every $a \in A$, t_a denotes $t(a)$[1]. An A-relation is a set of A-tuples. The projection of an A-tuple t to a subset $B \subseteq A$ is the restriction of t on B, denoted by $\pi_B(t)$. The projection of an A-relation R to B is the set of the projections of tuples in R to B.

A database instance D of schema σ consists of $R_1(D), \ldots, R_n(D)$, where each $R_i(D)$ is an A_i-relation, and A_i is the set of attributes of R_i.

A join query is of the form:

$$Q \quad := \quad R_1(A_1) \ \bowtie \ \cdots \ \bowtie \ R_n(A_n)$$

where A_i is the set of attributes of R_i. The schema of Q is $\{R_1, \ldots, R_m\}$, and the size of the query Q is $\sum_i |A_i|$. The set of attributes of Q is $A_Q := A_1 \cup \cdots \cup A_n$. For a database instance D of schema σ, the answer of Q on D is the A_Q-relation $Q(D)$ which consists of all the A_Q-tuple t such that $t_{A_i} \in R_i(D)$, for every $R_i(A_i)$.

References

[1] A. Atserias, M. Grohe, and D. Marx. Size bounds and query plans for relational joins. *SIAM Journal on Computing*, 42(4):1737–1767, 2013.

[1] That is, we only consider databases whose values are from the set of natural numbers $\mathbb{N}$.