

## Lesson 4: Datalog queries

**Theme:** Datalog queries and their evaluations.

**Syntax.** Let  $\tau$  be a vocabulary. A *datalog rule* is of the form:

$$S(\bar{x}) \leftarrow T_1(\bar{y}_1), \dots, T_m(\bar{y}_m)$$

where  $S, T_1, \dots, T_m$  are relation names (not necessarily from  $\tau$ ), the variables in  $\bar{x}$  come from  $\bar{y}_1, \dots, \bar{y}_m$ , and  $S$  is *not* from  $\tau$ .

The atom  $S(\bar{x})$  is called the *head* of the rule, and  $R_1(\bar{y}_1), \dots, R_k(\bar{y}_k)$  is called the *body*. A *datalog program* is a finite collection of datalog rules.

- The symbol  $P$  is usually reserved to denote a datalog program.
- An *intensional* relation in  $P$  is a relation that occurs as a head in a rule in  $P$ .
- An *extensional* relation is a relation that occurs only in the body of the rules in  $P$ .

**Semantics.** Let  $DB = (U, R_1, \dots, R_k)$  be a database over vocabulary  $\tau = \{R_1, \dots, R_k\}$ .

Let  $P$  be a datalog program with all the extensional relations from  $\tau$ . Let  $\{S_1, \dots, S_n\}$  be the intensional relations in  $P$ . A database  $DB' = (U, R_1, \dots, R_k, S_1, \dots, S_n)$  is the *result of  $P$  on  $DB$* , if it is the *smallest* database such that for every rule in  $P$ :

$$S(\bar{x}) \leftarrow T_1(\bar{y}_1), \dots, T_m(\bar{y}_m)$$

the following holds:

$$DB' \models \forall \bar{y}_1 \dots \forall \bar{y}_m \forall \bar{x} \left( T_1(\bar{y}_1) \wedge \dots \wedge T_m(\bar{y}_m) \rightarrow S_i(\bar{x}) \right)$$

The output relations of  $P$  are the relations  $S_1, \dots, S_n$  in  $DB'$ .

**Note.** For more details about Datalog programs, see [1, Chapters 12–15] and the references within.

## References

- [1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995. URL: <http://webdam.inria.fr/Alice/>.