

Course proposal for ESSLLI 2016

15–26 August, 2016

Title: Logical Foundations of Databases

Area: Logic & Computation

Level: Foundational (could possibly be upgraded to ‘Introductory’)

Lecturers:

Diego Figueira

Affiliation: CNRS, LaBRI, France

Email: diego.figueira@labri.fr

Homepage: <http://homepages.inf.ed.ac.uk/dfigueir/>

Gabriele Puppis

Affiliation: CNRS, LaBRI, France

Email: gabriele.puppis@labri.fr

Homepage: <http://www.labri.fr/perso/gpuppis/>

Abstract. The proposed course will present several fundamental results linking logics with database query languages. The course can thus serve as an introduction to finite model theory and database theory. More specifically, it will cover basic knowledge about first-order logic, Relational Algebra and Conjunctive Queries; complexity of fundamental decision problems, such as evaluation and satisfiability; as well as results about expressiveness of specification languages. Regarding the logical formalisms, we will focus mainly on two important query languages for relational databases, namely, Relational Algebra and Conjunctive Queries. We will not assume any prior knowledge on either logic or databases, but some familiarity with basic results in complexity theory will be required. The goal of the course is to give the fundamental tools that enable a deeper understanding of database query languages.

Motivation and description

Relational databases are arguably the most widely used database management systems. In these systems, data is logically organised into relations or ‘tables’. This relational data model has been a very successful model. Relational database theory has been intimately

linked to logic from its birth, since E. F. Codd's fundamental result in the 1970's establishing that Relational Algebra and the Predicate Calculus of First-order logic are essentially the same. Indeed, logics and query languages are so closely linked to the point that logics are now a standard starting point for developing database query languages. Thanks to these connections, several problems that are crucial to database applications (most notably, query evaluation, query optimisation, constraint satisfaction and entailment) can be rephrased and studied in a purely logical setting and they can thus profit from the knowledge and tools that are available nowadays in theoretical computer science. For example, an important branch of research in theoretical computer science, called finite model theory, emerged very early with the rise of database theory, and focuses on fundamental questions related to expressiveness and complexity of logics interpreted on finite structures.

This course aims at showing some important results in database theory following the logical perspective. Specifically, it will explain the links between logics and database query languages and give a number of tools originally developed in Finite Model Theory (e.g., Ehrenfeucht-Fraïssé games, 0-1 Law, Chandra-Merlin Lemma) that can be used to study the expressive power of query languages and the complexity of some fundamental problems.

Structure of the course and tentative schedule

The course will be given in English; however, if needed, questions can be posed and answered in French, Spanish, or Italian. The course will be structured into five lessons of about 90 minutes each.

Below is a tentative schedule of the covered subjects.

Day 1. First Order logic (FO) basics

- (a) First-order logic
- (b) Relational Algebra

Day 2. Evaluation and satisfiability

- (a) Data complexity, combined complexity
- (b) Complexity of model checking
- (c) Undecidability of satisfiability problem

Day 3. Expressiveness of FO

- (a) Ehrenfeucht-Fraïssé games
- (b) 0-1 Law
- (c) Hanf and Gaiffman locality

Day 4. Conjunctive Queries (CQ)

- (a) CQ = Select-Project-Join-Union fragment of Relational Algebra
- (b) CQ = Existential-Positive FO (EPFO)
- (c) Duality of CQ's and structures
- (d) Chandra-Merlin Lemma

Day 5. Complexity of problems on CQ

- (a) Evaluation problem
- (b) Containment and equivalence problems
- (c) Evaluation problem for bounded-width CQ

Requirements

The course will not require particular prior knowledge on logic or database theory. It will assume some familiarity with basic concepts and results in computation theory, e.g., halting problem, complexity classes P, NP, PSPACE, LOGSPACE. Nevertheless, all these concepts will be refreshed.

References

1. Serge Abiteboul, Richard Hull, Victor Vianu, Foundations of Databases. Addison Wesley. Available online at <http://webdam.inria.fr/Alice/>.
2. Leonid Libkin. Finite Model Theory. Springer, 2004.