TECHNISCHE UNIVERSITÄT DRESDEN

EXISTENTIAL RULES SEMINAR

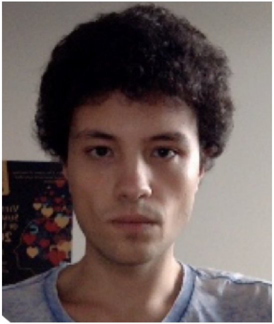
Lecture 1: Syntax and Semantics

David Carral, Markus Krötzsch

Knowledge-Based Systems

TU Dresden, 11th April 2018

Course Tutors



David Carral

Markus Krötzsch

Acknowledgements

Content

Michaël Thomazo and **Andreas Pieris**; Student Session at ESSLLI (more info at `http://esslli-stus-2015.phil.hhu.de/`)

Beamer Style

Markus Krötzsch

Tips on How to Run a Seminar

Lukas Schweizer

Organisation

Lectures

Wednesdays, DS 3 (11:10–12:40), APB E005

Web Page

`https://iccl.inf.tu-dresden.de/web/Existential_Rules_(SS2018)`

Lecture Notes

All slides will be available online.

Goals, Prerequisites, and Reading List

A good understanding of:

- The fundamentals of ontology-based query answering.
- State-of-the-art of research in the field.

(Non-)Prerequisites

- First-order logic (syntax and semantics).
- Complexity theory (complexity classes, reductions...).

Reading list

- Uwe Schöning: **Logic for Computer Scientists**; Birkhäuser.
- Michael Sipser: **Introduction to the Theory of Computation, International Edition**; Cengage Learning.

Structure of the Seminar and Evaluation

Lectures

- **Wednesday 11th (i.e., today):** Introductory lecture 1
- **Wednesday 18th:** Introductory lecture 2
- **Afterwards:** Office hours in 3035 and presentations

Evaluation

- **Paper summary:** self-selected research paper;[a] 15 pages
- **Presentation:** 45 minutes + discussion

[a] See the “Literature” tab at `https://iccl.inf.tu-dresden.de/web/Existential_Rules_(SS2018)`.

Motivation: Accessing Big Data

"Data is stored in various **heterogeneous** formats over many differently structured databases. As a result, the gathering of only relevant data spread over **disparate sources** becomes a very **time consuming task**." – Jim Crompton, W3C Workshop on Semantic Web in Oil & Gas Industry, 2008

More info at: `http://www.expertsystem.com/semantic-web-in-oil-gas-industry/`

Motivation: Accessing Big Data

Experts in geology and geophysics develop stratigraphic models of unexplored areas on the basis of data acquired from previous operations at nearby geographical locations.

Facts:

- 1000 TB of relational data
- Using diverse schemata
- Spread over 2000 tables, over multiple individual data bases

A Possible Solution

- Achieve transparency in accessing data using **logic** – e.g., **existential rules**!
- Manage data by exploiting Knowledge Representation techniques.
- Provide a conceptual, high level representation of the domain of interest of terms of an **ontology** (i.e., a logical theory).

A Simple Example

Example 1.1:

$$HasSon(x, y) \wedge HasSister(y, z) \rightarrow HasDaughter(x, z) \quad (1)$$

$$HasFather(x, y) \rightarrow HasSon(y, x) \quad (2)$$

$$Person(x) \rightarrow \exists y.HasFather(x, y) \quad (3)$$

$$Person(anakin) \quad (4)$$

$$HasFather(luke, anakin) \quad (5)$$

$$HasSister(luke, leia) \quad (6)$$

Is *leia* the daughter of *anakin*?

I.e., does $HasDaughter(anakin, leia)$ follow from (1-6)?

Syntax: Signature and Atoms

- A **signature** is a tuple $\langle \mathsf{P}, \mathsf{V}, \mathsf{C}, \mathsf{N} \rangle$ with P a set of **predicates**, V a set of **variables**, C a set of **constants**, and N a set of **nulls**.
- P, V, C, and N are mutually disjoint sets.
- Every predicate $P \in \mathsf{P}$ is associated to some arity $\mathsf{ar}(P) \geq 1$.
- $\mathsf{T} = \mathsf{V} \cup \mathsf{C} \cup \mathsf{N}$ is the set of **terms**.
- An **atom** is a formula of the form $P(\vec{t})$ with $P \in \mathsf{P}$, $\mathsf{ar}(P) = |\vec{t}|$, and $t \in \mathsf{T}$ for all $t \in \vec{t}$.

Example 1.2: Entities and atoms.

$$Person(x) \to \exists y.HasFather(x, y)$$

$$HasSister(luke, leia)$$

Syntax: Existential Rules

Definition 1.3: An **(existential) rule** is a formula of the form

$$\forall \vec{x}, \vec{z}.(\beta[\vec{x}, \vec{z}] \to \exists \vec{y}.\eta[\vec{x}, \vec{y}])$$

with β and η conjunctions of null-free atoms and $\vec{x}$, $\vec{y}$, and $\vec{z}$ mutually disjoint sequences of variables. A **fact** is a rule with an empty body that contains no occurrences of variables.

Formulas (1-6) from slide 9 are existential rules.
Formulas (4-6) are also facts.

Semantics: Interpretations

Definition 1.4: An **interpretation** $\mathcal{I}$ is a tuple $\langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ where

- $\Delta^{\mathcal{I}}$ is a non-empty set of domain elements, and
- $\cdot^{\mathcal{I}}$ is a function mapping all $c \in \mathsf{C}$ to some $c^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ and all $P \in \mathsf{P}$ to some $P^{\mathcal{I}} \subseteq (\Delta^{\mathcal{I}})^{\mathsf{ar}(P)}$.

Intuitively, an **interpretation** may be regarded as a **possible state of the world**. A **logical theory** – in our case, a rule set – is a set of conditions that restrict the valid states of the world.

Semantics: Models

Definition 1.5: A **(variable) assignment** $\mathcal{Z}$ for an interpretation $\mathcal{I}$ is a partial function $\mathcal{Z} : \mathsf{T} \to \Delta^{\mathcal{I}}$ with $Z(c) = c^{\mathcal{I}}$ for all $c \in \mathsf{C}$. Given a conjunction β of atoms, we write $\mathcal{I}, \mathcal{Z} \models \beta$ if $\langle \mathcal{Z}(\vec{t}) \rangle \in P^{\mathcal{I}}$ for all $P(\vec{t}) \in \beta$. Given a rule $\rho = \beta[\vec{x}, \vec{z}] \to \exists \vec{y}.\eta$, we write $\mathcal{I}, \mathcal{Z} \models \rho$ if

1. $\mathcal{I}, \mathcal{Z} \not\models \beta$ or
2. $\mathcal{I}, \mathcal{W} \models \eta$ for some assignment $\mathcal{W} \supseteq \mathcal{Z}$.

We write $\mathcal{I} \models \rho$ if $\mathcal{I}, \mathcal{Z} \models \rho$ for every assignment $\mathcal{Z}$ defined over $\vec{x} \cup \vec{z} \cup \mathsf{C}$. Interpretation $\mathcal{I}$ is a **model** of a rule set R if $\mathcal{I} \models \rho$ for all $\rho \in \mathsf{R}$.

Models are those interpretations that satisfy a particular logical theory – in our case, a given rule set. I.e., models are **valid states of the world** according to some logical theory.

Semantics: Queries and Entailment

Definition 1.6: A **query** is a formula of the form $\exists \vec{y}.\beta$ with β a conjunction of null-free atoms.

E.g., $q = \exists x.HasFather(anakin, x)$, and $q' = HasSister(leia, anakin)$.

Definition 1.7: A query q is **entailed** by an interpretation $\mathcal{I}$ if $\mathcal{I}, \mathcal{Z} \models \beta$ for some assignment $\mathcal{Z}$. A rule set R **entails** a query if $\mathcal{M} \models q$ for every model $\mathcal{M}$ of R.

Query q is entailed by R = $\{(1), \ldots, (6)\}$ whilst query q' is not.

Solving CQE: The two dimensions of infinity.

To determine if a query q is entailed by a rule set R, we have to determine that $\mathcal{M} \models q$ for every model $\mathcal{M}$ of R. Alas, this is not easy!

1. R may accept an infinite number of models.
2. Each one models may be of infinite size.

To address (1), we introduce the notion of **universal models** which can be used to solve conjunctive query entailment independently.

Brief recap

- Syntax
- Semantic
- Conjunctive query entailment

What's next?

- Universal models
- The chase algorithm