universiteit hasselt

KNOWLEDGE IN ACTION

On the Satisfiability Problem for SPARQL Patterns

Jan Van den Bussche
(Hasselt University, Belgium)

joint work with Xiaowang Zhang and François Picalausa

Journal of Artificial Intelligence Research 56 (2016) 403–428

Relational Databases and Queries

Friend

john	mary
john	paul
mary	anne
anne	paul
john	eve

Hobby

john	pingpong
mary	birding
mary	running
anne	pingpong
paul	running

A **query** computes a new relation from the given relations

Typically expressed in first-order logic (FO)

"List pairs of friends of John with a common hobby"

$$\{(x, y) \mid \exists z : \text{Friend}(\texttt{john}, x) \wedge \text{Friend}(\texttt{john}, y) \wedge x \neq y \wedge \text{Hobby}(x, z) \wedge \text{Hobby}(y, z)\}$$

Friend

john	mary
john	paul
mary	anne
anne	paul
john	eve

Hobby

john	pingpong
mary	birding
mary	running
anne	pingpong
paul	running

$$\{(x, y) \mid \exists z : \mathsf{Friend}(\mathtt{john}, x) \wedge \mathsf{Friend}(\mathtt{john}, y) \wedge x \neq y \wedge \mathsf{Hobby}(x, z) \wedge \mathsf{Hobby}(y, z)\}$$

Answer

x	y
mary	paul
paul	mary

RDF

Simplification of relational databases to facilitate exchanging, collecting, combining, data on the Web

One single, ternary, relation T

T

john	friend	mary
john	friend	paul
mary	friend	anne
anne	friend	paul
john	friend	eve
john	hobby	pingpong
mary	hobby	birding
mary	hobby	running
anne	hobby	pingpong
paul	hobby	running

Querying RDF

T

john	friend	mary
john	friend	paul
mary	friend	anne
anne	friend	paul
john	friend	eve
john	hobby	pingpong
mary	hobby	birding
mary	hobby	running
anne	hobby	pingpong
paul	hobby	running

$$\{(x, y) \mid \exists z : T(\texttt{john}, \texttt{friend}, x) \wedge T(\texttt{john}, \texttt{friend}, y) \wedge x \neq y \wedge T(x, \texttt{hobby}, z) \wedge T(y, \texttt{hobby}, z)\}$$

Practical problems with standard FO semantics

Disjunction:

$$\{(x, y) \mid T(\texttt{john}, \texttt{friend}, x) \vee T(\texttt{anne}, \texttt{friend}, y)\}$$

When x is a friend of John, y can be **anything**

Answer

x	y
mary	john
mary	friend
mary	pingpong
mary	...
paul	...
eve	...
...	paul

Alternative semantics: use partial assignments

$$\{(x, y) \mid T(\texttt{john}, \texttt{friend}, x) \vee T(\texttt{anne}, \texttt{friend}, y)\}$$

Answer

x	y
mary	$\perp$
paul	$\perp$
eve	$\perp$
$\perp$	paul

Conjunction

The semantics of conjunction becomes natural join ($\bowtie$)

$$T(\texttt{john}, \texttt{friend}, x) \wedge T(x, \texttt{hobby}, z)$$

x
mary
paul

$\bowtie$

x	z
john	pingpong
mary	birding
mary	running
anne	pingpong
paul	running

$=$

x	z
mary	birding
mary	running
paul	running

(The semantics of existential quantification becomes projection)

Negation

$$\{x \mid \neg T(\texttt{john}, \texttt{friend}, x)\}$$

Again x can be **anything** (except John's friends)

x
john
anne
friend
hobby
pingpong
...

Alternative: **disallow** negation altogether

Web data is open-ended, so, anyway,
negation is not very appropriate

Alternative: Optional matching (OPT)

"List friends of John and their hobbies, if any"

$$T(\texttt{john}, \texttt{friend}, x) \,\mathsf{OPT}\, T(x, \texttt{hobby}, z)$$

x	z
mary	birding
mary	running
paul	running
eve	$\perp$

OPT, in conjunction with undefinedness tests, can **simulate** negation

"List friends of John without any recorded hobby"

$$(T(\texttt{john}, \texttt{friend}, x) \,\mathsf{OPT}\, T(x, \texttt{hobby}, z)) \wedge z = \perp$$

SPARQL and the satisfiability problem

"SPARQL": FO on a ternary relation
with discussed adaptations

- partial assignments
- OPT
- no negation
- undefinedness tests

Satisfiability problem: given a SPARQL expression E, does there exist a ternary relation on which E returns a nonempty result?

For FO this is undecidable

For SPARQL too

Our results on SPARQL satisfiability

Undecidable fragments:

1. allowing constant equalities ($x = \texttt{eve}$)

2. allowing equalities ($x = y$) in conjunction with nonequalities ($x \neq y$)

Negation can still be simulated

All other cases are decidable

3. in NP, and NP-hard already for quantifier-free formulas using just T-atoms without constants, $\vee$, $\wedge$, $x \neq \bot$

"Well-designed" expressions:

4. PTIME