

From Horn-*SRIQ* to Datalog:
A Data-Independent Transformation
that Preserves Assertion Entailment

David Carral, Larry González, and Patrick Koopmann



**TECHNISCHE
UNIVERSITÄT
DRESDEN**

Introduction

Syntax

Horn-*SRIQ*

Datalog

Formulas

TBox Axioms	
$C_1 \sqcap \dots \sqcap C_n \sqsubseteq D$	$C \sqsubseteq \exists R.D$
$\exists R.C \sqsubseteq D$	$C \sqsubseteq \leq 1R.D$
$R_1 \circ \dots \circ R_n \sqsubseteq S$	$R^- \sqsubseteq S$

Rules
$P_1(\vec{x}_1) \wedge \dots \wedge P_n(\vec{x}_n) \rightarrow Q(\vec{y})$

ABox Axioms	
$C(a)$	$R(a, b)$

Facts
$P(\vec{c})$

Theories

Ontologies
$\mathcal{O} = (\mathcal{T}, \mathcal{F})$

Programs
$\mathcal{P} = (\mathcal{R}, \mathcal{F})$

From Horn-*SRIQ* to Datalog

Definition. A rule set \mathcal{R} is an **AR-rewriting** for a TBox \mathcal{T} iff, for all fact sets \mathcal{F} ,

- * the ontology $(\mathcal{T}, \mathcal{F})$ and the program $(\mathcal{R}, \mathcal{F})$ are equi-satisfiable and,
- * for all facts α over the signature of \mathcal{T} , $(\mathcal{T}, \mathcal{F})$ entails α iff $(\mathcal{R}, \mathcal{F})$ entails α .

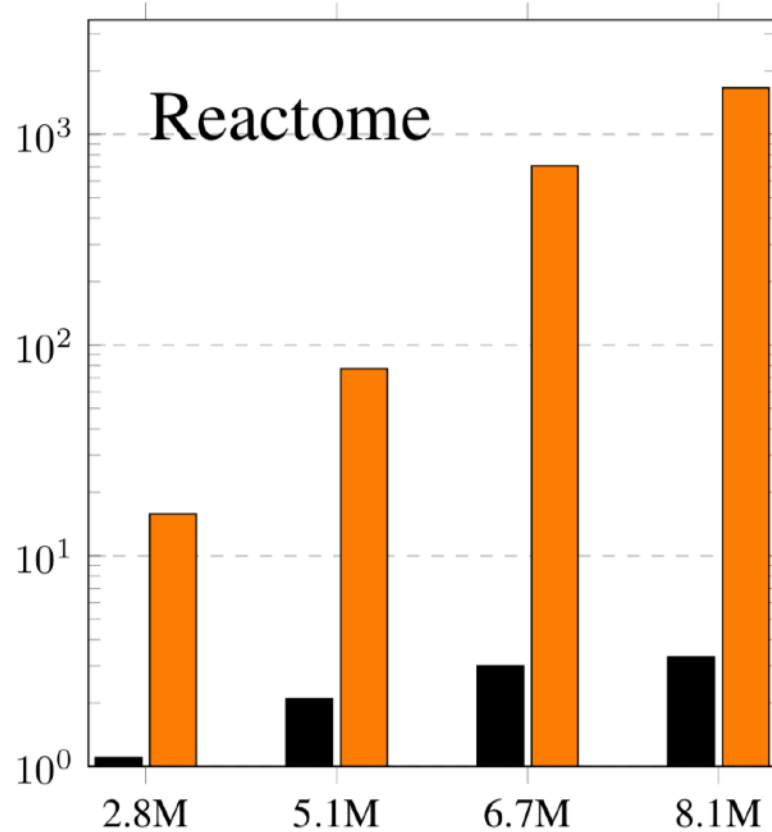
Can we compute **AR-rewritings**?

- * Reasoning in Description Logics by a Reduction to Disjunctive Datalog. Hustadt, Motik, and Sattler. In Journal of Autom. Reasoning 2007.
- * The Combined Approach to Query Answering in Horn-*ALCHOIQ*. Carral, Dragoste, and Krötzsch. In KR 2018.

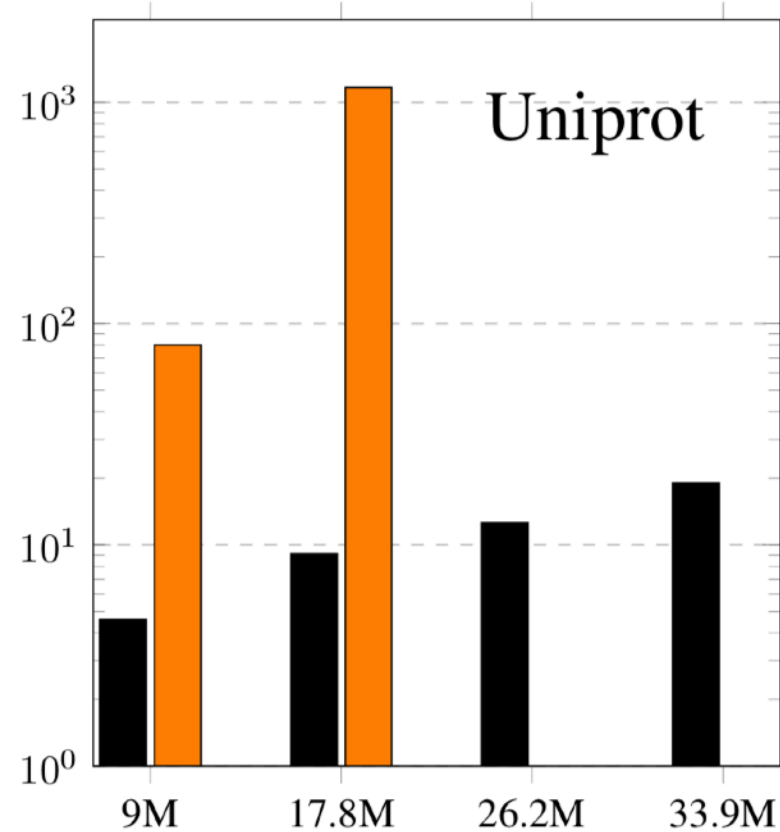
What about Horn-*SRIQ*? **Yes!** Wait... but why is this interesting?

Evaluation

Reasoning with Rewritings



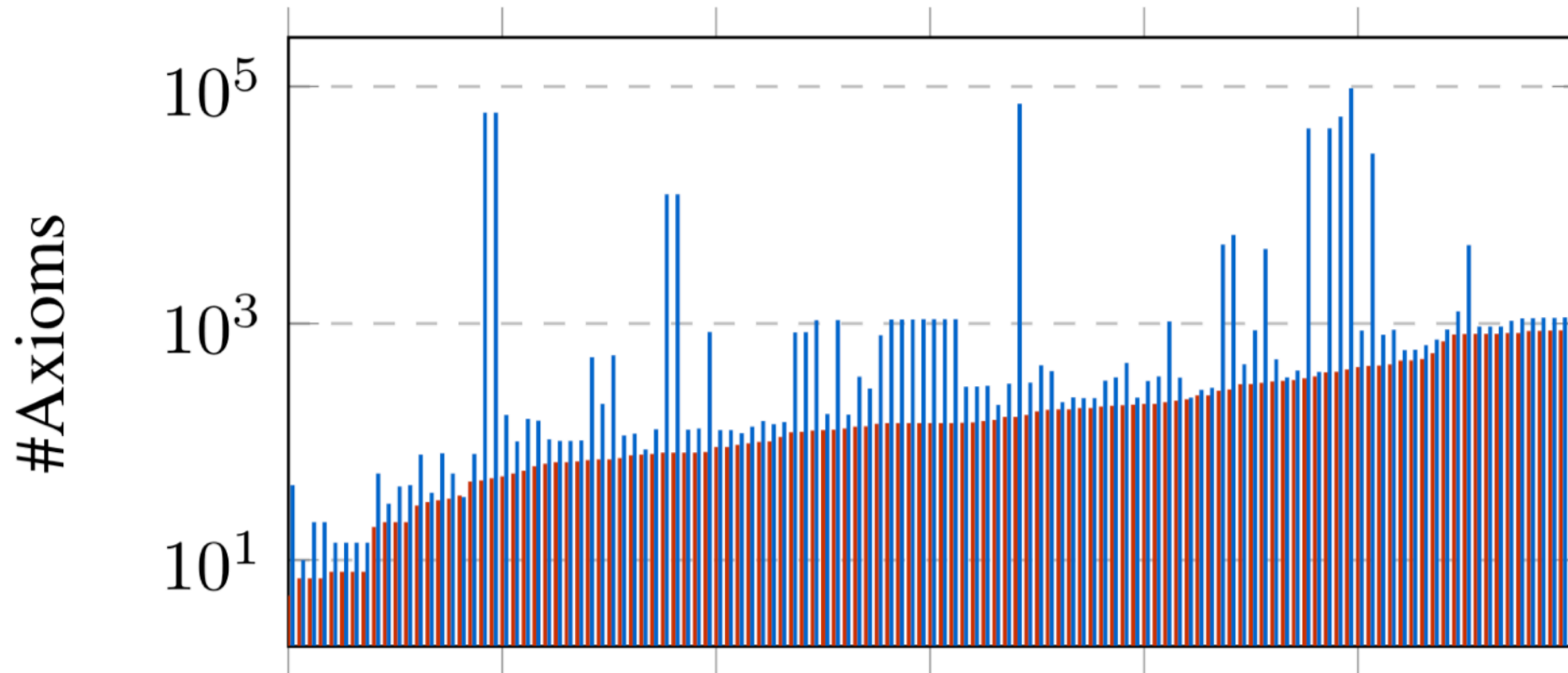
TBox size: 485
Rewriting size: 549
Time: 221s



TBox size: 304
Rewriting size: 367
Time: 182s



Size of Rewritings



- MOWLCorpus: TBoxes with less 1000 axioms and containing role chain axioms
- 187 TBoxes: 121 computed rewritings w/o OOM errors

From Horn-*ALCHIQ* to Datalog

$$R_1 \circ \dots \circ R_n \sqsubseteq S \quad \rightarrow \quad R \sqsubseteq S$$

Forest Model Property

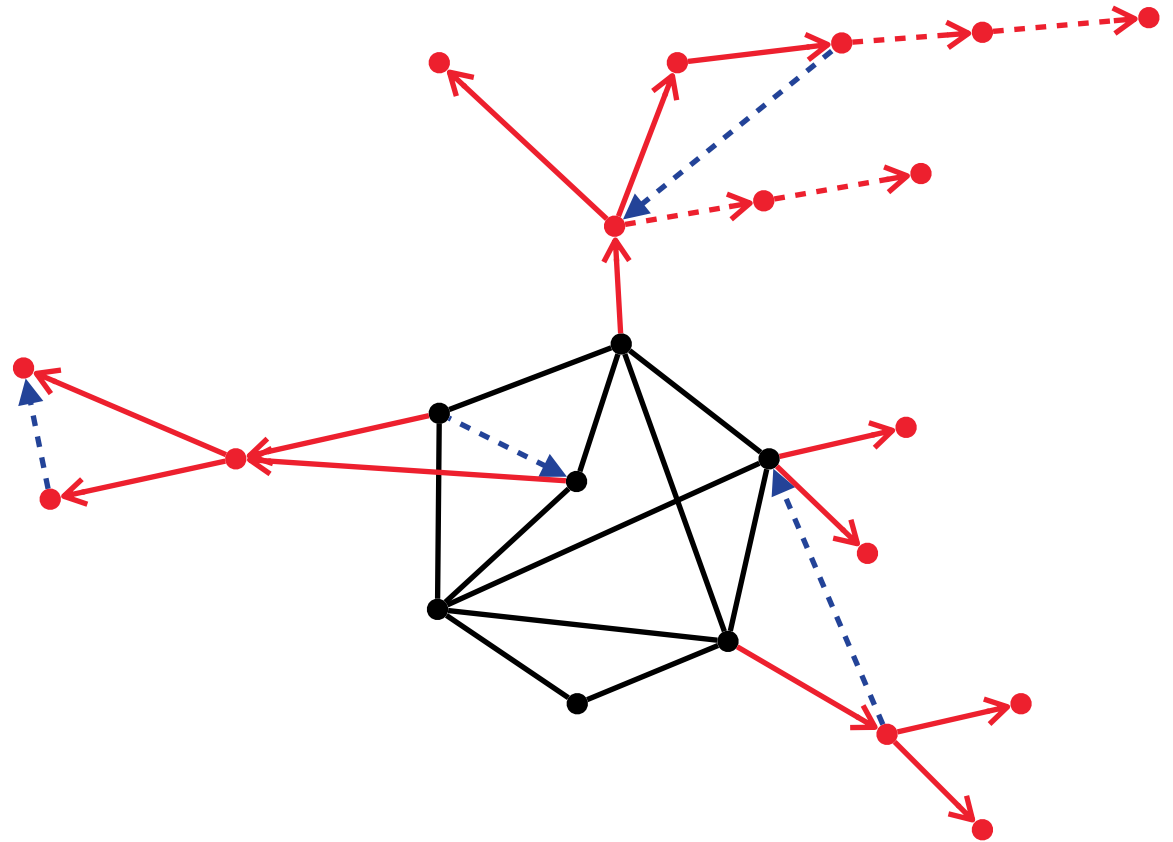
$$C_1 \sqcap \dots \sqcap C_n \sqsubseteq D$$

$$\exists R. C \sqsubseteq D$$

$$C \sqsubseteq \exists R. D$$

$$C \sqsubseteq \leq 1 R. D$$

$$R \sqsubseteq S$$



“Unnamed-to-Named” Consequences

Successor-to-predecessor

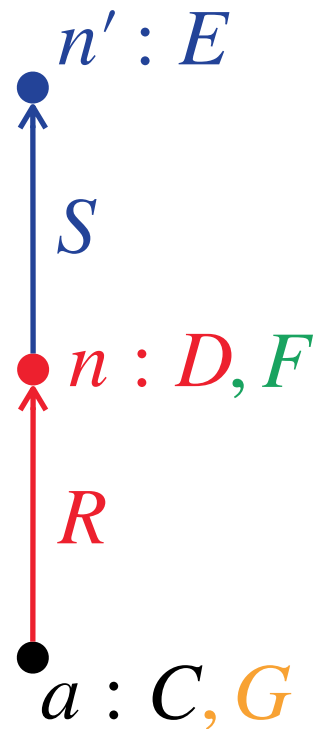
$$C \sqsubseteq \exists R.D$$

$$D \sqsubseteq \exists S.E$$

$$\exists S.E \sqsubseteq F$$

$$\exists R.F \sqsubseteq G$$

$$C(x) \rightarrow G(x)$$



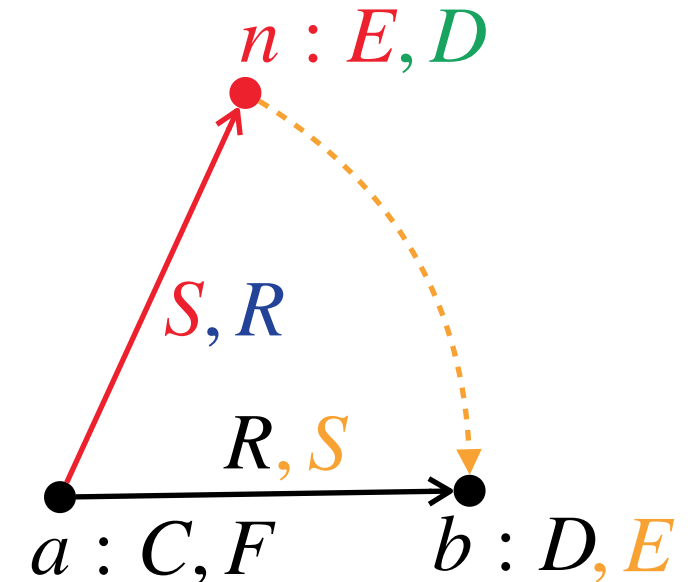
Folding

$$C \sqsubseteq \exists S.E$$

$$S \sqsubseteq R$$

$$E \sqsubseteq D$$

$$F \sqsubseteq \leq 1R.D$$



$$C(x) \wedge F(x) \wedge R(x, y) \wedge D(y) \rightarrow S(x, y)$$

$$C(x) \wedge F(x) \wedge R(x, y) \wedge D(y) \rightarrow E(y)$$

Computing AR-Rewritings for Horn-*ALCHIQ*

Definition. Consider some Horn-*ALCHIQ* TBox \mathcal{T} .

The rule set $\mathcal{R}_{\mathcal{T}}$, which is an AR-preserving rewriting for \mathcal{T} , is defined as follows :

1. For all $C \sqsubseteq \forall R.D \in \mathcal{T}$,
 $C(x) \wedge R(x, y) \rightarrow D(y) \in \mathcal{R}_{\mathcal{T}}$
2. For all $R \sqsubseteq S \in \mathcal{T}$,
 $R(x, y) \rightarrow S(x, y) \in \mathcal{R}_{\mathcal{T}}$
3. For all $R^- \sqsubseteq S \in \mathcal{T}$,
 $R(y, x) \rightarrow S(x, y) \in \mathcal{R}_{\mathcal{T}}$
4. For all $C_1 \sqcap \dots \sqcap C_n \sqsubseteq D \in \Omega(\mathcal{T})$, $C_1(x) \wedge \dots \wedge C_n(x) \rightarrow D(x) \in \mathcal{R}_{\mathcal{T}}$ **Successor-to-predecessor**
5. For all $C \sqsubseteq \leq 1R.D \in \mathcal{T}$,
 $C(x) \wedge R(x, y) \wedge D(y) \wedge R(x, z) \wedge D(z) \rightarrow y \approx z \in \mathcal{R}_{\mathcal{T}}$, **Folding**
 $C(x) \wedge C_1(x) \wedge \dots \wedge C_n(x) \wedge R(x, y) \wedge D(y) \rightarrow E(y) \in \mathcal{R}_{\mathcal{T}}$ if $C_1 \sqcap \dots \sqcap C_n \sqsubseteq \exists R.(D \sqcap E) \in \Omega(\mathcal{T})$, and
 $C(x) \wedge C_1(x) \wedge \dots \wedge C_n(x) \wedge R(x, y) \wedge D(y) \rightarrow S(x, y) \in \mathcal{R}_{\mathcal{T}}$ if $C_1 \sqcap \dots \sqcap C_n \sqsubseteq \exists(R \sqcap S).D \in \Omega(\mathcal{T})$

Definition. $\Omega(\mathcal{T})$ is the set of all axioms of either of the following forms entailed by \mathcal{T} .

$$C_1 \sqcap \dots \sqcap C_n \sqsubseteq D$$

$$C_1 \sqcap \dots \sqcap C_n \sqsubseteq \exists(R_1 \sqcap \dots \sqcap R_m).(D_1 \sqcap \dots \sqcap D_k)$$

Remarks

- * $\mathcal{R}_{\mathcal{T}}$ is exponential in \mathcal{T}
- * Compute $\Omega(\mathcal{T})$ using consequence-based

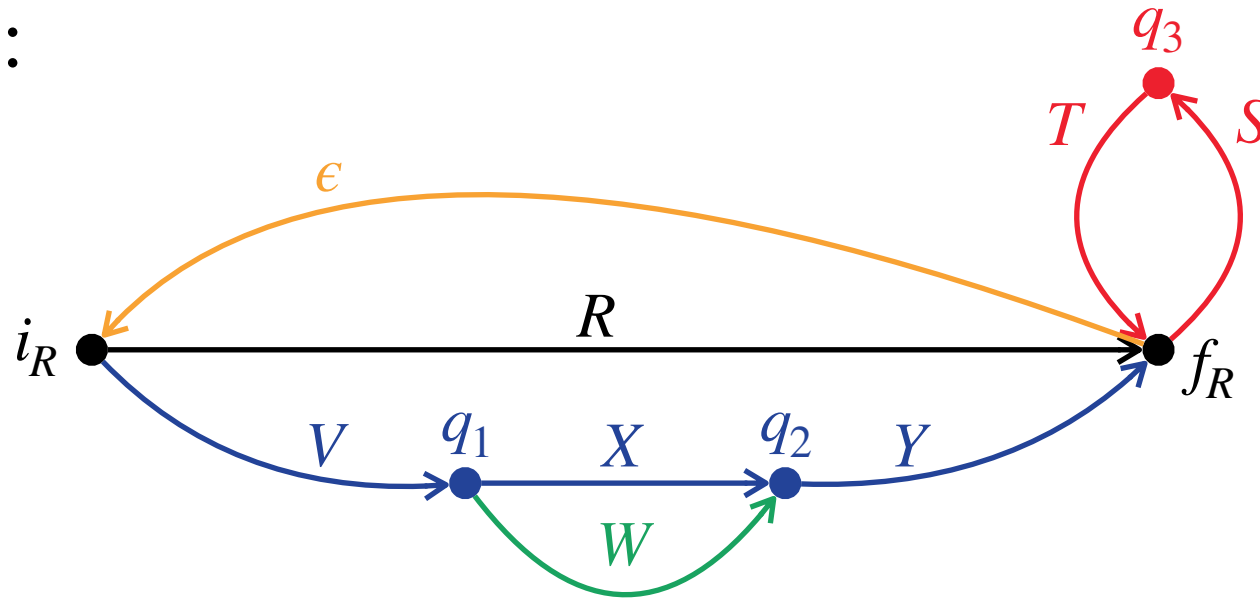
Query Rewriting for Horn-*SHIQ* plus Rules.
 Eiter, Ortiz, Simkus, Tran, and Xiao. In AAI 2012.

From Horn-*SRIQ* to Datalog

Complex Roles and NFA

$$\mathcal{T} = \{V \circ X \circ Y \sqsubseteq R, R \circ S \circ T \sqsubseteq R, W \sqsubseteq X, R \circ R \sqsubseteq R\}$$

$\mathcal{N}_{\mathcal{T}}(R)$:



Box Pushing

$$A \sqsubseteq \forall R . B \in \mathcal{T}$$

$$BP(\mathcal{T}) \supseteq \mathcal{T} \cup \{$$

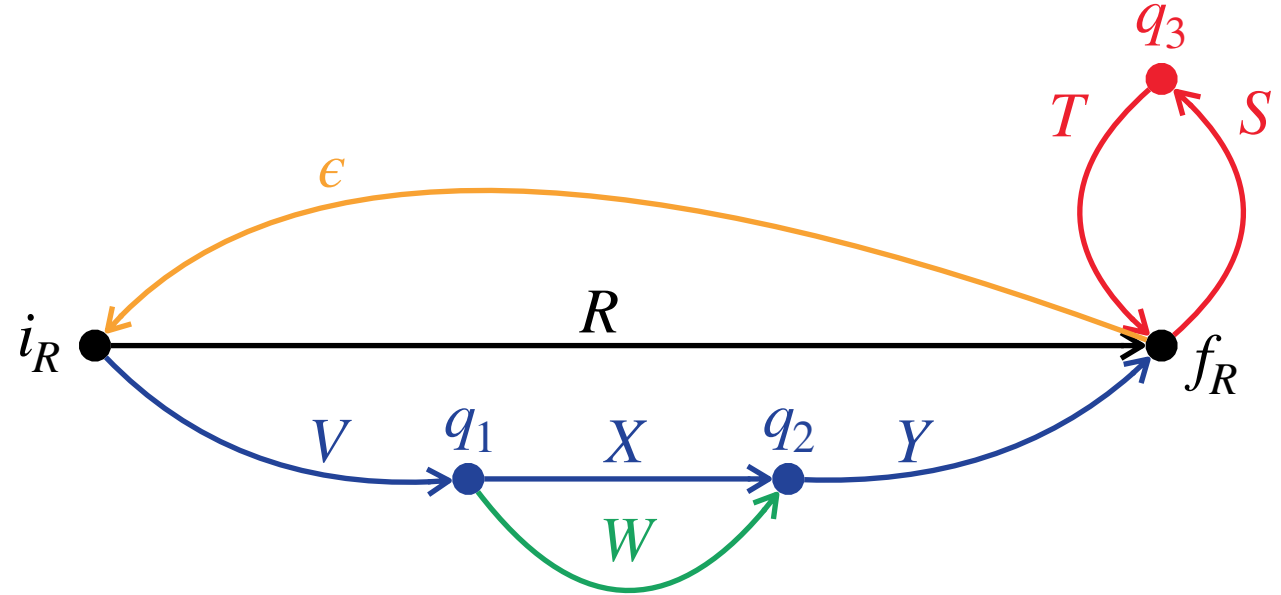
$$A \sqsubseteq B_{i_R}, B_{f_R} \sqsubseteq B,$$

$$B_{i_R} \sqsubseteq \forall R . B_{f_R},$$

$$B_{f_R} \sqsubseteq \forall S . B_{q_3}, B_{q_3} \sqsubseteq \forall T . B_{f_R},$$

$$B_{i_R} \sqsubseteq \forall V . B_{q_1}, B_{q_1} \sqsubseteq \forall X . B_{q_2}, B_{q_2} \sqsubseteq \forall Y . B_{f_R},$$

$$B_{q_1} \sqsubseteq \forall W . B_{q_2}, B_{f_R} \sqsubseteq B_{i_R} \}$$



Computing “AR-Rewritings” for Horn-*SRIQ*

Definition. Consider some Horn-*SRIQ* TBox \mathcal{T} .

Remarks

- * \mathcal{T}' is kind of an “AR-rewriting” for \mathcal{T} , but only for class assertions!
- * \mathcal{T}' is a Horn-*ALCHIQ* TBox

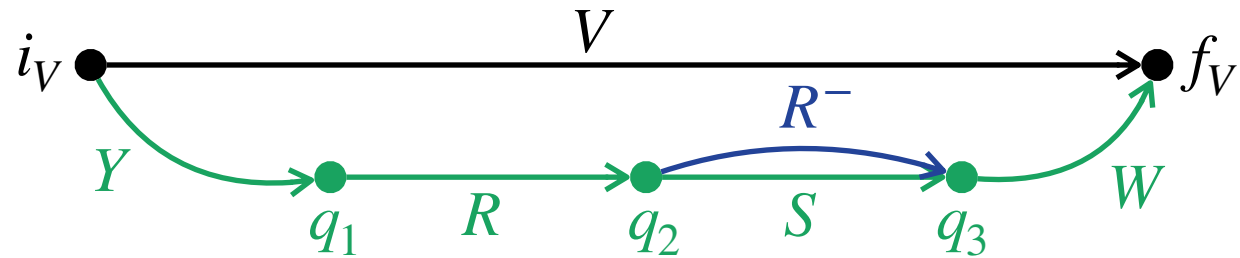
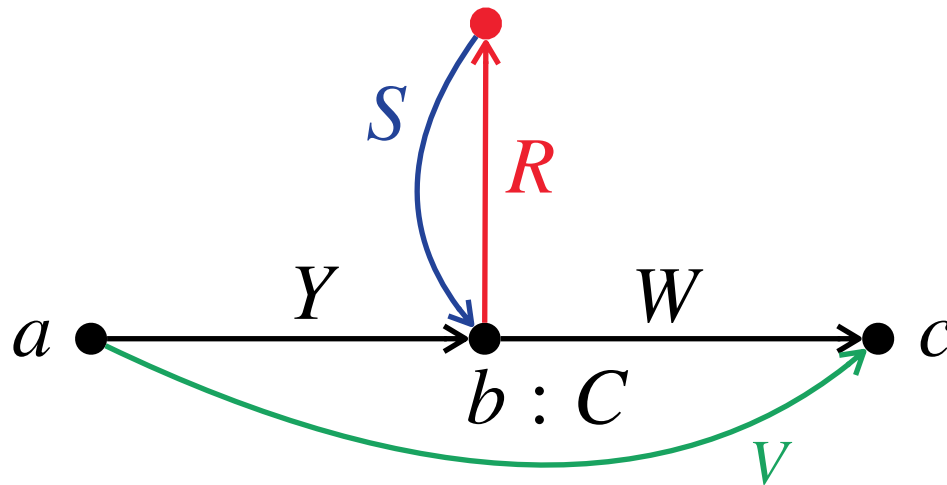
“Unnamed-to-Named” Role Consequences

Unnamed Paths

$$C \sqsubseteq \exists R. T$$

$$R^- \sqsubseteq S$$

$$Y \circ R \circ S \circ W \sqsubseteq V$$



$$Y(x, y) \rightarrow V_{q_1}(x, y)$$

$$V_{q_2}(x, y) \wedge S(y, z) \rightarrow V_{q_3}(x, z)$$

$$V_{q_1}(x, y) \wedge R(y, z) \rightarrow V_{q_2}(x, z)$$

$$V_{q_3}(x, y) \wedge W(y, z) \rightarrow V_{f_V}(x, z)$$

$$C(x) \rightarrow V_{q_1, q_3}(x)$$

$$V_{q_1}(x, y) \wedge V_{q_1, q_3}(y) \rightarrow V_{q_3}(x, z)$$

$$V_{f_V}(x, y) \rightarrow V(x, y)$$

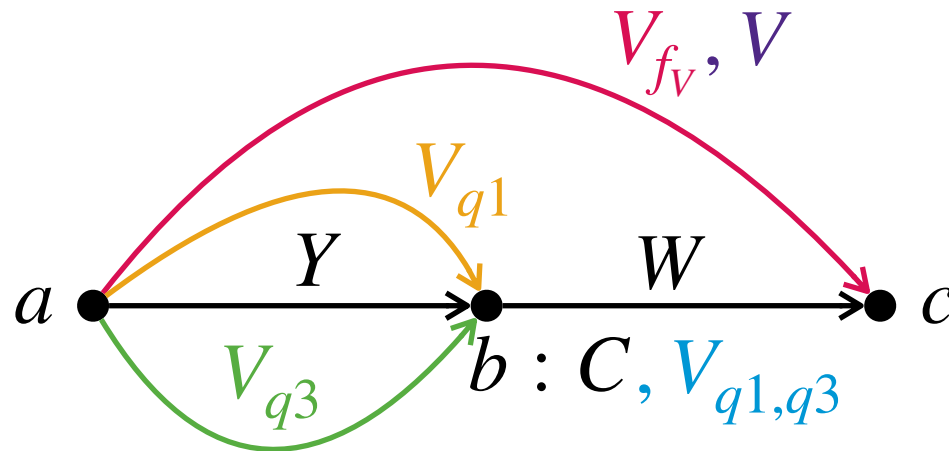
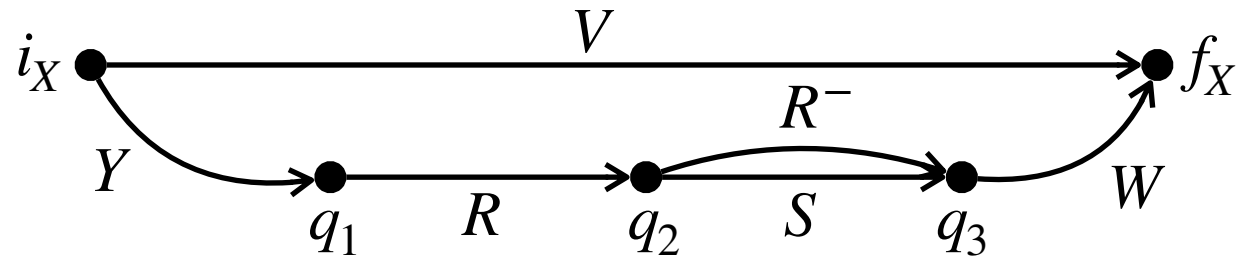
“Unnamed-to-Named” Role Consequences

Unnamed Paths

$$C \sqsubseteq \exists R . \top$$

$$R^- \sqsubseteq S$$

$$Y \circ R \circ S \circ W \sqsubseteq V$$



$$Y(x, y) \rightarrow V_{q_1}(x, y)$$

$$V_{q_2}(x, y) \wedge S(y, z) \rightarrow V_{q_3}(x, z)$$

$$V_{q_1}(x, y) \wedge R(y, z) \rightarrow V_{q_2}(x, z)$$

$$V_{q_3}(x, y) \wedge W(y, z) \rightarrow V_{f_V}(x, z)$$

$$C(x) \rightarrow V_{q_1, q_3}(x)$$

$$V_{q_1}(x, y) \wedge V_{q_1, q_3}(y) \rightarrow V_{q_3}(x, z)$$

$$V_{f_V}(x, y) \rightarrow V(x, y)$$

Computing AR-Rewritings for Horn-*SRIQ*

Definition. Consider some Horn-*SRIQ* TBox \mathcal{T} .

The rule set $\mathcal{R}_{\mathcal{T}}$, which is an AR-preserving rewriting for \mathcal{T} , is defined as follows :

- A. Let $\mathcal{T}_+ = \mathcal{T} \cup \{X \sqsubseteq \forall R. Y \mid R \text{ a role in } \mathcal{T}\}$ where X and Y are fresh class names.
- B. Let \mathcal{T}_\times be the TBox that results from extending \mathcal{T}_+ with all axioms obtained via "box pushing" and then removing every axiom with role chains.
- C. Add all of the rules in the AR-rewriting of \mathcal{T}_\times to $\mathcal{R}_{\mathcal{T}}$ (computed as shown in previous slides).
- D. For all roles R in \mathcal{T} , all states q and q' in $\mathcal{N}_{\mathcal{T}}(R)$, and all sets of concepts C_1, \dots, C_n ,
if $C_1 \sqcap \dots \sqcap C_n \sqcap Y_q \sqsubseteq Y_{q'} \in \Omega(\mathcal{T}_\times)$, then add $C_1(x) \wedge \dots \wedge C_n(x) \rightarrow R_{q,q'}(x) \in \mathcal{R}_{\mathcal{T}}$. Unnamed paths
- E. For all roles R occurring in \mathcal{T} ,
for all transitions $i_R \rightarrow_S^* q \in \mathcal{N}_{\mathcal{T}}(R)$ with i_R the initial state, add $S(x, y) \rightarrow R_q(x, y) \in \mathcal{R}_{\mathcal{T}}$,
for all states q in $\mathcal{N}_{\mathcal{T}}(R)$, add $R_{i_R, q}(x) \rightarrow R_q(x, x) \in \mathcal{R}_{\mathcal{T}}$,
for all transitions $q \rightarrow_S^* q' \in \mathcal{N}_{\mathcal{T}}(R)$, add $R_q(x, y) \wedge S(y, z) \rightarrow R_{q'}(x, z) \in \mathcal{R}_{\mathcal{T}}$,
for all states q and q' in $\mathcal{N}_{\mathcal{T}}(R)$, add $R_q(x, y) \wedge R_{q,q'}(y) \rightarrow R_{q'}(x, y) \in \mathcal{R}_{\mathcal{T}}$, and
add $R_{f_R}(x, y) \rightarrow R(x, y)$ with f_R the final state.

Conclusion

Summary

Title: From Horn-*SRIQ* to Datalog: A Data-Independent Transformation that Preserves Assertion Entailment

Authors: David Carral, Larry González, and Patrick Koopmann

Affiliation: TU Dresden

Contributions:

- * Theoretical: method to compute AR-rewritings for Horn-*SRIQ*
- * Practical: the use of rewritings results in performance gains; we can compute AR-rewritings for many real-world TBoxes

Future Work:

- * Develop AR-rewritings for more expressive DLs; consider different target and input languages for these rewritings
- * Optimise implementation to produce rewritings of smaller size

JOIN US IN DRESDEN

- Open position available immediately
- Ph.D. students & postdocs
- 100% English & International
- KR / Databases / AI

CONTACT

Markus Krötzsch markus.kroetzsch@tu-dresden.de