From Horn-$\mathcal{SRIQ}$ to Datalog: A Data-Independent Transformation that Preserves Assertion Entailment

David Carral, Larry González, and Patrick Koopmann

TECHNISCHE UNIVERSITÄT DRESDEN

Introduction

Syntax

	Horn-$\mathcal{SRIQ}$	Datalog
Formulas	TBox Axioms $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq D$ $\quad C \sqsubseteq \exists R . D$ $\exists R . C \sqsubseteq D$ $\quad C \sqsubseteq\ \leq 1 R . D$ $R_1 \circ \ldots \circ R_n \sqsubseteq S$ $\quad R^- \sqsubseteq S$	Rules $P_1(\overrightarrow{x}_1) \wedge \ldots \wedge P_n(\overrightarrow{x}_n) \rightarrow Q(\overrightarrow{y})$
	ABox Axioms $C(a)$ $\quad R(a, b)$	Facts $P(\overrightarrow{c})$
Theories	Ontologies $\mathcal{O} = (\mathcal{T}, \mathcal{F})$	Programs $\mathcal{P} = (\mathcal{R}, \mathcal{F})$

From Horn-$\mathcal{SRIQ}$ 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 $\boldsymbol{\alpha}$ over the signature of $\mathcal{T}$, $(\mathcal{T}, \mathcal{F})$ entails $\boldsymbol{\alpha}$ iff $(\mathcal{R}, \mathcal{F})$ entails $\boldsymbol{\alpha}$.

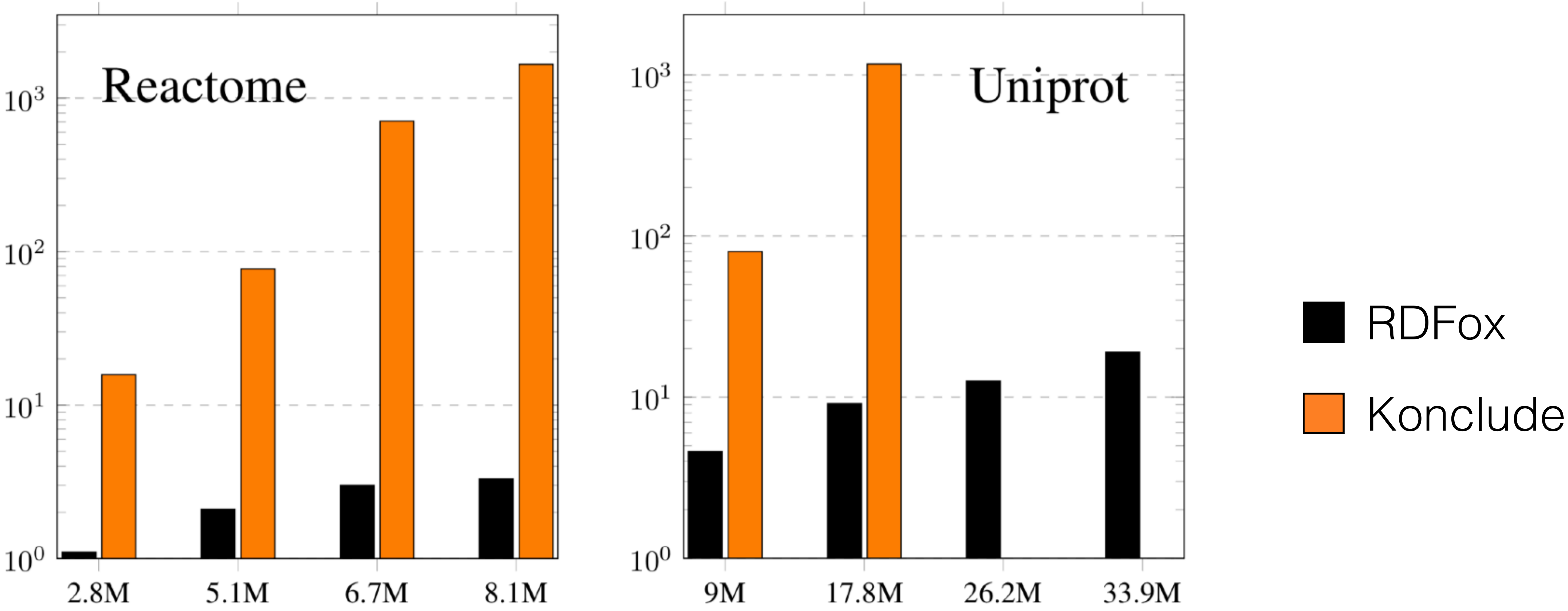
Can we compute **AR-rewritings**?

- 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-$\mathcal{ALCHOIQ}$. Carral, Dragoste, and Krötzsch. In KR 2018.

What about Horn-$\mathcal{SRIQ}$? Yes! Wait… but why is this interesting?

Evaluation

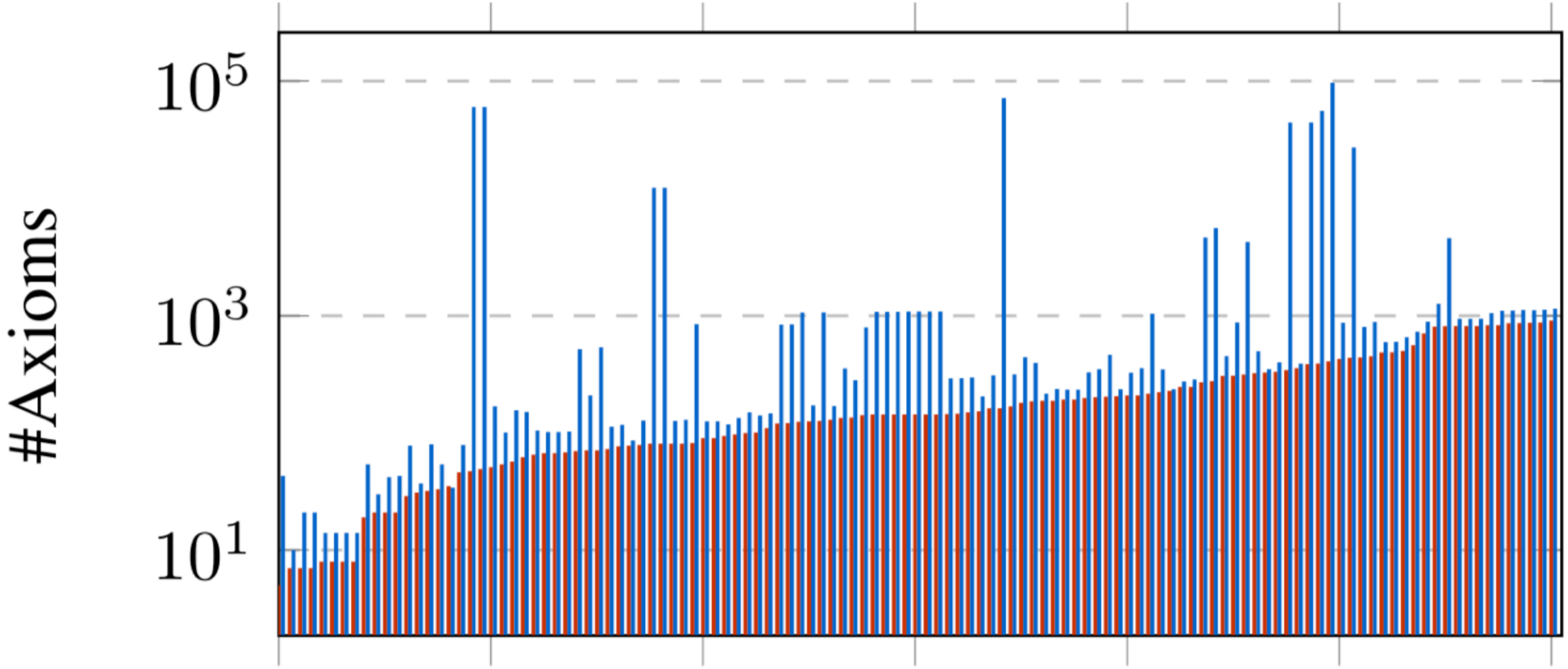
Reasoning with Rewritings



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-$\mathcal{ALCHIQ}$ to Datalog

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

Forest Model Property

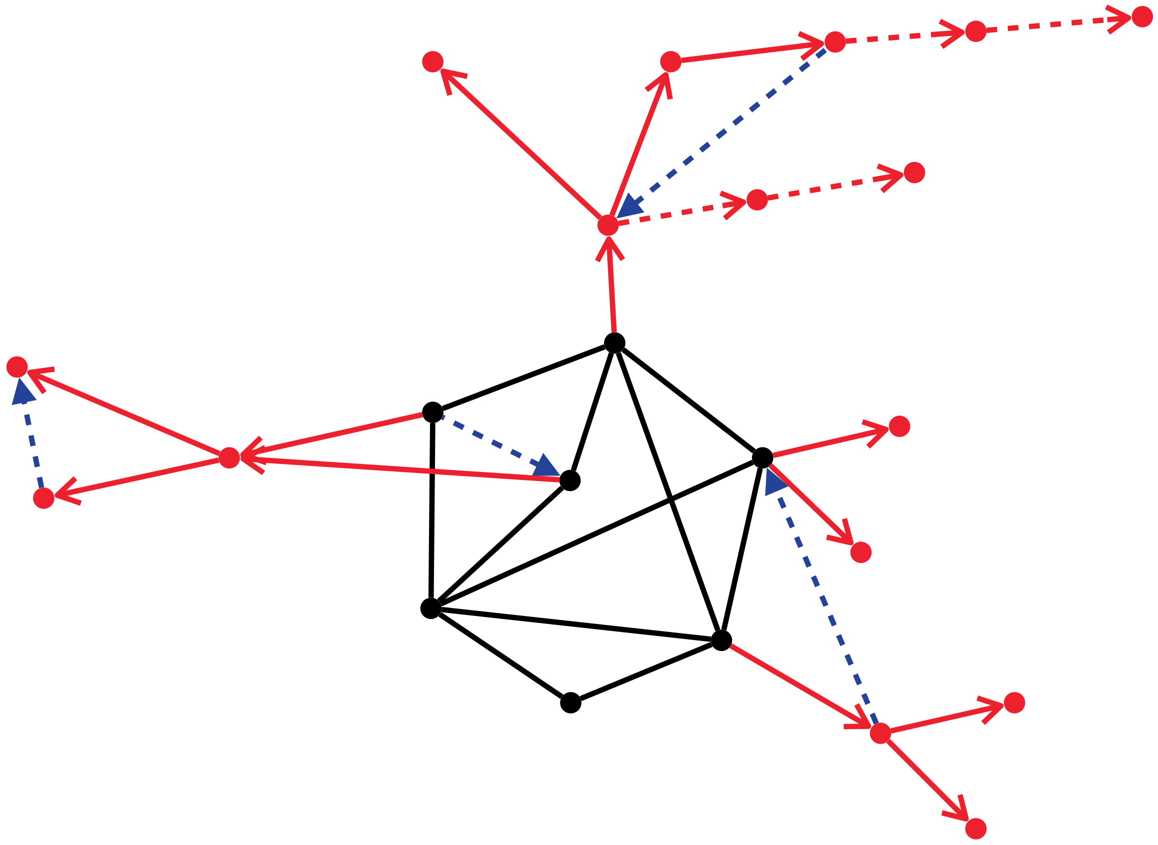
$C_1 \sqcap \ldots \sqcap C_n \sqsubseteq D$

$\exists R.C \sqsubseteq D$

$C \sqsubseteq \exists R.D$

$C \sqsubseteq \leq 1R.D$

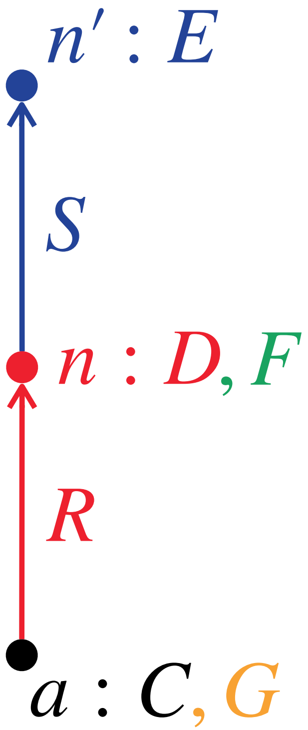
$R \sqsubseteq S$



"Unnamed-to-Named" Consequences

Successor-to-predecessor

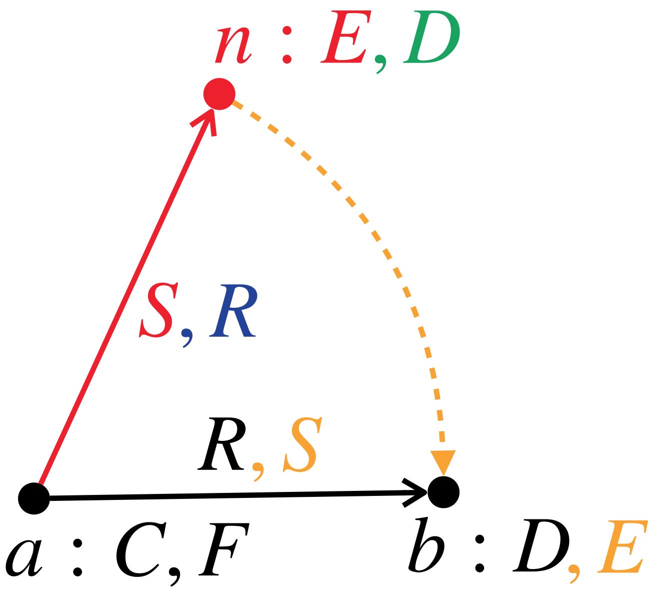
$C \sqsubseteq \exists R \,.\, D$
$D \sqsubseteq \exists S \,.\, E$
$\exists S \,.\, E \sqsubseteq F$
$\exists R \,.\, F \sqsubseteq G$



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

Folding

$C \sqsubseteq \exists S \,.\, E$
$S \sqsubseteq R$
$E \sqsubseteq D$
$F \sqsubseteq \leq 1 R \,.\, D$



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

Computing AR-Rewritings for Horn-$\mathcal{ALCHIQ}$

Definition. Consider some Horn-$\mathcal{ALCHIQ}$ 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 $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq D \in \Omega(\mathcal{T})$, $C_1(x) \wedge \ldots \wedge C_n(x) \rightarrow D(x) \in \mathcal{R}_{\mathcal{T}}$ — 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
   $C(x) \wedge C_1(x) \wedge \ldots \wedge C_n(x) \wedge R(x,y) \wedge D(y) \rightarrow E(y) \in \mathcal{R}_{\mathcal{T}}$ if $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq \exists R.(D \sqcap E) \in \Omega(\mathcal{T})$, and
   $C(x) \wedge C_1(x) \wedge \ldots \wedge C_n(x) \wedge R(x,y) \wedge D(y) \rightarrow S(x,y) \in \mathcal{R}_{\mathcal{T}}$ if $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq \exists(R \sqcap S).D \in \Omega(\mathcal{T})$

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

$C_1 \sqcap \ldots \sqcap C_n \sqsubseteq D$

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

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

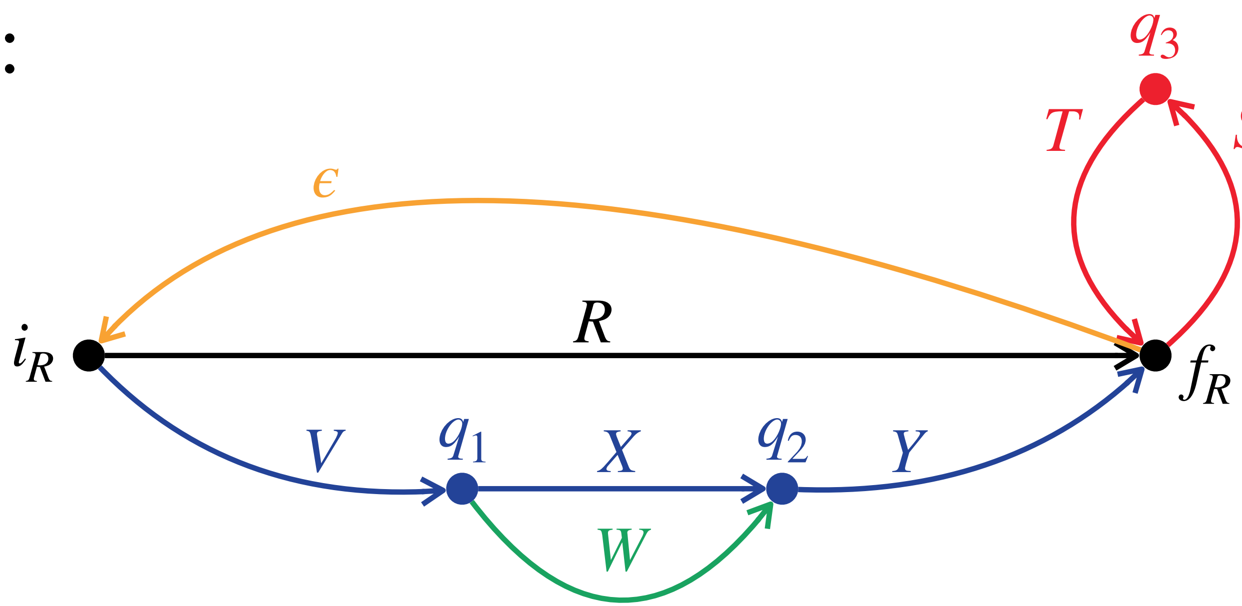
Query Rewriting for Horn-$\mathcal{SHIQ}$ plus Rules.
Eiter, Ortiz, Simkus, Tran, and Xiao. In AAAI 2012.

From Horn-$\mathcal{SRIQ}$ to Datalog

Complex Roles and NFA

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

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



Box Pushing

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

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

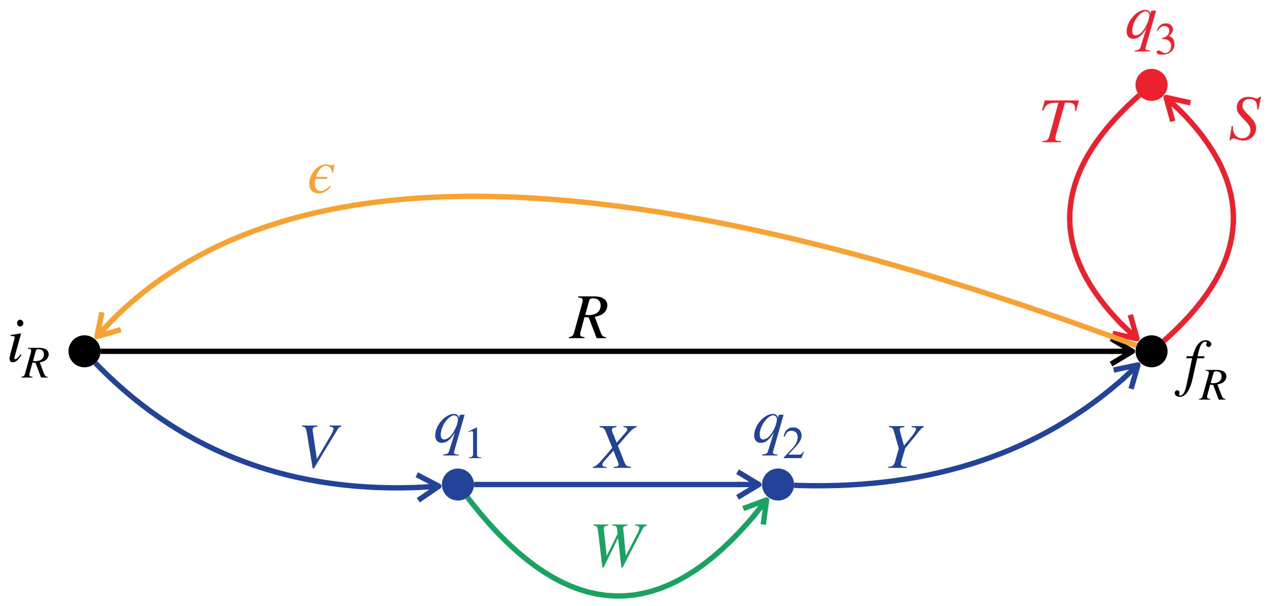
$$A \sqsubseteq B_{i_R}, B_{f_R} \sqsubseteq B,$$

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

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

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

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



Computing “AR-Rewritings” for Horn-$\mathcal{SRIQ}$

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

1. For all roles R in $\mathcal{T}$, compute the NFA $\mathcal{N}_{\mathcal{T}}(R)$.
2. Compute the TBox $\mathcal{T}'$ which results from adding all the axioms obtained via "box pushing", and then removing all axioms with role chains .
3. Compute the AR-rewriting $\mathcal{R}_{\mathcal{T}'}$ for the TBox $\mathcal{T}'$ (as defined in previous slides).
4. The rule set $\mathcal{R}_{\mathcal{T}'}$ can be used to solve class retrieval "in place" of $\mathcal{T}$.

Remarks

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

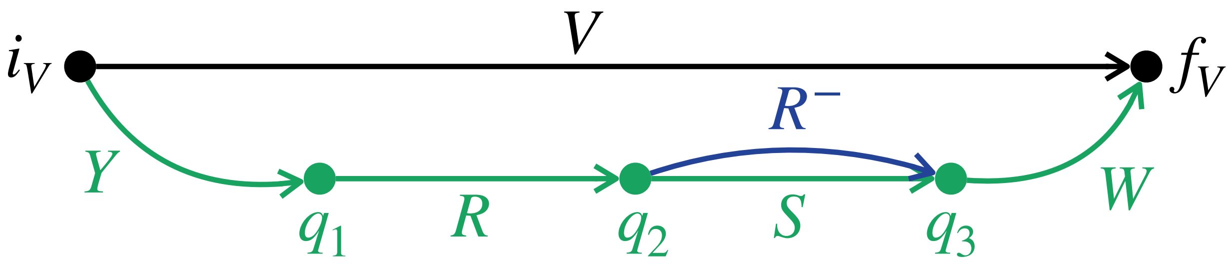
“Unnamed-to-Named” Role Consequences

Unnamed Paths

$C \sqsubseteq \exists R.\top$

$R^{-} \sqsubseteq S$

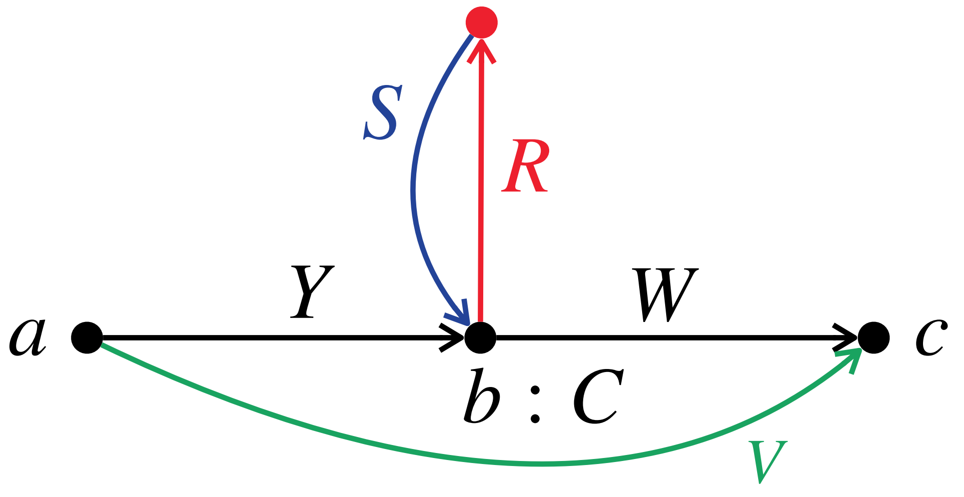
$Y \circ R \circ S \circ W \sqsubseteq V$



$$Y(x,y) \rightarrow V_{q_1}(x,y)$$
$$V_{q_2}(x,y) \wedge S(y,z) \rightarrow V_{q_3}(x,z)$$
$$V_{q_1}(x,y) \wedge R(y,z) \rightarrow V_{q_2}(x,z)$$
$$V_{q_3}(x,y) \wedge W(y,z) \rightarrow V_{f_V}(x,z)$$

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

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



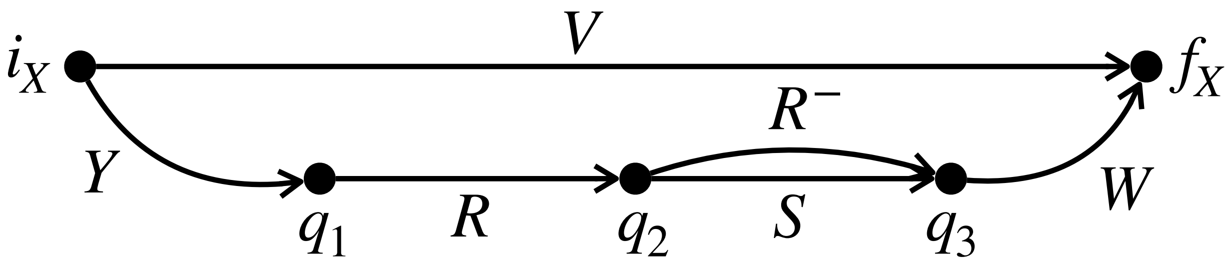
“Unnamed-to-Named” Role Consequences

Unnamed Paths

$C \sqsubseteq \exists R . \top$

$R^{-} \sqsubseteq S$

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



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

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

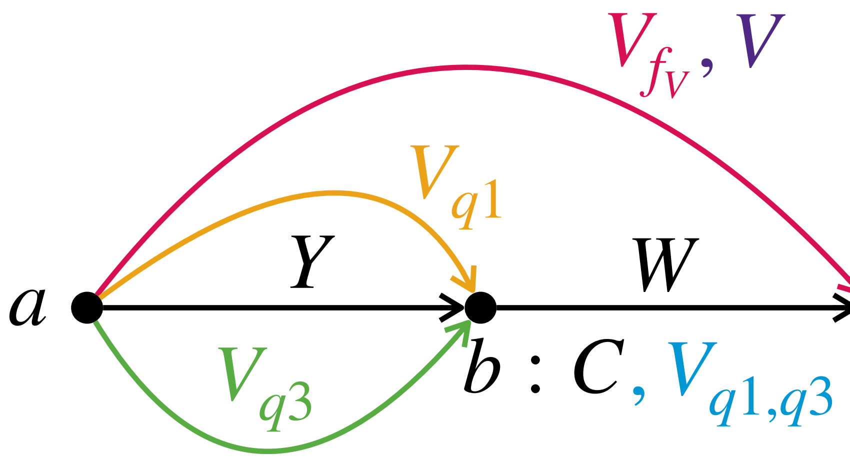
$V_{q_1}(x, y) \wedge R(y, z) \rightarrow V_{q_2}(x, z)$

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

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

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

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



Computing AR-Rewritings for Horn-$\mathcal{SRIQ}$

Definition. Consider some Horn-$\mathcal{SRIQ}$ 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 $\mathcal{T}_{+} = \mathcal{T} \cup \{X \sqsubseteq \forall R\,.\,Y \mid R$ 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 $C_1, \ldots, C_n$, if $C_1 \sqcap \ldots \sqcap C_n \sqcap Y_q \sqsubseteq Y_{q'} \in \Omega(\mathcal{T}_{\times})$, then add $C_1(x) \wedge \ldots \wedge C_n(x) \rightarrow R_{q,q'}(x) \in \mathcal{R}_{\mathcal{T}}$.

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-$\mathcal{SRIQ}$ 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-$\mathcal{SRIQ}$
- 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