



Goal Directed Tracing of Inferences in OWL EL Ontologies

Yevgeny Kazakov, Pavel Klinov

University of Ulm, Germany

Summary of this talk

A new technique for **inference tracing** in OWL EL
(a lightweight language for large terminologies):

- ▶ finding **all** inference steps used in **a proof** of a particular class subsumption ($A \sqsubseteq B$)
- ▶ doing it **fast**

The method can be used with **arbitrary rule-based procedures** but the main ideas are demonstrated for **EL** reasoning

Implemented **ELK 0.5.0+**, works well in practice

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$\mathbf{R}_0 \frac{}{C \sqsubseteq C} \quad \mathbf{R}_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad \mathbf{R}_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\mathbf{R}_{\top} \frac{}{C \sqsubseteq \top} \quad \mathbf{R}_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad \mathbf{R}_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$R_0 \vdash A \sqsubseteq A$$

$$R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \stackrel{R_{\sqsubseteq}}{\vdash} A \sqsubseteq \exists R.B$$

$$R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B$$

$$\begin{array}{c} \vdash \\ \mathbf{R}_0 \end{array} B \sqsubseteq B$$

$$\mathbf{R}_0 \frac{}{C \sqsubseteq C} \quad \mathbf{R}_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad \mathbf{R}_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\mathbf{R}_{\top} \frac{}{C \sqsubseteq \top} \quad \mathbf{R}_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad \mathbf{R}_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B$$

$$B \sqsubseteq B \xrightarrow{\mathbf{R}_{\sqsubseteq}} B \sqsubseteq D$$

$$\mathbf{R}_0 \frac{}{C \sqsubseteq C} \quad \mathbf{R}_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad \mathbf{R}_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\mathbf{R}_{\top} \frac{}{C \sqsubseteq \top} \quad \mathbf{R}_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad \mathbf{R}_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$\begin{array}{ccc}
 A \sqsubseteq A & A \sqsubseteq \exists R.B & \begin{array}{c} \mathbf{R}_{\exists} \\ \vdash \\ \hline \end{array} A \sqsubseteq \exists R.D \\
 & B \sqsubseteq B & B \sqsubseteq D
 \end{array}$$

$$\mathbf{R}_0 \frac{}{C \sqsubseteq C} \quad \mathbf{R}_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad \mathbf{R}_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\mathbf{R}_{\top} \frac{}{C \sqsubseteq \top} \quad \mathbf{R}_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad \mathbf{R}_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B \quad A \sqsubseteq \exists R.D \stackrel{R \sqsubseteq}{\mapsto} A \sqsubseteq C$$

$$B \sqsubseteq B \quad B \sqsubseteq D$$

$$R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B \quad A \sqsubseteq \exists R.D \quad A \sqsubseteq C$$

$$B \sqsubseteq B \quad B \sqsubseteq D \quad \xrightarrow{R_{\exists}} \quad B \sqsubseteq \exists S.A$$

$$R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$\begin{array}{cccc}
 A \sqsubseteq A & A \sqsubseteq \exists R.B & A \sqsubseteq \exists R.D & A \sqsubseteq C \\
 & B \sqsubseteq B & B \sqsubseteq D & B \sqsubseteq \exists S.A \quad \downarrow \text{R}_{\exists} \quad B \sqsubseteq \exists S.C
 \end{array}$$

$$\text{R}_0 \frac{}{C \sqsubseteq C} \quad \text{R}_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad \text{R}_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\text{R}_{\top} \frac{}{C \sqsubseteq \top} \quad \text{R}_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad \text{R}_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B \quad A \sqsubseteq \exists R.D \quad A \sqsubseteq C$$

$$B \sqsubseteq B \quad B \sqsubseteq D \quad B \sqsubseteq \exists S.A \quad B \sqsubseteq \exists S.C \xrightarrow{R_{\exists}} B \sqsubseteq C$$

$$R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B \quad A \sqsubseteq \exists R.D \quad A \sqsubseteq C$$

$$B \sqsubseteq B \quad B \sqsubseteq D \quad B \sqsubseteq \exists S.A \quad B \sqsubseteq \exists S.C \quad B \sqsubseteq C$$

Explain: how was $B \sqsubseteq C$ derived?

Consequence-based reasoning in \mathcal{EL} : example

$$\mathcal{O} = \{A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C\}$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B \quad A \sqsubseteq \exists R.D \quad A \sqsubseteq C$$

$$B \sqsubseteq B \quad B \sqsubseteq D \quad B \sqsubseteq \exists S.A \quad B \sqsubseteq \exists S.C \quad B \sqsubseteq C$$

Explain: how was $B \sqsubseteq C$ derived?

Closure under the \mathcal{EL} rules contains

- ▶ enough information on **what** is derived
- ▶ no information on **how** or **why** it is derived

Explanations are important (e.g. for debugging), hard and require **tool support**

Approaches to explanations

Justifications: minimal subsets of \mathcal{O} which entail the axiom

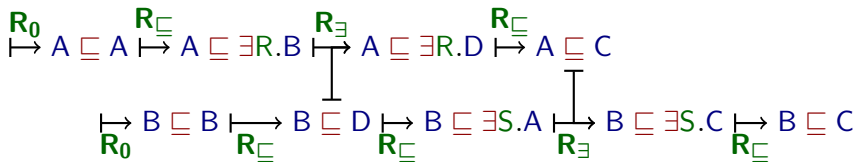
- ▶ don't provide information on **how** the result follows
- ▶ **intractable** even for very simple languages (in the **worst** case)
often work fine in practice

Approaches to explanations

Justifications: minimal subsets of \mathcal{O} which entail the axiom

- ▶ don't provide information on **how** the result follows
- ▶ **intractable** even for very simple languages (in the **worst** case)
often work fine in practice

Proof-based explanations present steps made by the reasoning procedure to derive the result



Methods do **not** have to be tied to a particular inference system

Problem statement

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

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

$B \sqsubseteq B$ $B \sqsubseteq D$ $B \sqsubseteq \exists S.A$ $B \sqsubseteq \exists S.C$ $B \sqsubseteq C$

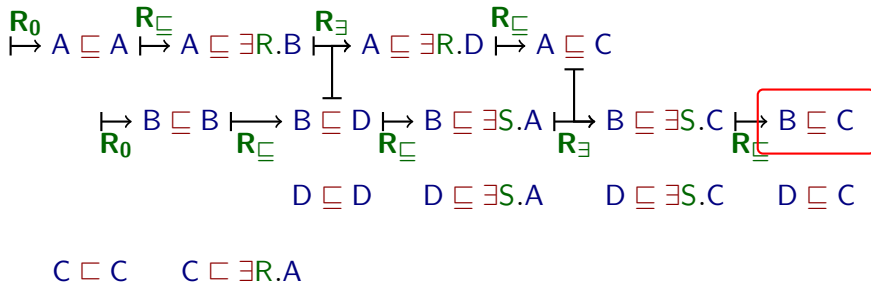
$D \sqsubseteq D$ $D \sqsubseteq \exists S.A$ $D \sqsubseteq \exists S.C$ $D \sqsubseteq C$

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

How to **re-construct** proofs of $B \sqsubseteq C$ with **minimal** effort?
(e.g. without **full tracing**)

Problem statement

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$



How to **re-construct** proofs of $B \sqsubseteq C$ with **minimal** effort?
 (e.g. without **full tracing**)

Goal-directed tracing: main ideas

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

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

$B \sqsubseteq B$ $B \sqsubseteq D$ $B \sqsubseteq \exists S.A$ $B \sqsubseteq \exists S.C$ $B \sqsubseteq C$

$D \sqsubseteq D$ $D \sqsubseteq \exists S.A$ $D \sqsubseteq \exists S.C$ $D \sqsubseteq C$

Goal-directed tracing: main ideas

$$A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C, \quad C \sqsubseteq \exists R.A$$

$$A \sqsubseteq A \quad A \sqsubseteq \exists R.B \quad A \sqsubseteq \exists R.D \quad A \sqsubseteq C$$

$$B \sqsubseteq B \quad B \sqsubseteq D \quad B \sqsubseteq \exists S.A \quad B \sqsubseteq \exists S.C \quad B \sqsubseteq C$$

$$D \sqsubseteq D \quad D \sqsubseteq \exists S.A \quad D \sqsubseteq \exists S.C \quad D \sqsubseteq C$$

$$\begin{array}{l}
 R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{\boxed{C} \sqsubseteq D_1 \sqcap D_2}{\boxed{C} \sqsubseteq D_1 \quad \boxed{C} \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{\boxed{C} \sqsubseteq D}{\boxed{C} \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O} \\
 R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{\boxed{C} \sqsubseteq D_1 \quad \boxed{C} \sqsubseteq D_2}{\boxed{C} \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{\boxed{E} \sqsubseteq \exists R.C \quad C \sqsubseteq D}{\boxed{E} \sqsubseteq \exists R.D}
 \end{array}$$

Goal-directed tracing: main ideas

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

$A \sqsubseteq A$

$A \sqsubseteq \exists R.B$

$A \sqsubseteq \exists R.D$

$A \sqsubseteq C$

$\vdash B \sqsubseteq B$
 R_0

$\vdash B \sqsubseteq D$
 R_{\sqsubseteq}

$\vdash B \sqsubseteq \exists S.A$
 R_{\sqsubseteq}

$\vdash B \sqsubseteq \exists S.C$
 R_{\exists}

$\vdash B \sqsubseteq C$
 R_{\sqsubseteq}

$D \sqsubseteq D$

$D \sqsubseteq \exists S.A$

$D \sqsubseteq \exists S.C$

$D \sqsubseteq C$

To find inferences for $B \sqsubseteq C$ it is sufficient to re-apply (and record) inferences applicable to subsumptions of the form $B \sqsubseteq X$

Goal-directed tracing: main ideas

$$A \sqsubseteq \exists R.B, \quad B \sqsubseteq D, \quad D \sqsubseteq \exists S.A, \quad \exists R.D \sqsubseteq C, \quad \exists S.C \sqsubseteq C, \quad C \sqsubseteq \exists R.A$$

 $A \sqsubseteq A$
 $A \sqsubseteq \exists R.B$
 $A \sqsubseteq \exists R.D$
 $A \sqsubseteq C$
 \vdash_{R_0}
 $B \sqsubseteq B$
 $\vdash_{R_{\sqsubseteq}}$
 $B \sqsubseteq D$
 $\vdash_{R_{\sqsubseteq}}$
 $B \sqsubseteq \exists S.A$
 $\vdash_{R_{\exists}}$
 $B \sqsubseteq \exists S.C$
 $\vdash_{R_{\sqsubseteq}}$
 $B \sqsubseteq C$
 $D \sqsubseteq D$
 $D \sqsubseteq \exists S.A$
 $D \sqsubseteq \exists S.C$
 $D \sqsubseteq C$

To find inferences for $B \sqsubseteq C$ it is sufficient to re-apply (and record) inferences applicable to subsumptions of the form $B \sqsubseteq X$

Other subsumptions can be **used**, e.g., $A \sqsubseteq C$, but not **derived**

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

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

$B \sqsubseteq B$ $B \sqsubseteq D$ $B \sqsubseteq \exists S.A$ $B \sqsubseteq \exists S.C$

$B \sqsubseteq C$

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

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

$\xrightarrow{R_0} B \sqsubseteq B$ $B \sqsubseteq D$ $B \sqsubseteq \exists S.A$ $B \sqsubseteq \exists S.C$ $B \sqsubseteq C$

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

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

$\xrightarrow{R_0} B \sqsubseteq B \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq D$ $B \sqsubseteq \exists S.A$ $B \sqsubseteq \exists S.C$ $B \sqsubseteq C$

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

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

$\xrightarrow{R_0} B \sqsubseteq B \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq D \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq \exists S.A$ $B \sqsubseteq \exists S.C$ $B \sqsubseteq C$

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

$A \sqsubseteq A$ $A \sqsubseteq \exists R.B$ $A \sqsubseteq \exists R.D$

$A \sqsubseteq C$

$\xrightarrow{R_0} B \sqsubseteq B \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq D \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq \exists S.A \xrightarrow{R_{\exists}} B \sqsubseteq \exists S.C \quad B \sqsubseteq C$

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

$A \sqsubseteq A$ $A \sqsubseteq \exists R.B$ $A \sqsubseteq \exists R.D$

$A \sqsubseteq C$

$\xrightarrow{R_0} B \sqsubseteq B \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq D \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq \exists S.A \xrightarrow{R_{\exists}} B \sqsubseteq \exists S.C \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq C$

Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$

$A \sqsubseteq A$ $A \sqsubseteq \exists R.B$ $A \sqsubseteq \exists R.D$

$A \sqsubseteq C$

$\xrightarrow{R_0} B \sqsubseteq B \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq D \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq \exists S.A \xrightarrow{R_{\exists}} B \sqsubseteq \exists S.C \xrightarrow{R_{\sqsubseteq}} B \sqsubseteq C$

Then, **if needed**, we can recursively request inferences for $A \sqsubseteq C$

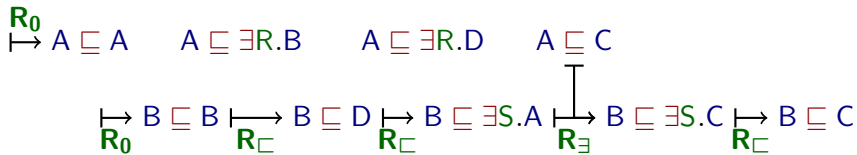
Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$



Then, **if needed**, we can recursively request inferences for $A \sqsubseteq C$

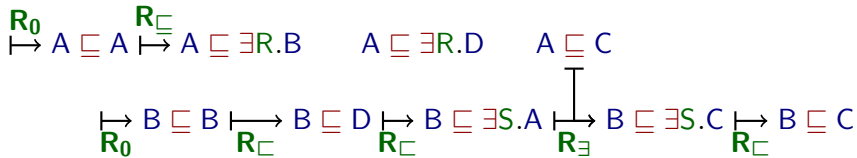
Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$



Then, **if needed**, we can recursively request inferences for $A \sqsubseteq C$

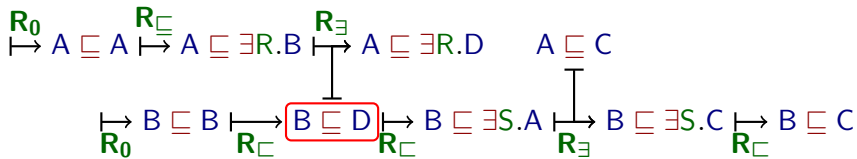
Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$



Then, **if needed**, we can recursively request inferences for $A \sqsubseteq C$

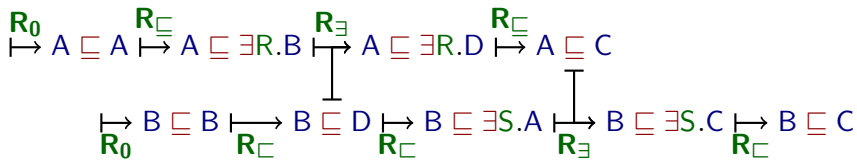
Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$



Then, **if needed**, we can recursively request inferences for $A \sqsubseteq C$

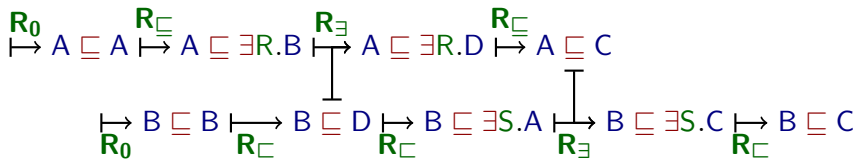
Goal-directed tracing: example

$A \sqsubseteq \exists R.B$, $B \sqsubseteq D$, $D \sqsubseteq \exists S.A$, $\exists R.D \sqsubseteq C$, $\exists S.C \sqsubseteq C$, $C \sqsubseteq \exists R.A$

Goal: reconstruct proofs of $B \sqsubseteq C$

Given: computed subsumptions (without tracing information)

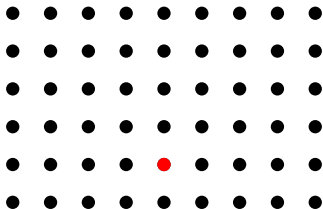
$C \sqsubseteq C$ $C \sqsubseteq \exists R.A$



Then, **if needed**, we can recursively request inferences for $A \sqsubseteq C$

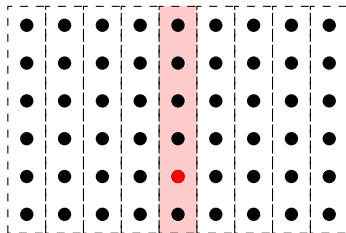
If we request inferences for $C \sqsubseteq \exists R.A$, we won't have to trace subsumptions with A or B on the left

Goal-directed tracing: the general procedure



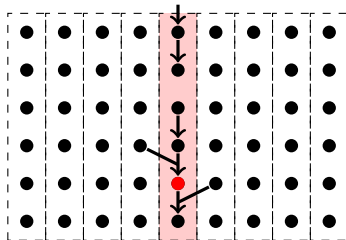
Goal-directed tracing: the general procedure

- ▶ **partition** the set of all derived expressions



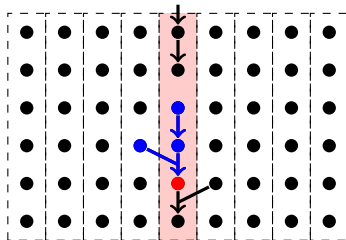
Goal-directed tracing: the general procedure

- ▶ **partition** the set of all derived expressions
- ▶ apply and trace all inferences **within** the target partition



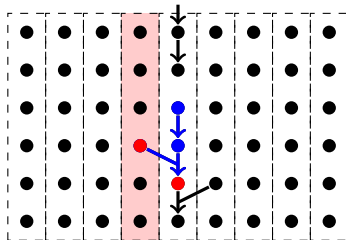
Goal-directed tracing: the general procedure

- ▶ **partition** the set of all derived expressions
- ▶ apply and trace all inferences **within** the target partition
- ▶ identify expressions used as **a part of a proof**



Goal-directed tracing: the general procedure

- ▶ **partition** the set of all derived expressions
- ▶ apply and trace all inferences **within** the target partition
- ▶ identify expressions used as **a part of a proof**
- ▶ **recursively** trace such expressions as new targets



\mathcal{EL} partitions for goal-directed tracing

We use the **left hand-side** as the partition ID for each subsumption

$$C \sqsubseteq D \mapsto C$$

\mathbf{R}_0 $\frac{}{C \sqsubseteq C}$	\mathbf{R}_{\sqcap}^- $\frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}$	\mathbf{R}_{\sqsubseteq} $\frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$
\mathbf{R}_{\top} $\frac{}{C \sqsubseteq \top}$	\mathbf{R}_{\sqcap}^+ $\frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2}$	\mathbf{R}_{\exists} $\frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$

\mathcal{EL} partitions for goal-directed tracing

We use the **left hand-side** as the partition ID for each subsumption

$$C \sqsubseteq D \mapsto C$$

The **key** property of this partitioning function:

- ▶ each inference's conclusion belongs to the same partition as one of the premises.

$\mathbf{R}_0 \frac{}{C \sqsubseteq C}$	$\mathbf{R}_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}$	$\mathbf{R}_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$
$\mathbf{R}_{\top} \frac{}{C \sqsubseteq \top}$	$\mathbf{R}_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2}$	$\mathbf{R}_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$

\mathcal{EL} partitions for goal-directed tracing

We use the **left hand-side** as the partition ID for each subsumption

$$C \sqsubseteq D \mapsto C$$

The **key** property of this partitioning function:

- ▶ each inference's conclusion belongs to the same partition as one of the premises.
- ▶ won't miss anything when applying rules **only** to premises in the target partition (using other expressions as **set of support**).

$$R_0 \frac{}{C \sqsubseteq C} \quad R_{\sqcap}^- \frac{C \sqsubseteq D_1 \sqcap D_2}{C \sqsubseteq D_1 \quad C \sqsubseteq D_2} \quad R_{\sqsubseteq} \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$R_{\top} \frac{}{C \sqsubseteq \top} \quad R_{\sqcap}^+ \frac{C \sqsubseteq D_1 \quad C \sqsubseteq D_2}{C \sqsubseteq D_1 \sqcap D_2} \quad R_{\exists} \frac{E \sqsubseteq \exists R.C \quad C \sqsubseteq D}{E \sqsubseteq \exists R.D}$$

Evaluation

Common \mathcal{EL}^+ ontologies: SNOMED CT, EL-GALEN, GO-EXT

Evaluation

Common \mathcal{EL}^+ ontologies: SNOMED CT, EL-GALEN, GO-EXT

Some classification statistics:

- ▶ **many** more inferences applied than derived subsumptions:
54.5M vs 19.9M for SNOMED, 3.9M vs 1.9M for GALEN

Evaluation

Common \mathcal{EL}^+ ontologies: SNOMED CT, EL-GALEN, GO-EXT

Some classification statistics:

- ▶ **many** more inferences applied than derived subsumptions:
54.5M vs 19.9M for SNOMED, 3.9M vs 1.9M for GALEN
- ▶ \mathcal{EL} partitioning is **fine-grained**:
avg partition is 73 subsumptions for GALEN, 54 for SNOMED
max partition is only 1,350 and 484, respectively

Evaluation

Common \mathcal{EL}^+ ontologies: SNOMED CT, EL-GALEN, GO-EXT

Some classification statistics:

- ▶ **many** more inferences applied than derived subsumptions:
54.5M vs 19.9M for SNOMED, 3.9M vs 1.9M for GALEN
- ▶ \mathcal{EL} partitioning is **fine-grained**:
avg partition is 73 subsumptions for GALEN, 54 for SNOMED
max partition is only 1,350 and 484, respectively

Hypothesis for goal-directed tracing evaluation:

- ▶ **few** partitions are traced
- ▶ **few** inferences are re-applied

Goal-directed tracing: experimental setup

- ▶ classify the ontology **without** tracing

Goal-directed tracing: experimental setup

- ▶ classify the ontology **without** tracing
- ▶ separately trace each subsumption between named concepts (and average over all of them)

Goal-directed tracing: experimental setup

- ▶ classify the ontology **without** tracing
- ▶ separately trace each subsumption between named concepts (and average over all of them)
- ▶ with full recursive proof unwinding

Goal-directed tracing: experimental setup

- ▶ classify the ontology **without** tracing
- ▶ separately trace each subsumption between named concepts (and average over all of them)
- ▶ with full recursive proof unwinding
- ▶ separately measure **traced** and **used** inferences
 - ▶ **used** means “used as a part of a proof”

Goal-directed tracing: results

GO: 5.7M inferences

GALEN: 3.9M

SNOMED CT: 54.5M

Ontology	Total partitions	# of traced partitions	# of traced inferences	# of inferences used in proofs	# of used \sqsubseteq axioms	Time (in ms.)
GO	46,8K	3.7	456.2	94.9	18.6	2.0
GALEN	25.9K	1.9	414.4	21.7	5.2	1.9
SNOMED CT	371.6K	8.6	788.3	385.9	9.7	10.1

Goal-directed tracing is fast.

Goal-directed tracing: results

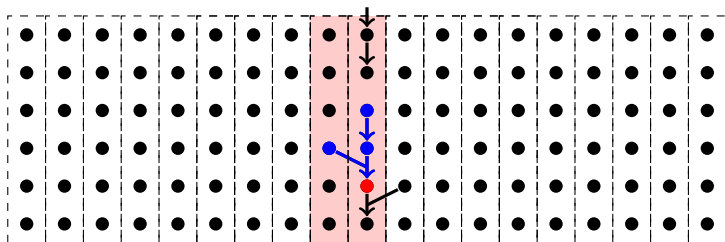
GO: 5.7M inferences

GALEN: 3.9M

SNOMED CT: 54.5M

Ontology	Total partitions	# of traced partitions	# of traced inferences	# of inferences used in proofs	# of used \sqsubseteq axioms	Time (in ms.)
GO	46.8K	3.7	456.2	94.9	18.6	2.0
GALEN	25.9K	1.9	414.4	21.7	5.2	1.9
SNOMED CT	371.6K	8.6	788.3	385.9	9.7	10.1

Only few partitions/inferences are traced.



Goal-directed tracing: results

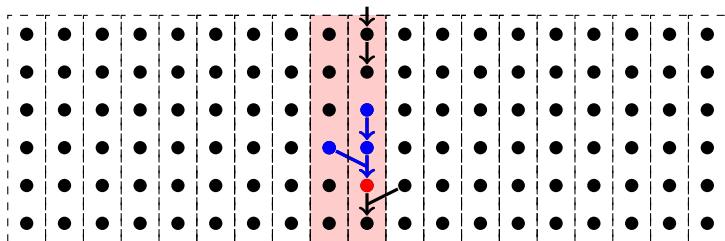
GO: 5.7M inferences

GALEN: 3.9M

SNOMED CT: 54.5M

Ontology	Total partitions	# of traced partitions	# of traced inferences	# of inferences used in proofs	# of used \sqsubseteq axioms	Time (in ms.)
GO	46,8K	3.7	456.2	94.9	18.6	2.0
GALEN	25.9K	1.9	414.4	21.7	5.2	1.9
SNOMED CT	371.6K	8.6	788.3	385.9	9.7	10.1

More inferences are stored than used in proofs. Doesn't matter.



Summary

Presented a new method for tracing which is

- ▶ agnostic to the rule system used (efficiency depends on **partitioning**)
- ▶ very simple and efficient for *EL*
- ▶ does **not** slow down the base procedure

Exploits the **granularity** property of the *EL* procedure, as for:

- ▶ concurrency
- ▶ incremental reasoning

Method can be used for:

- ▶ proof reconstruction, understanding, verification. . .
- ▶ interactive step-wise debugging
- ▶ speeding up justifications

Questions?

Tracing and justifications are different

Compute **different** things

Justifications:

- ▶ **minimal** subsets entailing the result
- ▶ **intractable**

Tracing:

- ▶ **compact** representation of all proofs
- ▶ not necessarily only **minimal** proofs!
- ▶ **tractable**

Tracing and justifications can be used together

Approach 1: use traces to reduce search space

- ▶ to the union over all axioms used in proofs
- ▶ i.e. like **modules**, just smaller

Approach 2: directly read justifications off proofs

- ▶ not every minimal proof corresponds to a justification
- ▶ ... but every justification corresponds to a proof!
- ▶ enumerate (via unwinding) all proofs and read justifications
- ▶ can be done **offline**, without the reasoner

Tracing and locality-based modules

Ontology modularity often used with justifications

- ▶ any justification is a subset of the module for the subsumption
- ▶ modules can be efficiently computed (and are **usually** small)
- ▶ justifications can be found inside the module

That's similar to searching for justifications inside the traces

Tracing and locality-based modules

Ontology modularity often used with justifications

- ▶ any justification is a subset of the module for the subsumption
- ▶ modules can be efficiently computed (and are **usually** small)
- ▶ justifications can be found inside the module

That's similar to searching for justifications inside the traces

And there's a relationship:

- ▶ modules **cannot** be smaller than traces
if $A \sqsubseteq B$ is derived then $M_B(\mathcal{O}) \subseteq M_A(\mathcal{O})$
- ▶ but they can be **significantly** larger

Tracing and locality-based modules

The GALEN case: the entire anatomical part is **connected**

$$A_0 \equiv B_0 \sqcap \exists R.A_1, \dots, A_i \equiv B_i \sqcap \exists R.A_{i+1}$$

The module for A_0 will contain modules for B_i and A_i

More than **10K** axioms on average for EL-GALEN

The trace for A_0 will **not** if partitions for most of A_i and B_i do not have inferences used in any proof