TECHNISCHE UNIVERSITÄT DRESDEN

CPEC
CENTER FOR PERSPICUOUS COMPUTING

Christian Alrabbaa   Stefan Borgwardt   Patrick Koopmann   Alisa Kovtunova

Finding Proofs for Description Logic Entailments in Practice

Based on "Finding Small Proofs for Description Logic Entailments—Theory and Practice" (LPAR'20) // Explainable Logic-Based Knowledge Representation (XLoKR 2020), September 14, 2020

Description Logics and Ontologies

Syntax of DL $\mathcal{ALC}$

Concepts: $C \ ::= \ A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \exists r.C \mid \forall r.C$

Axioms: $\alpha \ ::= \ C \sqsubseteq C \mid C \equiv C$

Description Logics

- Well-established formalism for specifying terminological knowledge in Ontologies
- Used for many large-scale ontologies
  - SNOMED CT: over 300,000 concepts
  - BioPortal: repository of bio-medical ontologies, currently hosting 889 ontologies defining 12,084,317 terms
  - MOWLCorp: ontologies obtained by web-crawling, containing 21,000 ontologies

Description Logics and Ontologies

Syntax of DL $\mathcal{ALC}$

Concepts: $C ::= A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \exists r.C \mid \forall r.C$
Axioms: $\alpha ::= C \sqsubseteq C \mid C \equiv C$

Description Logics

- Well-established formalism for specifying terminological knowledge in Ontologies
- Used for many large-scale ontologies
- With increasing complexity of the ontology, understanding entailments becomes both crucial and difficult

Description Logics and Ontologies

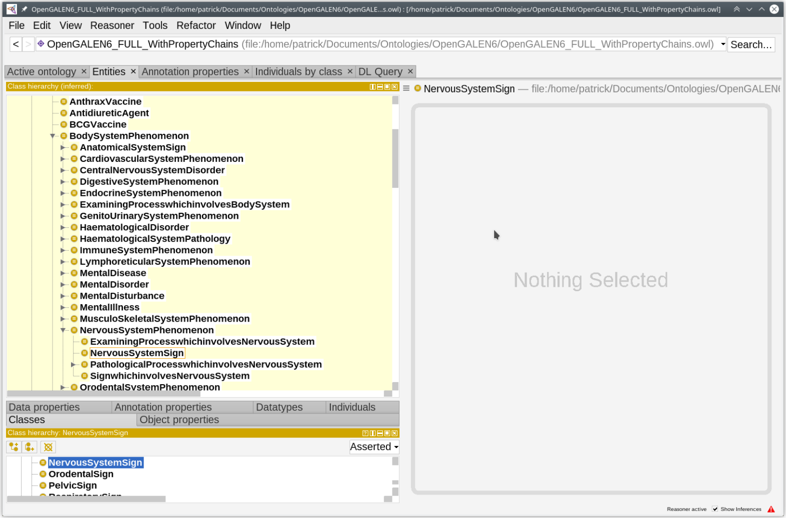
Syntax of DL $\mathcal{ALC}$

Concepts: $C ::= A \mid \neg C \mid C \sqcap C \mid C \sqcup C \mid \exists r.C \mid \forall r.C$
Axioms: $\alpha ::= C \sqsubseteq C \mid C \equiv C$

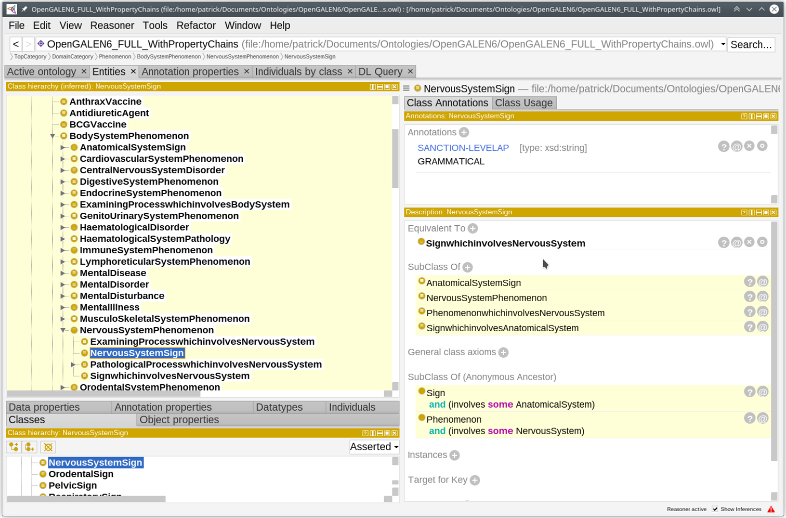
Description Logics

- Well-established formalism for specifying terminological knowledge in Ontologies
- Used for many large-scale ontologies
- With increasing complexity of the ontology, understanding entailments becomes both crucial and difficult
- One typical reasoning task is classification
  - compute all entailed axioms of form $A \sqsubseteq B$
  - obtain concept hierarchy

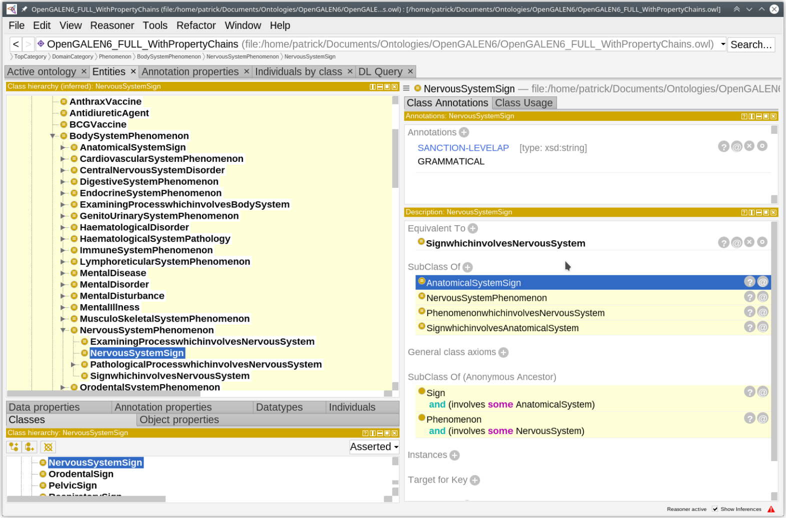
Current Tool of Choice: Justifications



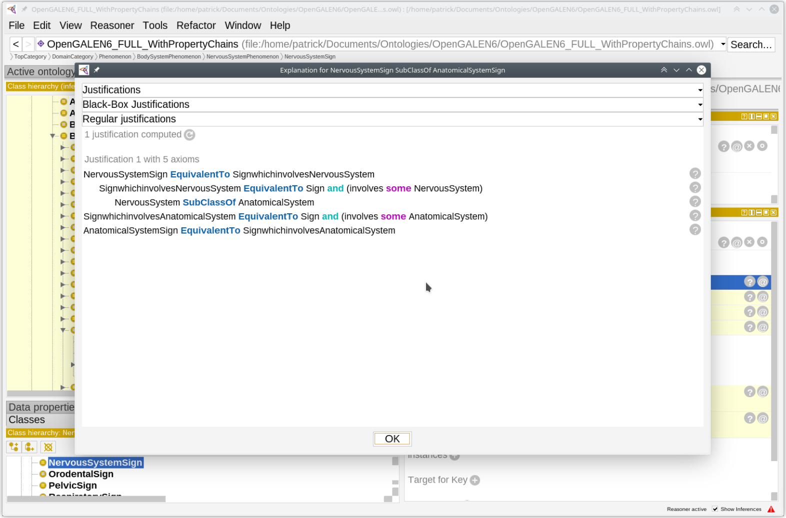
Current Tool of Choice: Justifications



Current Tool of Choice: Justifications



Current Tool of Choice: Justifications



Justifications

Justifications: Minimal subsets entailing given subsumption

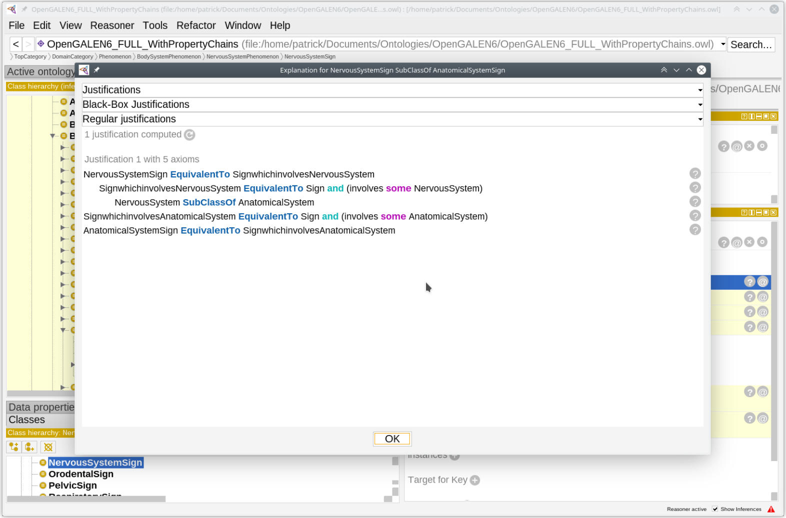
In practice often insufficient :

- can be large
- inferences often not obvious

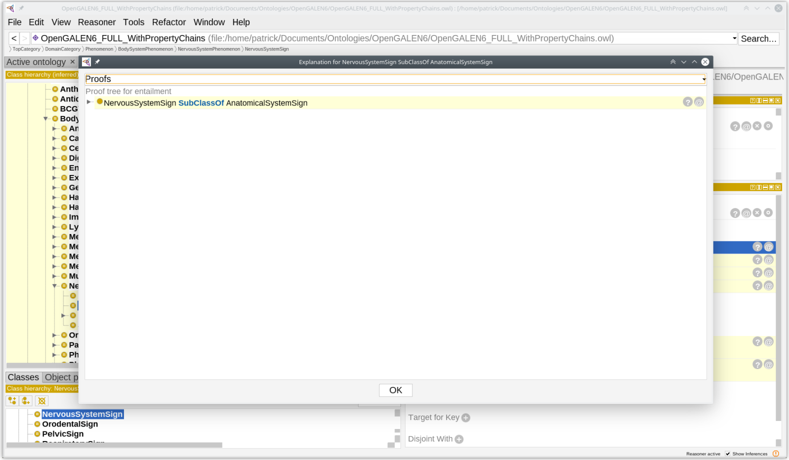
Showing *how* to obtain the inference would be better

- simple reasoning steps leading to conclusion
- generally known as proof

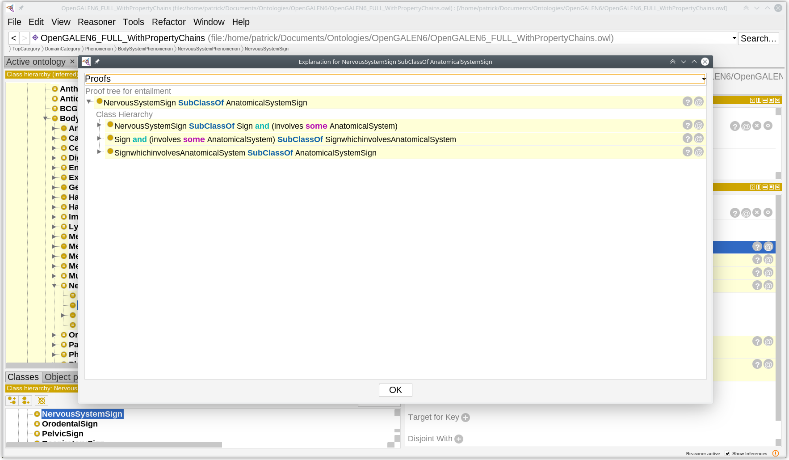
Proofs for ELK in Protege



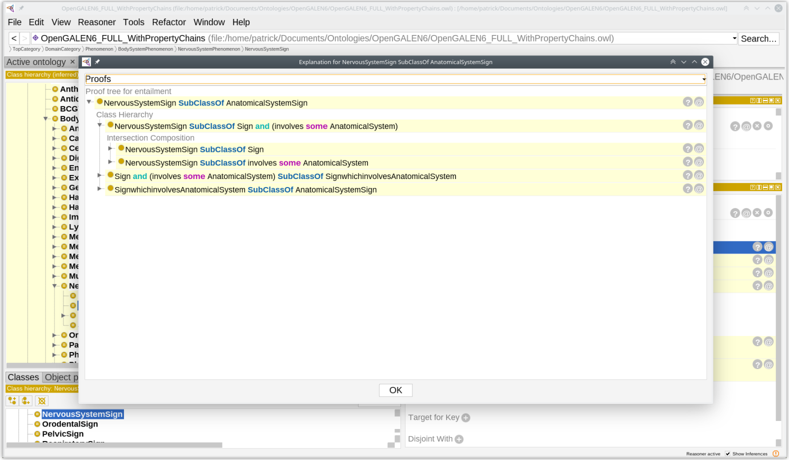
Proofs for ELK in Protege



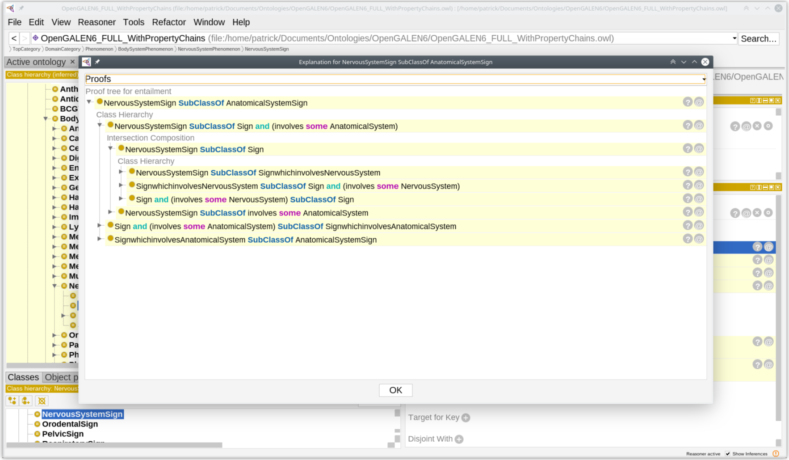
Proofs for ELK in Protege



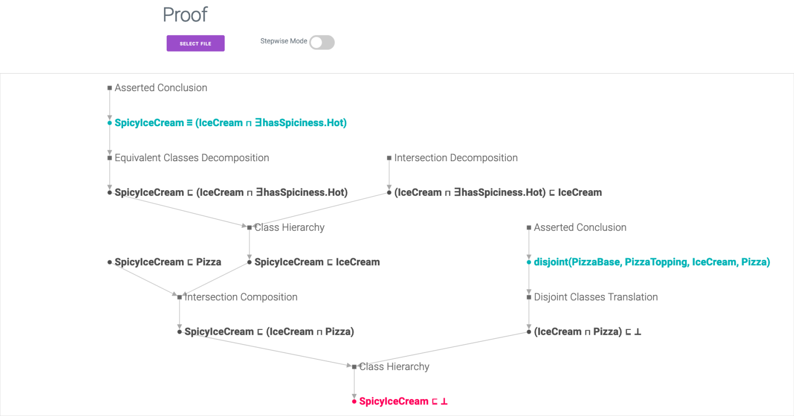
Proofs for ELK in Protege



Proofs for ELK in Protege



Proofs using Evonne (work in progress)



C. Alrabbaa, F. Baader, R. Dachselt, T. Flemisch, P. Koopmann: *Visualizing Proofs and the Modular Structure for Ontology Repair*, DL 2020.

Proofs with ELK

- ELK using a *consequence-based* reasoning procedure

⇒ inferences performed using a calculus

$$\mathbf{R}_0 \frac{}{C \sqsubseteq C} \qquad \mathbf{R}_\top \frac{}{C \sqsubseteq \top} \qquad \mathbf{R}_\sqcap^- \frac{C \sqsubseteq D \sqcap E}{C \sqsubseteq D \quad C \sqsubseteq E}$$

$$\mathbf{R}_\sqcap^+ \frac{C \sqsubseteq D \quad C \sqsubseteq E}{C \sqsubseteq D \sqcap E} \qquad \mathbf{R}_\exists \frac{C \sqsubseteq \exists r.D \quad D \sqsubseteq E}{C \sqsubseteq \exists r.E}$$

$$\mathbf{R}_\bot \frac{C \sqsubseteq \exists r.D \quad D \sqsubseteq \bot}{C \sqsubseteq \bot} \qquad \mathbf{R}_\sqsubseteq \frac{C \sqsubseteq D}{C \sqsubseteq E} : D \sqsubseteq E \in \mathcal{O}$$

$$\mathbf{R}_\circ \frac{C_0 \sqsubseteq \exists r_1.C_1, C_1 \sqsubseteq \exists r_2.C_2 \ \ldots \ C_{n-1} \sqsubseteq \exists r_n.C_n}{C_0 \sqsubseteq \exists r.C_n} : r_1 \circ \ldots \circ r_n \sqsubseteq r \in \mathcal{O}$$

⇒ proofs generated as part of reasoning process

Proofs For More Expressive DLs

- Currently, ELK is the only DL reasoner supporting proof generation
- ELK supports only a limited fragment of OWL, OWL EL
- More expressive reasoner often use other reasoning procedures
  - for a long time prominent: tableau reasoning
  - less convenient for understanding entailments
- Existing consequence-based reasoning for expressive DLs
  - often involved in complex systems
  - often combined with other reasoning paradigms
  - may use normal forms requiring different syntax

  $\Rightarrow$ generation of proofs not obvious
- Can we generate proofs without a calculus?

Justification Based Proofs

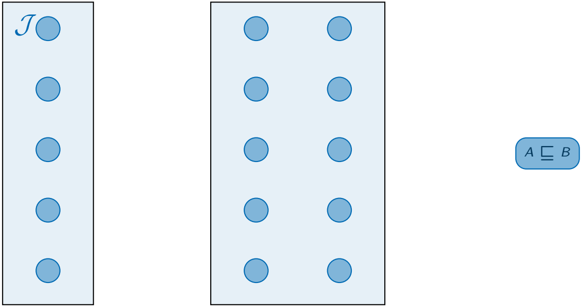
Justification-Based Proofs (Matthew Horridge 2011)

- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof

Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

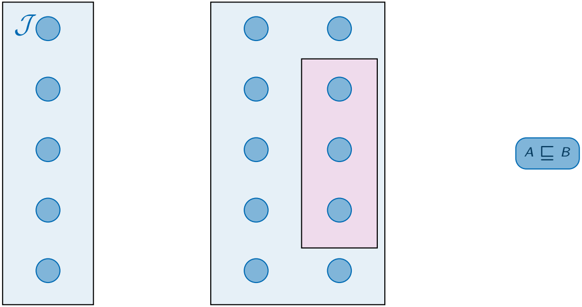
- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

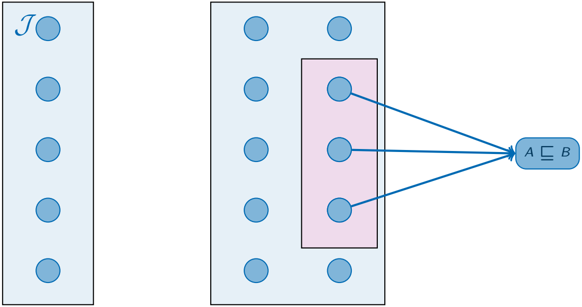
- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

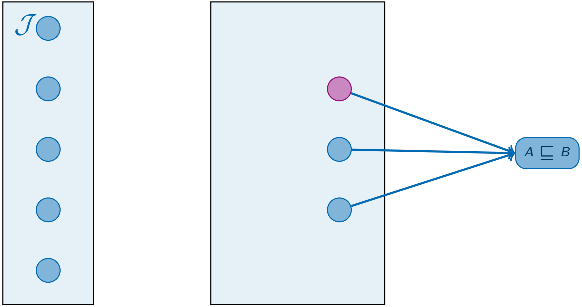
- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

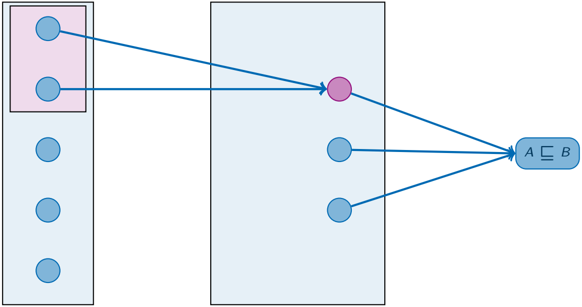
- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

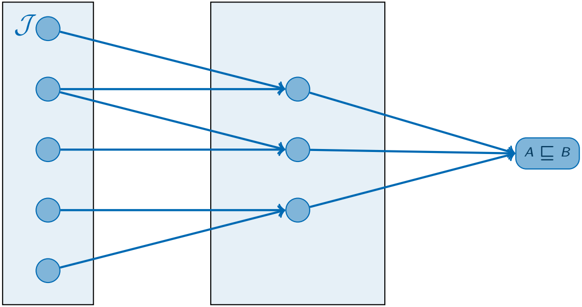
- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

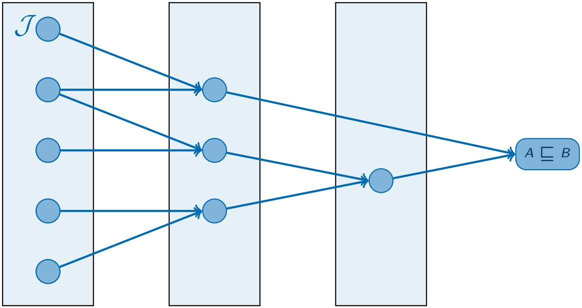
- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof



Justification Based Proofs

Justification-Based Proofs (Matthew Horridge 2011)

- derive intermediate steps between conclusion and justification
- consider all axioms of some predefined shapes
- justification-relation between allows to construct a proof
- involved ranking function allows to select axioms to be used in proof

Advantage of approach:

- generates best proof according to ranking

Disadvantage of approach:

- no clear inference principle
- hard to implement
- strongly depends on ranking function

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$b \qquad a \vee b \qquad \neg b \vee c \qquad \neg b \vee \neg c \qquad \neg a \vee c$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$b \qquad a \vee b \qquad \neg b \vee c \qquad \neg b \vee \neg c \qquad \neg a \vee c$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$b \qquad \neg b \vee c \qquad \neg b \vee \neg c \qquad b \vee c$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$b \qquad \neg b \vee c \qquad \neg b \vee \neg c \qquad b \vee c$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$c \qquad \neg c$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$C \qquad \neg C$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

⇒ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq C \qquad C \sqsubseteq \exists r.D \qquad \exists r.\top \sqsubseteq B$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq C \qquad C \sqsubseteq \exists r.D \qquad \exists r.\top \sqsubseteq B$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq C \qquad C \sqsubseteq \exists r.\top \qquad \exists r.\top \sqsubseteq B$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

⇒ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq C \qquad C \sqsubseteq \exists r.\top \qquad \exists r.\top \sqsubseteq B$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq C \qquad C \sqsubseteq B$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq C \qquad C \sqsubseteq B$$

Forgetting-Based Proofs

Forgetting Based Proofs

- Idea from propositional resolution:

$$\frac{q_1 \vee p \qquad q_2 \vee \neg p}{p_1 \vee p_2}$$

$\Rightarrow$ inference through elimination of p

- Decide satisfiability by eliminating names one after the other:

$$\bot$$

- Idea: Use similar approach to prove axioms of form $A \sqsubseteq B$

$$A \sqsubseteq B$$

Forgetting

Definition

Let $\mathcal{O}$ be an ontology and X a predicate name. Then, $\mathcal{O}^{-X}$ is a *result of forgetting* X iff

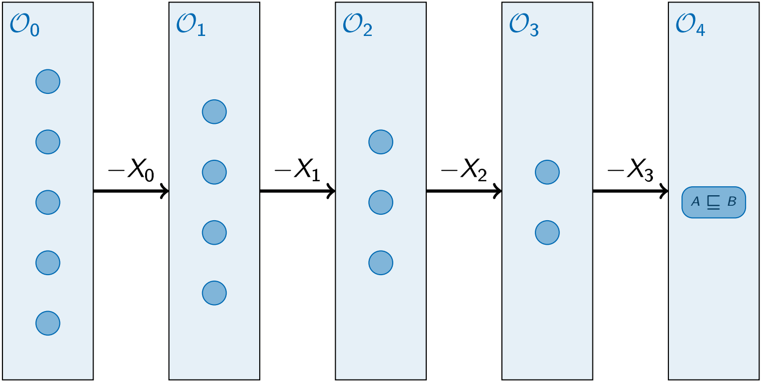
- X does not occur in $\mathcal{O}^{-X}$
- for every axiom α in which X does not occur, $\mathcal{O} \models \alpha$ iff $\mathcal{O}^{-X} \models \alpha$

$\Rightarrow$ strongest ontology without X entailed by $\mathcal{O}$

- which α to preserve also depends on underlying DL

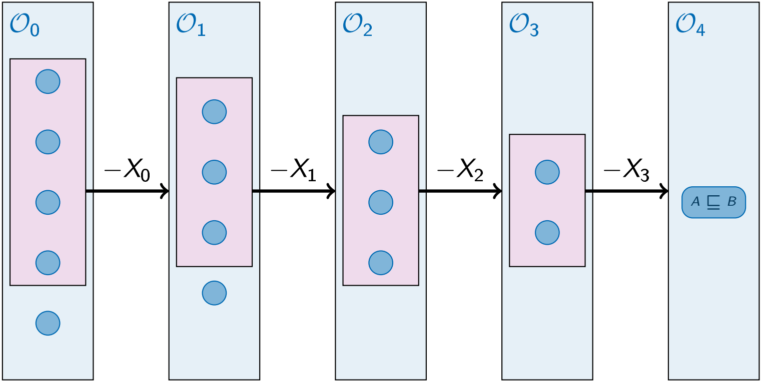
Forgetting based proofs

- Use forgetting to produce sequence of ontologies



Forgetting based proofs

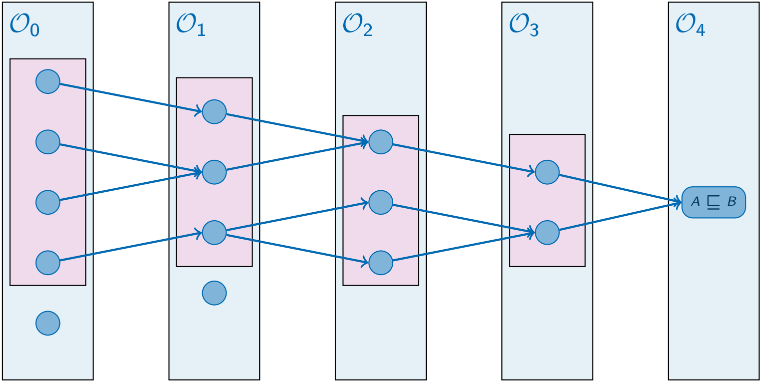
- Use forgetting to produce sequence of ontologies



- In each step, recompute justification for $A \sqsubseteq B$

Forgetting based proofs

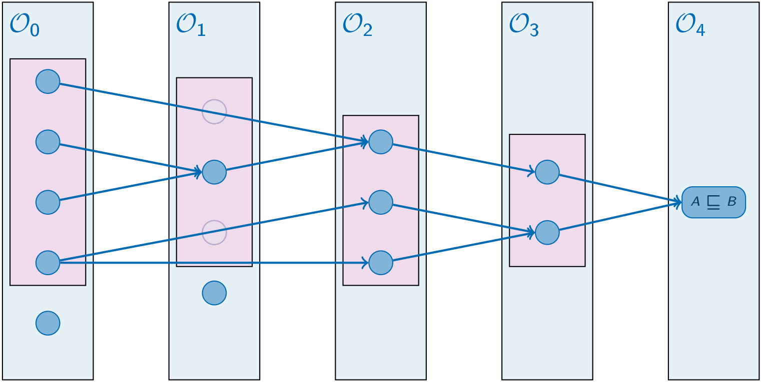
- Use forgetting to produce sequence of ontologies



- In each step, recompute justification for $A \sqsubseteq B$
- Finally, reconstruct proof using justifications

Forgetting based proofs

- Use forgetting to produce sequence of ontologies



- In each step, recompute justification for $A \sqsubseteq B$
- Finally, reconstruct proof using justifications, skipping steps if it makes sense

Forgetting Based Proof

$$
(C)\,\dfrac{A \sqsubseteq C \qquad (D)\,\dfrac{C \sqsubseteq \exists r.D}{C \sqsubseteq \exists r.\top}}{(r)\,\dfrac{A \sqsubseteq \exists r.\top \qquad\qquad \exists r.\top \sqsubseteq B}{A \sqsubseteq B}}
$$

Forgetting Order

Forgetting order, as well as selection of justification, affects proof

- Forgetting D first:

$$(r)\ \dfrac{(C)\ \dfrac{A \sqsubseteq C \qquad (D)\ \dfrac{C \sqsubseteq \exists r.D}{C \sqsubseteq \exists r.\top}}{A \sqsubseteq \exists r.\top} \qquad \exists r.\top \sqsubseteq B}{A \sqsubseteq B}$$

- Forgetting r first:

$$(C)\ \dfrac{A \sqsubseteq C \qquad (r)\ \dfrac{C \sqsubseteq \exists r.D \qquad \exists r.\top \sqsubseteq B}{C \sqsubseteq B}}{A \sqsubseteq B}$$

Forgetting Order

To obtain nicer proofs *practically*, we process names using the following heuristics:

- role with non-trivial fillers last:
  - otherwise may hide most of inference:

$$(r)\ \frac{A \sqsubseteq \exists r.B \qquad A \sqsubseteq \forall r.C \sqcup D \qquad B \sqsubseteq \exists s.D \qquad C \sqsubseteq \forall s.\neg D}{A \sqsubseteq D}$$

- unnested names first
  - delay complex inferences
- less frequent names first
  - delay expensive forgetting operations
  - (also used by existing forgetting procedures)

Evaluation

- Implemented approach in modular fashion
  - easy exchange of different forgetting procedures, provided they produce OWL ontologies
  - easy comparison with proofs generated by ELK
    - Dijkstra-based search to extract shortest proof
  - use 2 forgetting tools in the evaluation
    - $\mathcal{ALCH}$ variant of LETHE 0.6
    - $\mathcal{ALCOI}$ variant of FAME 1.0[1]

[1] there is a much improved version FAME 2.0, but it often creates ontologies outside of the OWL-standard

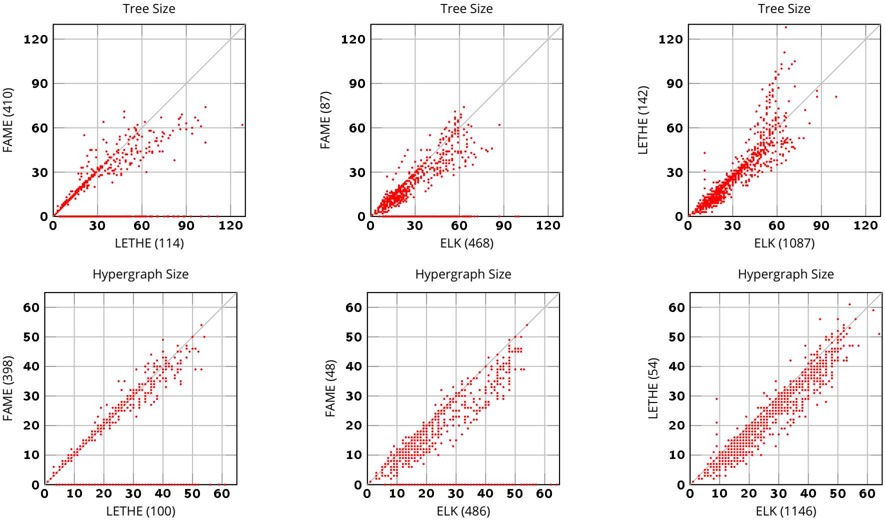
Evaluation: Corpus

- Focus on proofs in $\mathcal{ELH}$
  - to be able to compare with ELK
  - easier extraction of *justification patterns* (see below)
- Use ontologies from the OWL Reasoner Evaluation 2015, OWL EL Classification Track
  - well-balanced mix of ontologies from different repositories
- Extracted 1,573 *justification patterns*
  - all entailments of form $A \sqsubseteq B$ or $A \equiv B$
  - all justifications for these entailments
  - abstract away concept and role names
  - remove resulting duplicates

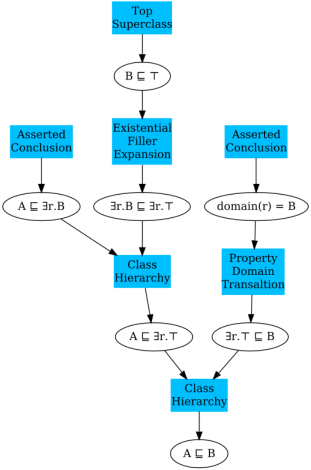
Evaluation: Metrics

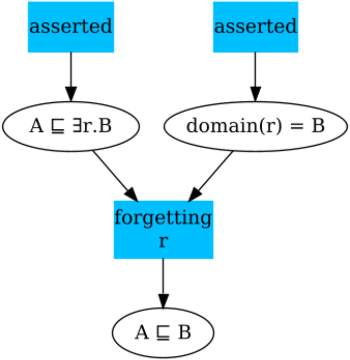
- Hypergraph-size
  - number of distinct axioms used in the proof
- Tree-size
  - sub-proofs count as often as they are used
- Justification Complexity
  - Matthew et al. 2013: "Toward cognitive support for OWL justifications"
  - attempt to measure cognitive complexity of justification
  - provides value for each proof step
  - we measured maximum and sum for each proof

Evaluation: Proof size

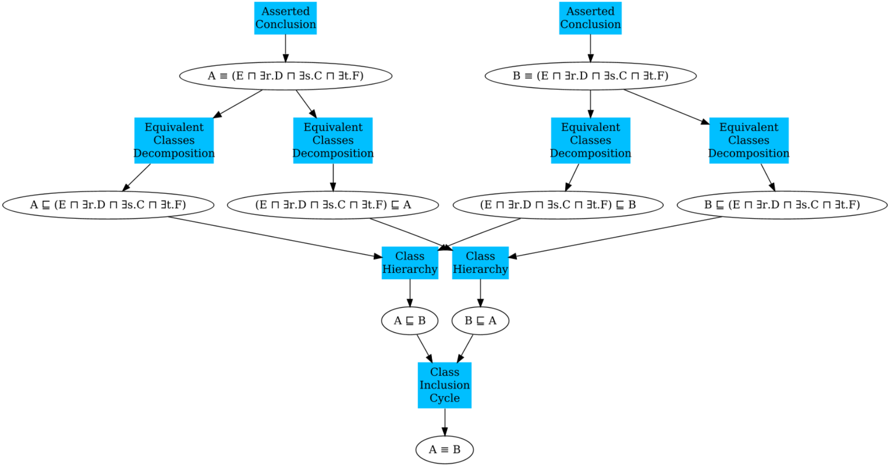


Evaluation: ELK vs. Forgetting-Based Proofs

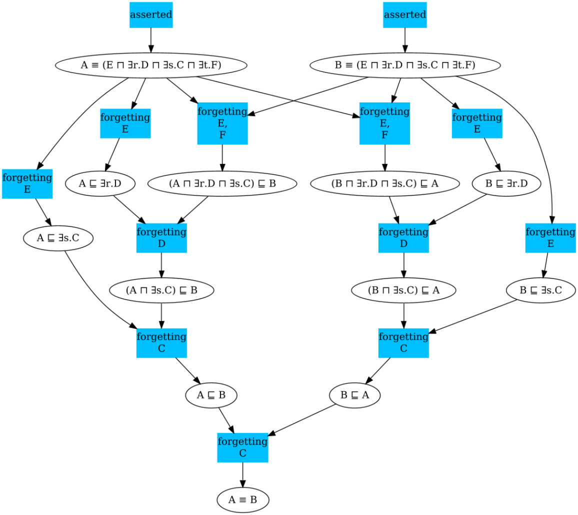




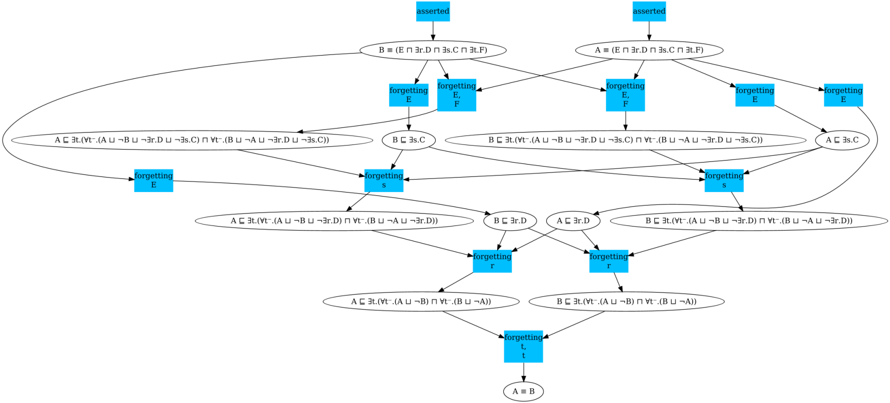
Evaluation: ELK Proof



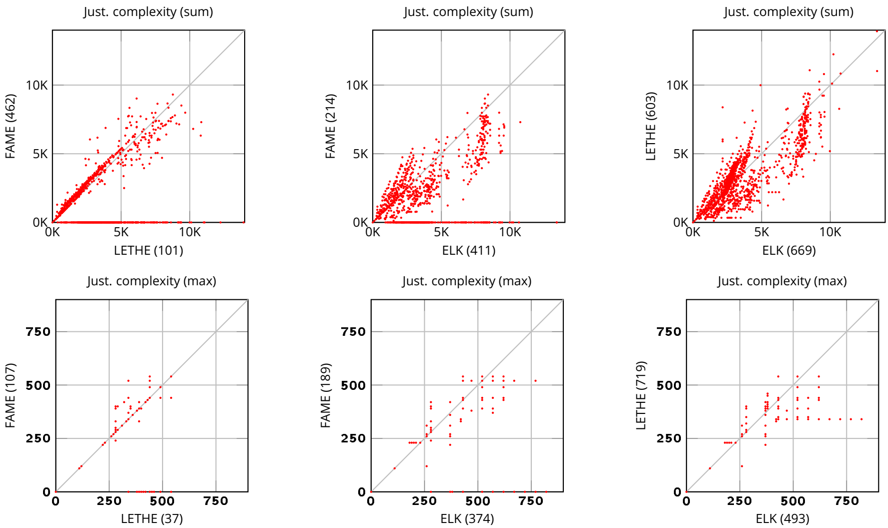
Evaluation: LETHE Proof



Evaluation: FAME Proof



Evaluation: Justification Complexity



Evaluation: Complexity of Inferences

- Both LETHE and FAME may use logical operators outside of $\mathcal{EL}$
- Large number of distinct "inference rules" was used:
  - LETHE: 362 different rules
  - FAME: 281 different rules

Conclusion

- New proof generation procedure based on forgetting
- Generate proof by repeated use of forgetting and justification
- Proofs for expressive DLs without calculus
- Can sometimes even compete with ELK
- Several possibilities to improve:
  - better heuristics on forgetting order or when to skip steps
  - dynamic selection of forgetting order
  - use learned “rules” to shorten proof computation times
  - integrate newer version of FAME