

Stefan Borgwardt

Technische Universität Dresden

Explaining Description Logic Reasoning

RuleML+RR'24, București, România, 18th September, 2024

Team

Logical Reasoning

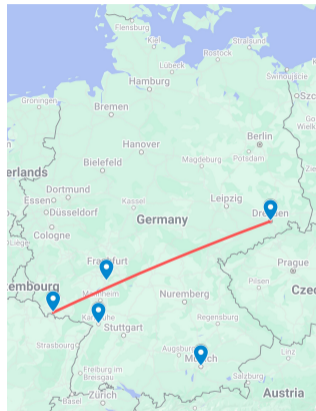
Human-Computer Interaction

Natural Language Generation



Undergraduate students:

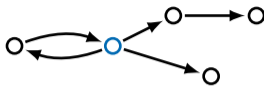
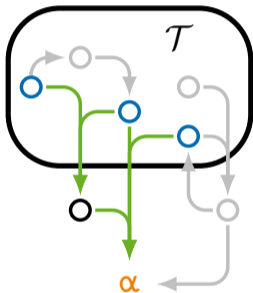
Tom Friese, Mikhail Kotlov, Alexej Popovič, Steffen Breuer, Anna Milena Rothermel, Duy Nhu, Kathryn Chapman, Hui-Syuan Yeh, Lukas Schmitz



Map data ©2024 GeoBasis-DE/BKG (©2009), Google, Inst. Geogr. Nacional

Outline

Proofs



Missing entailments

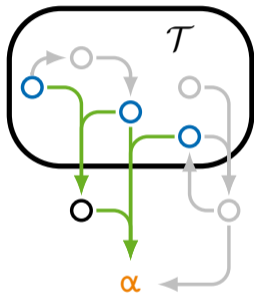


User studies



Verbalization

Justifications



TBox (ontology, knowledge base, logical theory) \mathcal{T} :
Collection of axioms (formulae, rules, facts)

Entailment (consequence, query) $\mathcal{T} \models \alpha$:
Axiom (formula, rule, fact) that follows logically from the TBox

Justification (MUS, minimal conflict set, why-provenance):
Minimal subset of \mathcal{T} from which α follows
(Horridge, Parsia, and Sattler 2010; Horridge, Bail, Parsia, and Sattler 2013)

Justifications are sometimes not enough for ...

complex reasoning

novice users

LLM-generated explanations

Explanation 135 Display laconic explanation

- Explanation for: 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' SubClassOf 'regulation of cellular ketone metabolic process by positive regulation of transcription from RNA polymerase II promoter'
- 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' SubClassOf 'regulation of fatty acid biosynthetic process by regulation of transcription from RNA polymerase II promoter' In 27 other justifications
 - 'regulation of fatty acid biosynthetic process by regulation of transcription from RNA polymerase II promoter' SubClassOf regulates some 'regulation of fatty acid biosynthetic process by transcription from RNA polymerase II promoter' In 9 other justifications
 - 'regulation of fatty acid biosynthetic process by transcription from RNA polymerase II promoter' SubClassOf 'regulation of fatty acid biosynthetic process' In 10 other justifications
 - 'regulation of fatty acid biosynthetic process' SubClassOf regulates some 'fatty acid biosynthetic process' In 16 other justifications
 - 'fatty acid biosynthetic process' SubClassOf 'fatty acid metabolic process' In 65 other justifications
 - Transitive: regulates In 18 other justifications
 - 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' SubClassOf biological_process and ('positively regulates' some 'transcription by RNA polymerase II') In NO other justifications
 - 'regulation of cellular ketone metabolic process' SubClassOf regulates some 'cellular ketone metabolic process' In 67 other justifications
 - 'regulation of fatty acid metabolic process' SubClassOf 'regulation of cellular ketone metabolic process' In 135 other justifications
 - biological_process and (regulates some 'fatty acid metabolic process') SubClassOf 'regulation of fatty acid metabolic process' In NO other justifications
 - biological_process and (regulates some 'cellular ketone metabolic process') and ('positively regulates' some 'transcription by RNA polymerase II') SubClassOf 'regulation of cellular ketone metabolic process by positive regulation of transcription from RNA polymerase II promoter' In NO other justifications

Justification in the ontology editor PROTÉGÉ

Justifications are sometimes not enough for ...

complex reasoning

novice users

LLM-generated explanations

Proof tree for entailment

● 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' **SubClassOf** 'regulation of cellular ketone metabolic process by positive regulation of transcription from RNA polymerase II promoter'

Class Hierarchy

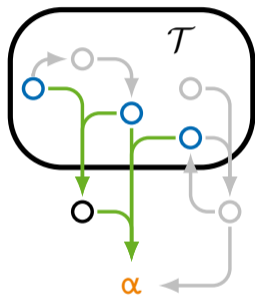
- 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' **SubClassOf** biological_process and (regulates some 'cellular ketone metabolic process') and ('positively regulates' some 'transcription by RNA polymerase II')

Intersection Composition

- 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' **SubClassOf** biological_process
- 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' **SubClassOf** regulates some 'cellular ketone metabolic process'
- 'positive regulation of unsaturated fatty acid biosynthetic process by positive regulation of transcription from RNA polymerase II promoter' **SubClassOf** 'positively regulates' some 'transcription by RNA polymerase II'
- biological_process and (regulates some 'cellular ketone metabolic process') and ('positively regulates' some 'transcription by RNA polymerase II') **SubClassOf** 'regulation of cellular ketone metabolic process by positive regulation of transcription from RNA polymerase II promoter'

Proof in the ontology editor PROTEGE

Proofs are hypergraphs



Inference step (hyperedge): List of premises + conclusion

Proof: Acyclic, non-redundant hypergraph with sink α

Hypergraph $\mathcal{D}(\mathcal{T}, \alpha)$ of **all admissible inference steps** for $\mathcal{T} \models \alpha$

Goal: Find optimal proofs in $\mathcal{D}(\mathcal{T}, \alpha)$

Measures: Size (5), tree size (6), depth (2), ...

Complexity of finding proofs

“Is there a proof in $\mathcal{D}(\mathcal{T}, \alpha)$ with measure $\leq p$?”

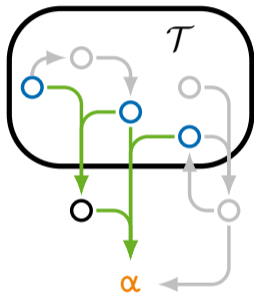
(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020, 2021)

	polynomial unary	polynomial binary	exponential unary	exponential binary
Size	NP	NP	NP	NExpTime
Monotone recursive Φ -measures	$\leq P$	$\leq P$	$\leq \text{ExpTime}$	$\leq \text{ExpTime}$
Tree size	P	P	NP	PSpace
Depth	P	P	PSpace	ExpTime
Logarithmic depth	P	P	ExpTime	ExpTime

guess proof in $\mathcal{D}(\mathcal{T}, \alpha)$

polynomial Dijkstra-like algorithm for directed hypergraphs

Computing proofs



Consequence-based reasoners (ELK, CLIPPER, SEQUOIA)

(Kazakov, Krötzsch, and Simancik 2014; Eiter, Ortiz, Simkus, Tran, and Xiao 2012; Tena Cucala, Cuenca Grau, and Horrocks 2021)

Justification-oriented proofs

(Horridge, Parsia, and Sattler 2010)

Concept interpolation: $A \sqsubseteq ? \sqsubseteq B$

(Schlobach 2004; Peuter 2024)

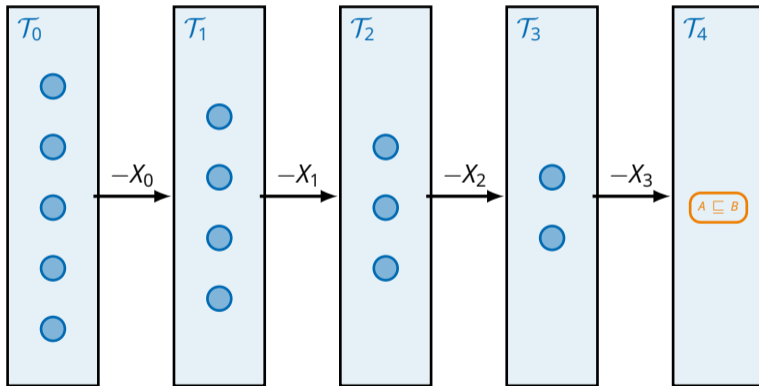
Elimination proofs based on forgetting (LETHE, FAME)

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020)

Computing proofs

Elimination proofs based on forgetting (LETHE, FAME)

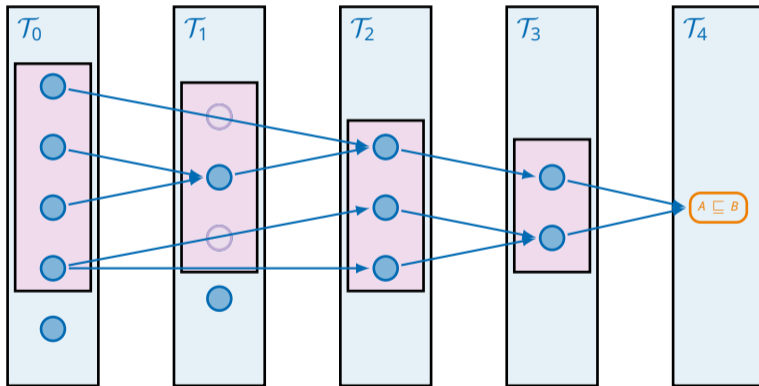
(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020)



Computing proofs

Elimination proofs based on forgetting (LETHE, FAME)

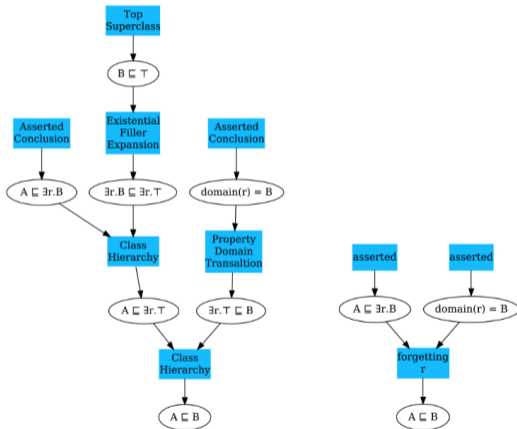
(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020)



Computing proofs

Elimination proofs based on forgetting (LETHE, FAME)

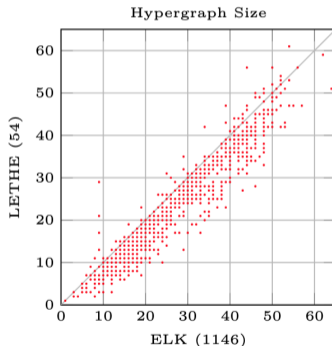
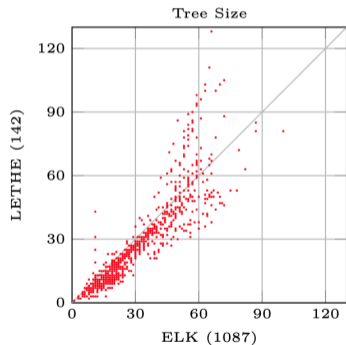
(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020)



Size of proofs

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020)

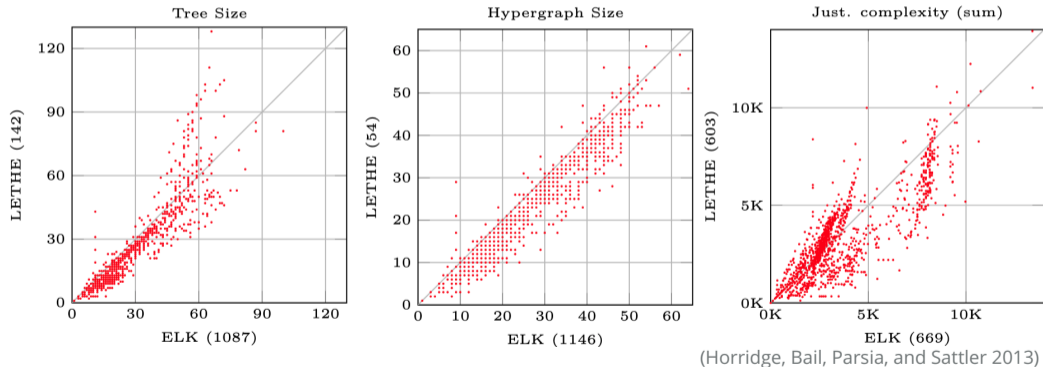
- Prototypical proofs from ORE 2015 ontologies (\mathcal{ELH})
- Proof of minimal size = proof of minimal tree size



Size of proofs

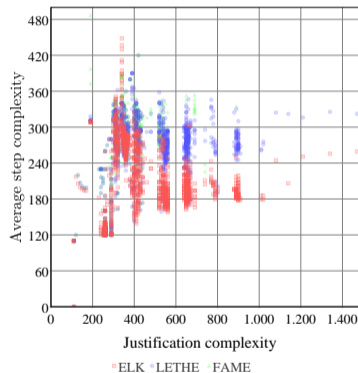
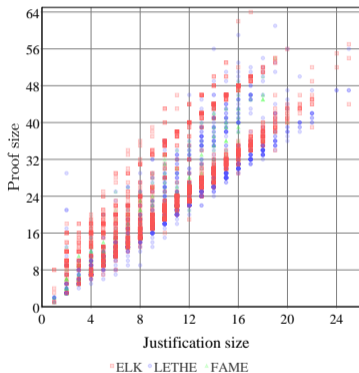
(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2020)

- Prototypical proofs from ORE 2015 ontologies (\mathcal{ELH})
- Proof of minimal size = proof of minimal tree size



Proofs vs. justifications

(B 2021)



Proofs for ontology-mediated queries

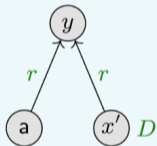
(Borgida, Calvanese, and Rodriguez-Muro 2008; Alrabbaa, B, Koopmann, and Kovtunova 2022)

ABox: $A(a)$

TBox: $A \sqsubseteq \exists r.B$, $B \sqsubseteq \exists s.A$, $\exists s \sqcap \exists r^- \sqsubseteq E$, $\exists r.E \sqsubseteq D$

Query:

$\exists x', y. r(x, y) \wedge r(x', y) \wedge D(x')$



Proofs for ontology-mediated queries

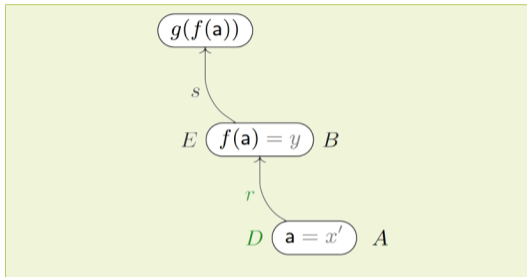
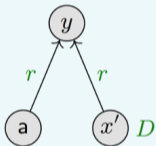
(Borgida, Calvanese, and Rodriguez-Muro 2008; Alrabbaa, B, Koopmann, and Kovtunova 2022)

ABox: $A(a)$

TBox: $A \sqsubseteq \exists r.B$, $B \sqsubseteq \exists s.A$, $\exists s \sqcap \exists r^- \sqsubseteq E$, $\exists r.E \sqsubseteq D$

Query:

$\exists x', y. r(x, y) \wedge r(x', y) \wedge D(x')$



Proofs for ontology-mediated queries

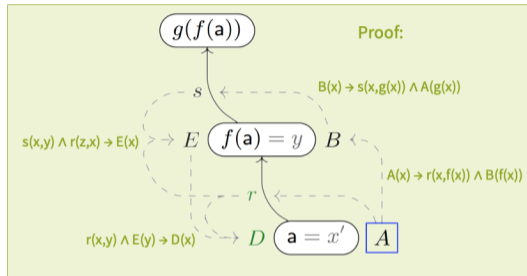
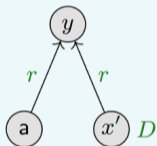
(Borgida, Calvanese, and Rodriguez-Muro 2008; Alrabbaa, B, Koopmann, and Kovtunova 2022)

ABox: $A(a)$

TBox: $A \sqsubseteq \exists r.B, B \sqsubseteq \exists s.A, \exists s \sqcap \exists r^- \sqsubseteq E, \exists r.E \sqsubseteq D$

Query:

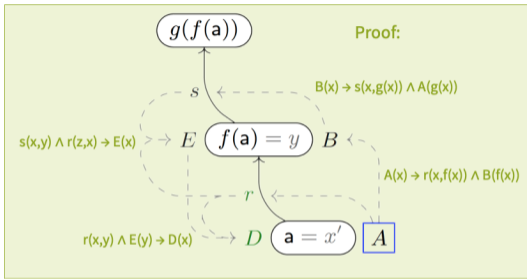
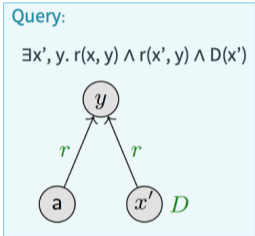
$\exists x', y. r(x, y) \wedge r(x', y) \wedge D(x')$



Proofs for ontology-mediated queries

(Borgida, Calvanese, and Rodriguez-Muro 2008; Alrabbaa, B, Koopmann, and Kovtunova 2022)

ABox: $A(a)$
 TBox: $A \sqsubseteq \exists r.B, B \sqsubseteq \exists s.A, \exists s \sqcap \exists r^- \sqsubseteq E, \exists r.E \sqsubseteq D$



	DL-Lite		\mathcal{EL}		Horn- <i>ALCHOI</i>
	ELQ	CQ	IQ	CQ	CQ
Domain size	NP-c		in ExpTime		in NExpTime
Tree size	in P	NP-c	P-c	NP-c	in PSpace
Proof size	NP-c		in ExpTime		
Proof size bound	polynomial		exponential	double exponential	

Proofs with concrete domains

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2023)

$$[\text{fuel_level} = 0.1 \cdot \text{fuel_capacity}] \sqsubseteq \text{LowFuel}$$

Proofs with concrete domains

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2023)

$$[\text{fuel_level} = 0.1 \cdot \text{fuel_capacity}] \sqsubseteq \text{LowFuel} \quad \rightsquigarrow \quad A' \sqsubseteq \text{LowFuel}$$

Proofs with concrete domains

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2023)

$[\text{fuel_level} = 0.1 \cdot \text{fuel_capacity}] \sqsubseteq \text{LowFuel} \quad \rightsquigarrow \quad A' \sqsubseteq \text{LowFuel}$

$$\frac{\frac{C \sqsubseteq A_\alpha \quad C \sqsubseteq A_\beta}{C \sqsubseteq A_\alpha \sqcap A_\beta} \quad A_\alpha \sqcap A_\beta \sqsubseteq A_\gamma}{C \sqsubseteq A_\gamma}$$

Proofs with concrete domains

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2023)

$[\text{fuel_level} = 0.1 \cdot \text{fuel_capacity}] \sqsubseteq \text{LowFuel} \quad \rightsquigarrow \quad A' \sqsubseteq \text{LowFuel}$

$$\frac{\frac{C \sqsubseteq A_\alpha \quad C \sqsubseteq A_\beta}{C \sqsubseteq A_\alpha \sqcap A_\beta} \quad A_\alpha \sqcap A_\beta \sqsubseteq A_\gamma}{C \sqsubseteq A_\gamma}$$

$$\frac{2x + 3y = 5 \quad \frac{4y = 3}{-12y = -9} \quad [-3]}{4x - 6y = 1} \quad [2, 1]$$

Proofs with concrete domains

(Alrabbaa, Baader, B, Koopmann, and Kovtunova 2023)

$[\text{fuel_level} = 0.1 \cdot \text{fuel_capacity}] \sqsubseteq \text{LowFuel} \rightsquigarrow A' \sqsubseteq \text{LowFuel}$

$$\frac{\frac{C \sqsubseteq A_\alpha \quad C \sqsubseteq A_\beta}{C \sqsubseteq A_\alpha \sqcap A_\beta} \quad A_\alpha \sqcap A_\beta \sqsubseteq A_\gamma}{C \sqsubseteq A_\gamma}$$

$$\frac{2x + 3y = 5 \quad \frac{4y = 3}{-12y = -9} \quad [-3]}{4x - 6y = 1} \quad [2, 1]$$

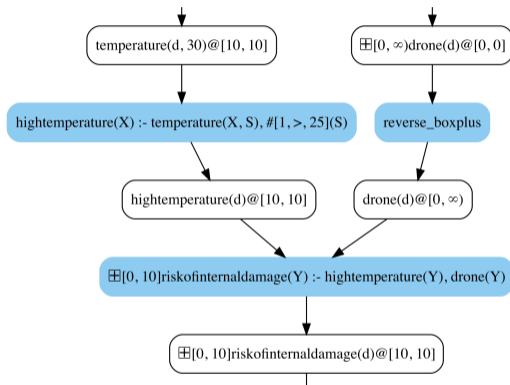
\Rightarrow

$$\frac{C \sqsubseteq [2x + 3y = 5] \quad \frac{C \sqsubseteq [4y = 3]}{C \sqsubseteq [-12y = -9]} \quad [-3]}{C \sqsubseteq [4x - 6y = 1]} \quad [2, 1]$$

Outlook: Proofs for rules

Proofs for DatalogMTL in METEOR

(B, Demberg, Jobanputra, Kovtunova, and Nhu 2024)

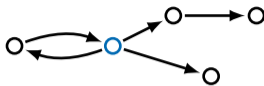
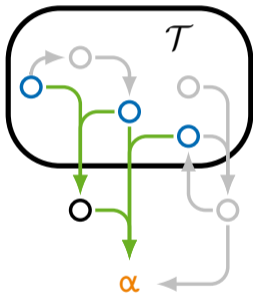


Proofs for Datalog and existential rules in NEMO

(Ivliev, Ellmauthaler, Gerlach, Marx, Meißner, Meusel, and Krötzsch 2023)

Outline

Proofs



Missing entailments



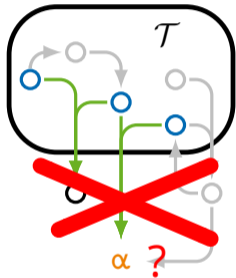
User studies



Verbalization

Explaining missing entailments

Missing entailment: $\mathcal{T} \not\models \alpha$



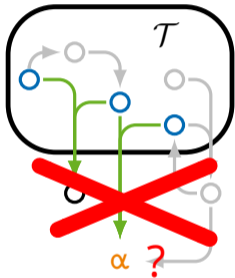
Counterexample:

Model of \mathcal{T} that does not satisfy α

(Bauer, Sattler, and Parsia 2009; Alrabbaa and Hieke 2022; Alrabbaa, B, Friese, Koopmann, and Kotlov 2023)

Explaining missing entailments

Missing entailment: $\mathcal{T} \not\models \alpha$



Counterexample:

Model of \mathcal{T} that does not satisfy α

(Bauer, Sattler, and Parsia 2009; Alrabbaa and Hieke 2022; Alrabbaa, B, Friese, Koopmann, and Kotlov 2023)

Abductive hypothesis:

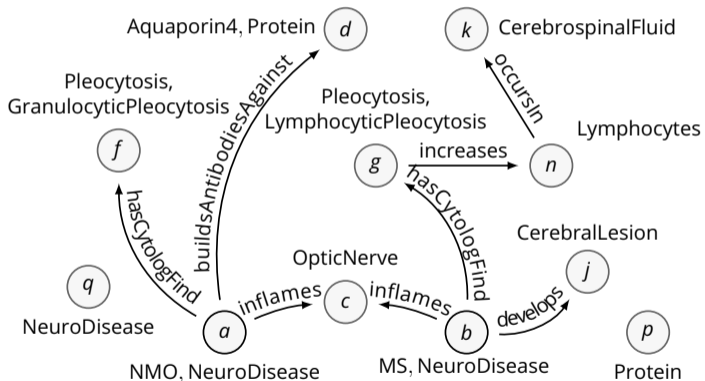
\mathcal{H} such that $\mathcal{T} \cup \mathcal{H} \models \alpha$ ($\mathcal{T} \cup \mathcal{H} \not\models \perp$ and $\mathcal{H} \neq \{\alpha\}$)

(Calvanese, Ortiz, Simkus, and Stefanoni 2013; Wei-Kleiner, Dragisic, and Lambrix 2014; Koopmann, Del-Pinto, Tourret, and Schmidt 2020; Haifani, Koopmann, Tourret, and Weidenbach 2022)

Relevant counterexamples

Model of \mathcal{T} that does not satisfy **NeuromyelitisOptica** \sqsubseteq **MultipleSclerosis**

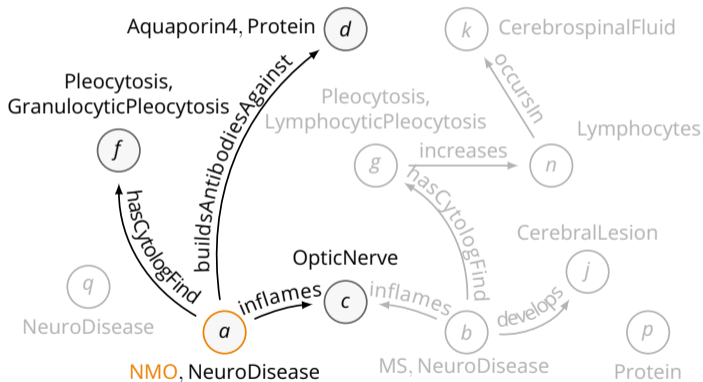
(Alrabbaa and Hieke 2022)



Relevant counterexamples

Part of a model of \mathcal{T} that does not satisfy $\text{NeuromyelitisOptica} \sqsubseteq \text{MultipleSclerosis}$

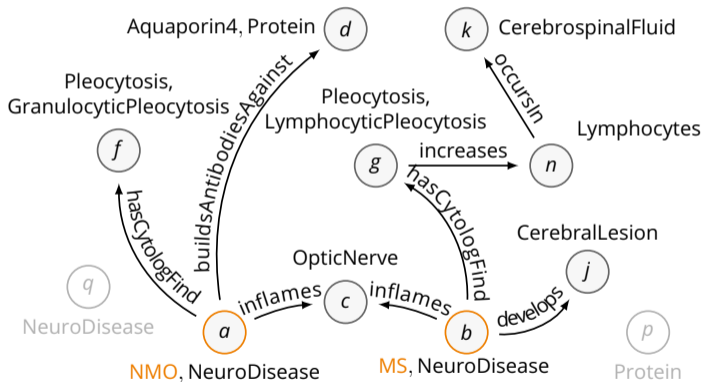
(Alrabbaa and Hieke 2022)



Relevant counterexamples

Part of a model of \mathcal{T} that does not satisfy $\text{NeuromyelitisOptica} \sqsubseteq \text{MultipleSclerosis}$

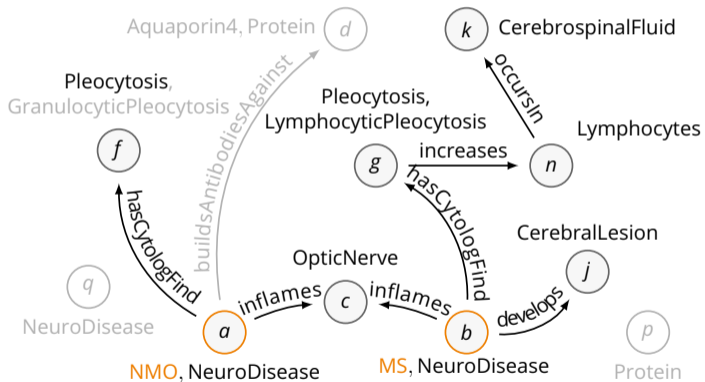
(Alrabbaa and Hieke 2022)



Relevant counterexamples

Part of a model of \mathcal{T} that does not satisfy $\text{NeuromyelitisOptica} \sqsubseteq \text{MultipleSclerosis}$

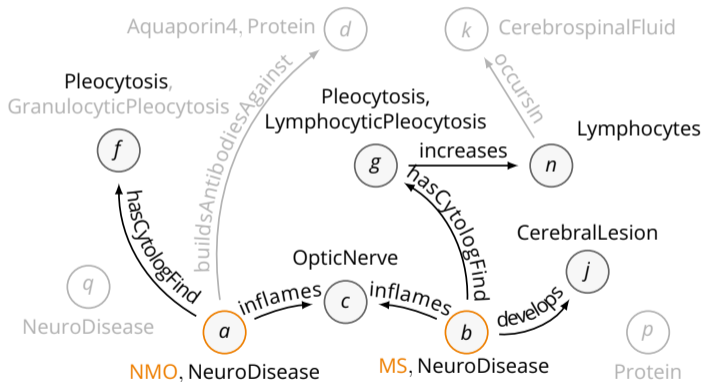
(Alrabbaa and Hieke 2022)



Relevant counterexamples

Part of a model of \mathcal{T} that does not satisfy $\text{NeuromyelitisOptica} \sqsubseteq \text{MultipleSclerosis}$

(Alrabbaa and Hieke 2022)



Abduction

\mathcal{H} such that $\mathcal{T} \cup \mathcal{H} \models \alpha$ ($\mathcal{T} \cup \mathcal{H} \not\models \perp$ and $\mathcal{H} \not\equiv \{\alpha\}$)

Complete signature-based abduction

(Koopmann, Del-Pinto, Tourret, and Schmidt 2020)

- Restrict \mathcal{H} to a signature Σ
- High expressivity: \mathcal{T} in \mathcal{ALC} , \mathcal{H} with disjunction of axioms, inverse roles, nominals, fixpoint concepts
- Implementation based on LETHE

Abduction

\mathcal{H} such that $\mathcal{T} \cup \mathcal{H} \models \alpha$ ($\mathcal{T} \cup \mathcal{H} \not\models \perp$ and $\mathcal{H} \not\equiv \{\alpha\}$)

Complete signature-based abduction

(Koopmann, Del-Pinto, Tourret, and Schmidt 2020)

- Restrict \mathcal{H} to a signature Σ
- High expressivity: \mathcal{T} in \mathcal{ALC} , \mathcal{H} with disjunction of axioms, inverse roles, nominals, fixpoint concepts
- Implementation based on LETHE

Connection-minimal abduction

(Haifani, Koopmann, Tourret, and Weidenbach 2022)

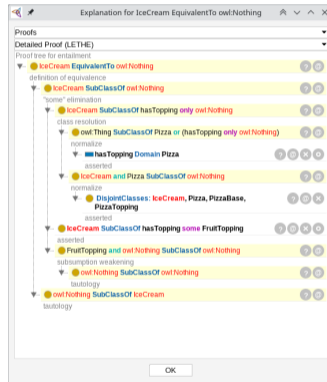
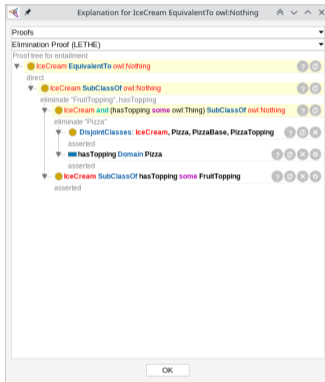
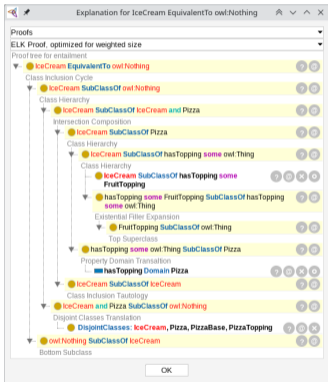
- \mathcal{H} is “connected” to α
- Low expressivity: \mathcal{T} in \mathcal{EL} , \mathcal{H} only conjunctions of concept names
- Implementation (CAPI) based on SPASS

(Non)commercial break

EVEE – a collection of PROTÉGÉ plugins



github.com/
de-tu-dresden-inf-lat/evee

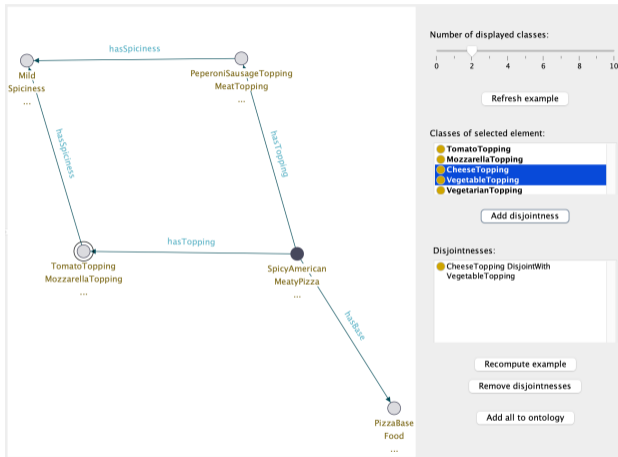


(Alrabbaa, B, Friese, Hirsch, Knieriemen, Koopmann, Kovtunova, Krüger, Popovič, and Siahaan 2024)

EVEE – a collection of PROTÉGÉ plugins



[github.com/
de-tu-dresden-inf-lat/evee](https://github.com/de-tu-dresden-inf-lat/evee)



(Alrabbaa, B, Friese, Hirsch, Knieriemen, Koopmann, Kovtunova, Krüger, Popovič, and Siahaan 2024)

EVEE – a collection of PROTÉGÉ plugins



[github.com/
de-tu-dresden-inf-lat/evee](https://github.com/de-tu-dresden-inf-lat/evee)

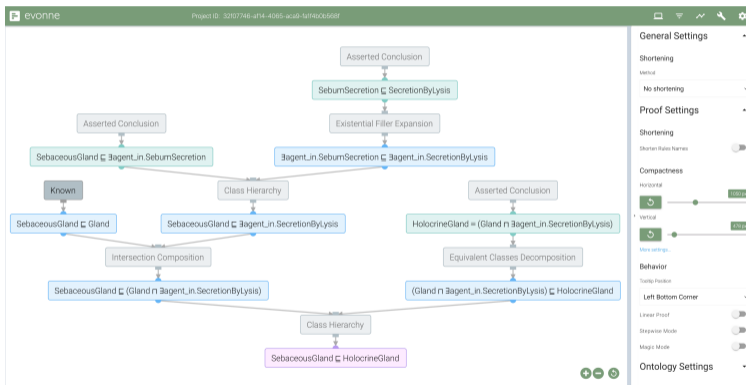
The screenshot displays the EVEE interface with two main panels. The left panel, titled 'Vocabulary', is divided into 'Permitted vocabulary' and 'Forbidden vocabulary'. Under 'Permitted vocabulary', there are tabs for 'Classes', 'Object properties', and 'Individuals'. A list of classes includes American, CheeseTopping, ChilliTopping, Country, DeepPanBase, DomainThing, Food, and HamTopping. Below this list are navigation arrows. Under 'Forbidden vocabulary', there are also tabs for 'Classes', 'Object properties', and 'Individuals', with a single entry 'SpicyPizza'. The right panel, titled 'Hypotheses:', contains two hypothesis boxes. The first box, 'Hypothesis 1', shows a hypothesis: 'Mild SubClassOf Hot'. The second box, 'Hypothesis 2', shows a hypothesis: 'SpicyAmerican SubClassOf (hasTopping some ChilliTopping) or (hasTopping some SpicySalamiTopping)'. Each hypothesis box has three buttons: 'Explain', 'Forbid vocabulary', and 'Add to ontology'.

(Alrabbaa, B, Friese, Hirsch, Knieriemen, Koopmann, Kovtunova, Krüger, Popovič, and Siahaan 2024)

EVONNE – EVEE's bigger sister



imld.de/evonne

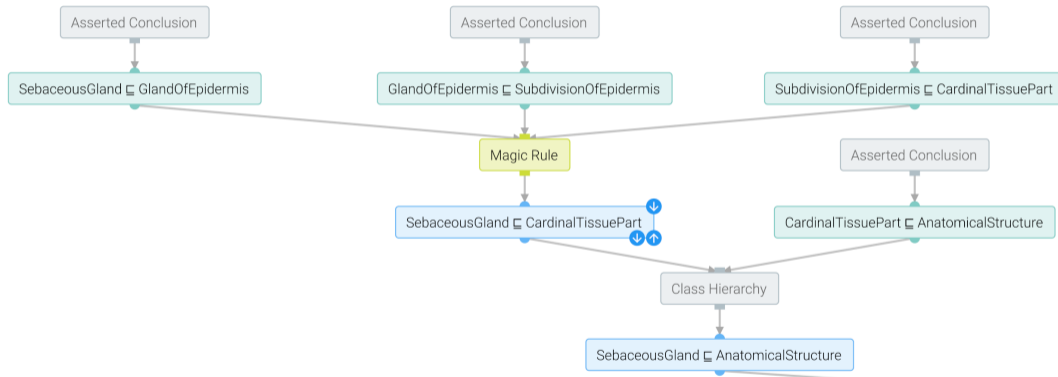


(Méndez, Alrabbaa, Koopmann, Langner, Baader, and Dachsel 2023)

EVONNE – EVEC’s bigger sister



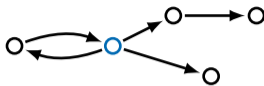
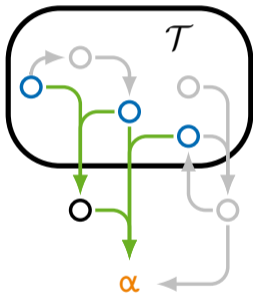
imld.de/evonne



(Méndez, Alrabbaa, Koopmann, Langner, Baader, and Dachsel 2023)

Outline

Proofs



Missing entailments

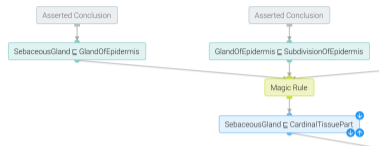


User studies



Verbalization

User studies

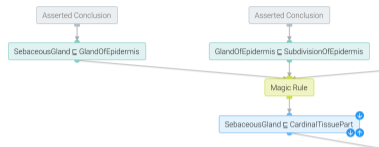


Proof format

(Alrabbaa, B, Hirsch, Knieriemen, Kovtunova, Rothermel, and Wiehr 2022)

- Many prefer **shorter**, **tree-shaped**, **interactive** proofs
- Preferences are **subjective**
- IQ test (ICAR16) predicts performance on proofs

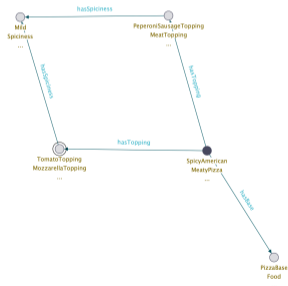
User studies



Proof format

(Alrabbaa, B, Hirsch, Knieriemen, Kovtunova, Rothermel, and Wiehr 2022)

- Many prefer **shorter**, **tree-shaped**, **interactive** proofs
- Preferences are **subjective**
- IQ test (ICAR16) predicts performance on proofs



Usability of E-VEE

(Alrabbaa, B, Friese, Hirsch, Knieriemen, Koopmann, Kovtunova, Krüger, Popovič, and Siahaan 2024)

- Preferred explanation methods depend on **users** and **use cases**

Designing user studies

domain-specific / abstract performance / preferences experts / students / laypeople

Since every cell culture is a material object and every material object contains an atom, every cell culture contains an atom. From the facts that every cell culture contains an atom and that every cell culture contains a cell and contains only cells, it follows that every cell culture contains something which is both an atom and a cell.

Every cell is a compound. Thus, any object which is an atom and a cell at the same time is also an atom and a compound. There is no object which is an atom and a compound at the same time. Therefore, there is no object which is both an atom and a cell.

Furthermore, since every cell culture contains something which is both an atom and a cell and there is no object which is both an atom and a cell, there is no cell culture.

Designing user studies

domain-specific / abstract performance / preferences experts / students / laypeople

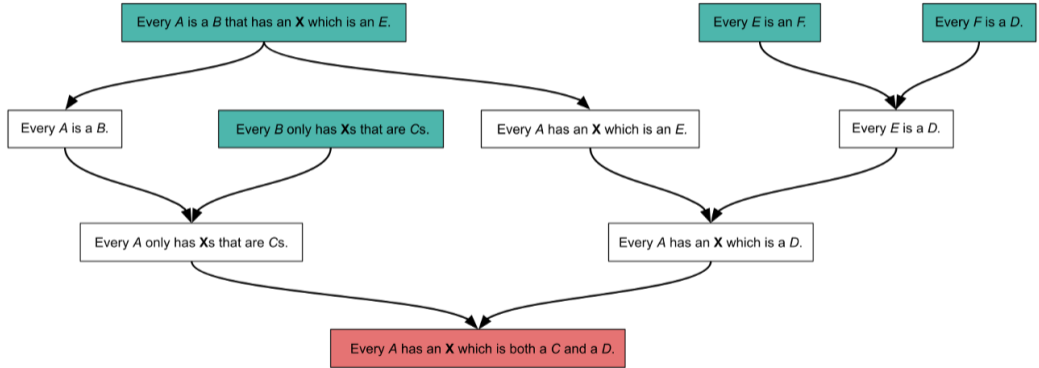
$$\frac{\frac{\frac{C\text{It} \sqsubseteq \exists \text{ct}.C \sqcap \forall \text{ct}.C}{C\text{It} \sqsubseteq \exists \text{ct}.(At \sqcap C)} \quad \frac{\frac{C\text{It} \sqsubseteq \text{MaObj} \quad \text{MaObj} \sqsubseteq \exists \text{ct}.At}{C\text{It} \sqsubseteq \exists \text{ct}.At}}{C\text{It} \sqsubseteq \perp} \quad \frac{\frac{C \sqsubseteq \text{Cmp}}{At \sqcap C \sqsubseteq At \sqcap \text{Cmp}} \quad At \sqcap \text{Cmp} \sqsubseteq \perp}{At \sqcap C \sqsubseteq \perp}}{C\text{It} \sqsubseteq \perp}$$

Designing user studies

domain-specific / abstract

performance / preferences

experts / students / laypeople



Designing user studies

domain-specific / abstract performance / preferences experts / students / laypeople

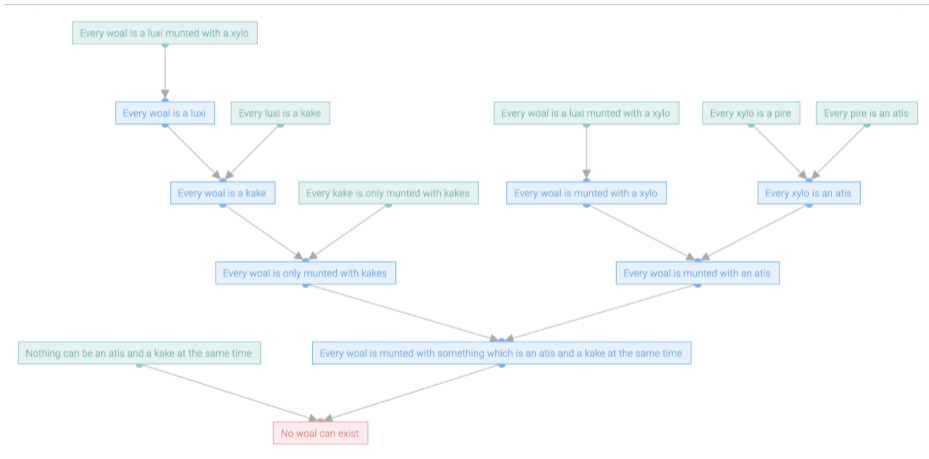
Since every woal is a luxi that is munted with a xylo,
every woal is a luxi.
Since every luxi is a kake,
every woal is a kake.
Moreover, every kake is only munted with kakes,
and thus every woal is only munted with kakes.

Every xylo is a pire
and every pire is an atis,
which means that every xylo is an atis.
Since every woal is a luxi that is munted with a xylo,
every woal is munted with a xylo.
Since every xylo is an atis,
we know that every woal is munted with an atis.

We have inferred that every woal is only munted with kakes
and that every woal is munted with an atis.
Therefore, every woal is munted with something which is both an atis and a kake.
However, nothing can be an atis and a kake at the same time
which lets us conclude that woals cannot exist.

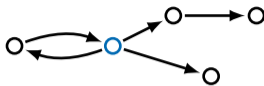
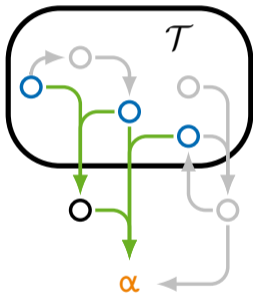
Designing user studies

domain-specific / abstract performance / preferences experts / students / laypeople



Outline

Proofs



Missing entailments



User studies



Verbalization

Drone handover scenario



Handover from autonomous drone to human operator
requires timely, short warning of a critical situation

(Wiehr, Hirsch, Schmitz, Knieriemen, Krüger, Kovtunova, B, Chang, Demberg, Steinmetz, and Hoffmann 2021)

Drone handover scenario



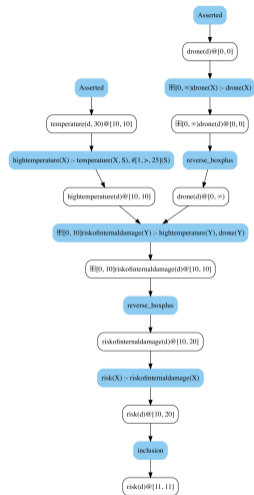
Handover from autonomous drone to human operator
requires timely, short warning of a critical situation

(Wiehr, Hirsch, Schmitz, Knieriemen, Krüger, Kovtunova, B, Chang, Demberg, Steinmetz, and Hoffmann 2021)

Verbalization of entailments and explanations:

- controlled natural language (Power 2012)
- template-based verbalization (Schiller, Schiller, and Glimm 2017)
- transformer models (Chang, Kovtunova, B, Demberg, Chapman, and Yeh 2022)
- large language models (B, Demberg, Jobanputra, Kovtunova, and Nhu 2024)

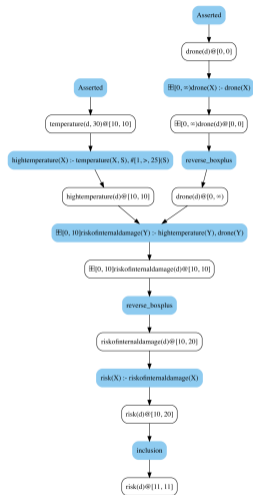
Proof verbalization with LLMs



(B, Demberg, Jobanputra, Kovtunova, and Nhu 2024)

Proofs provide more information than justifications

Proof verbalization with LLMs

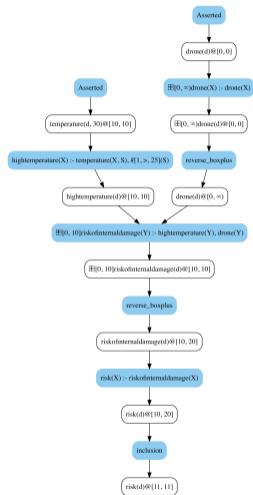


(B, Demberg, Jobanputra, Kovtunova, and Nhu 2024)

Proofs provide more information than justifications

Challenge: Find the important parts of the proof

Proof verbalization with LLMs



(B, Demberg, Jobanputra, Kovtunova, and Nhu 2024)

Proofs provide more information than justifications

Challenge: Find the important parts of the proof

“At time 10, there is a risk of internal damage to your drone (d) because it has a high temperature of 30 degrees Celsius. This is likely because the drone is flying.”

Inconsistent outputs, hallucinations

Summary

- Justifications are often enough, but sometimes proofs are needed
- Missing entailments also require explanations
- Computing explanations can take a long time
- Formal explanations are mainly useful for experts

Summary

- Justifications are often enough, but sometimes proofs are needed
- Missing entailments also require explanations
- Computing explanations can take a long time
- Formal explanations are mainly useful for experts

Future work:

- Find the most important steps in a proof
- Explain proofs to laypeople



Summary

- Justifications are often enough, but sometimes proofs are needed
- Missing entailments also require explanations
- Computing explanations can take a long time
- Formal explanations are mainly useful for experts

Future work:

- Find the most important steps in a proof
- Explain proofs to laypeople



Thank you!

References I

- Alrabbaa, Christian, Stefan B, Tom Frieze, Anke Hirsch, Nina Knieriemen, Patrick Koopmann, Alisa Kovtunova, Antonio Krüger, Alexej Popovič, and Ida Siahaan (2024). "Explaining Reasoning Results for OWL Ontologies with Eeve". In: *KR*. To appear.
- Alrabbaa, Christian, Stefan B, Tom Frieze, Patrick Koopmann, and Mikhail Kotlov (2023). "Why Not? Explaining Missing Entailments with Eeve". In: *DL*. URL: <https://ceur-ws.org/Vol-3515/paper-1.pdf>.
- Alrabbaa, Christian, Stefan B, Anke Hirsch, Nina Knieriemen, Alisa Kovtunova, Anna Milena Rothermel, and Frederik Wiehr (2022). "In the Head of the Beholder: Comparing Different Proof Representations". In: *RuleML+RR 2022*. DOI: 10.1007/978-3-031-21541-4_14.
- Alrabbaa, Christian, Stefan B, Patrick Koopmann, and Alisa Kovtunova (2022). "Finding Good Proofs for Answers to Conjunctive Queries Mediated by Lightweight Ontologies". In: *DL*. URL: <https://ceur-ws.org/Vol-3263/paper-3.pdf>.
- Alrabbaa, Christian, Franz Baader, Stefan B, Patrick Koopmann, and Alisa Kovtunova (2020). "Finding Small Proofs for Description Logic Entailments: Theory and Practice". In: *LPAR*. DOI: 10.29007/nhpp.
- (2021). "Finding Good Proofs for Description Logic Entailments using Recursive Quality Measures". In: *CADE*. DOI: 10.1007/978-3-030-79876-5_17.
- (2023). "Combining Proofs for Description Logic and Concrete Domain Reasoning". In: *RuleML+RR*. DOI: 10.1007/978-3-031-45072-3_4.
- Alrabbaa, Christian and Willi Hieke (2022). "Explaining Non-Entailment by Model Transformation for the Description Logic EL". In: *IJCKG*. DOI: 10.1145/3579051.3579060.
- B, Stefan (2021). *Concise Justifications Versus Detailed Proofs for Description Logic Entailments*. LTCS-Report 21-05. DOI: 10.25368/2023.225.
- B, Stefan, Vera Demberg, Mayank Jobanputra, Alisa Kovtunova, and Duy Nhu (2024). "Explaining Critical Situations Over Sensor Data Streams Using Proofs and Natural Language". In: *DL*. URL: <https://ceur-ws.org/Vol-3739/paper-2.pdf>.

References II

- Bauer, Johannes, Ulrike Sattler, and Bijan Parsia (2009). "Explaining by Example: Model Exploration for Ontology Comprehension". In: *DL*. URL: https://ceur-ws.org/Vol-477/paper_37.pdf.
- Borgida, Alexander, Diego Calvanese, and Mariano Rodriguez-Muro (2008). "Explanation in the DL-Lite Family of Description Logics". In: *OTM*. DOI: 10.1007/978-3-540-88873-4_35.
- Calvanese, Diego, Magdalena Ortiz, Mantas Simkus, and Giorgio Stefanoni (2013). "Reasoning about Explanations for Negative Query Answers in DL-Lite". In: *J. Artif. Intell. Res.* 48. DOI: 10.1613/jair.3870.
- Chang, Ernie, Alisa Kovtunova, Stefan B, Vera Demberg, Kathryn Chapman, and Hui-Syuan Yeh (2022). "Logic-Guided Message Generation from Raw Real-Time Sensor Data". In: *LREC*. URL: <https://aclanthology.org/2022.lrec-1.745>.
- Eiter, Thomas, Magdalena Ortiz, Mantas Simkus, Trung-Kien Tran, and Guohui Xiao (2012). "Query Rewriting for Horn-SHIQ Plus Rules". In: *AAAI*. DOI: 10.1609/aaai.v26i1.8219.
- Haifani, Fajar, Patrick Koopmann, Sophie Touret, and Christoph Weidenbach (2022). "Connection-Minimal Abduction in *EL* via Translation to FOL". In: *IJCAR*. DOI: 10.1007/978-3-031-10769-6_12.
- Horridge, Matthew, Samantha Bail, Bijan Parsia, and Uli Sattler (2013). "Toward cognitive support for OWL justifications". In: *KBS* 53. DOI: 10.1016/j.knosys.2013.08.021.
- Horridge, Matthew, Bijan Parsia, and Ulrike Sattler (2010). "Justification Oriented Proofs in OWL". In: *ISWC*. DOI: 10.1007/978-3-642-17746-0_23.
- Ivliev, Alex, Stefan Ellmauthaler, Lukas Gerlach, Maximilian Marx, Matthias Meißner, Simon Meusel, and Markus Krötzsch (2023). "Nemo: First Glimpse of a New Rule Engine". In: *ICLP*. DOI: 10.4204/eptcs.385.35.
- Kazakov, Yevgeny, Markus Krötzsch, and Frantisek Simancik (2014). "The Incredible ELK - From Polynomial Procedures to Efficient Reasoning with \mathcal{EL} Ontologies". In: *J. Autom. Reason.* 53.1. DOI: 10.1007/s10817-013-9296-3.

References III

- Koopmann, Patrick, Warren Del-Pinto, Sophie Turret, and Renate A. Schmidt (2020). "Signature-Based Abduction for Expressive Description Logics". In: *KR*. DOI: 10.24963/kr.2020/59.
- Méndez, Julián, Christian Alrabbaa, Patrick Koopmann, Ricardo Langner, Franz Baader, and Raimund Dachsel (2023). "Evonne: A Visual Tool for Explaining Reasoning with OWL Ontologies and Supporting Interactive Debugging". In: *Comput. Graph. Forum* 42.6. DOI: 10.1111/cgf.14730.
- Peuter, Dennis (2024). "Applications for Symbol Elimination in Combination with Hierarchical Reasoning". PhD thesis. URL: <https://uko.opus.hbz-nrw.de/frontdoor/index/index/docId/2507>.
- Power, Richard (2012). "OWL Simplified English: A Finite-State Language for Ontology Editing". In: *CNL*. DOI: 10.1007/978-3-642-32612-7_4.
- Schiller, Marvin R. G., Florian Schiller, and Birte Glimm (2017). "Testing the Adequacy of Automated Explanations of EL Subsumptions". In: *DL*. URL: <https://ceur-ws.org/Vol-1879/paper43.pdf>.
- Schlobach, Stefan (2004). "Explaining Subsumption by Optimal Interpolation". In: *JELIA*. DOI: 10.1007/978-3-540-30227-8_35.
- Tena Cucala, David, Bernardo Cuenca Grau, and Ian Horrocks (2021). "Pay-as-you-go consequence-based reasoning for the description logic SROIQ". In: *Artif. Intell.* 298. DOI: 10.1016/j.artint.2021.103518.
- Wei-Kleiner, Fang, Zlatan Dragisic, and Patrick Lambrix (2014). "Abduction Framework for Repairing Incomplete EL Ontologies: Complexity Results and Algorithms". In: *AAAI*. DOI: 10.1609/aaai.v28i1.8858.
- Wiehr, Frederik, Anke Hirsch, Lukas Schmitz, Nina Knieriem, Antonio Krüger, Alisa Kovtunova, Stefan B, Ernie Chang, Vera Demberg, Marcel Steinmetz, and Jörg Hoffmann (2021). "Why Do I Have to Take Over Control? Evaluating Safe Handovers with Advance Notice and Explanations in HAD". In: *ICMI*. DOI: 10.1145/3462244.3479884.

References IV

Images:

Drones by [JanBaby on Pixabay](#), [CC0](#)

Users by [danilo on Openclipart](#), [CC0](#)

Speech bubble by [gramzon on Openclipart](#), [CC0](#)

Control tower by [Annabel_P on Pixabay](#), [CC0](#)

All other images/illustrations © CPEC