Efficient Axiom Pinpointing with EL2MCS

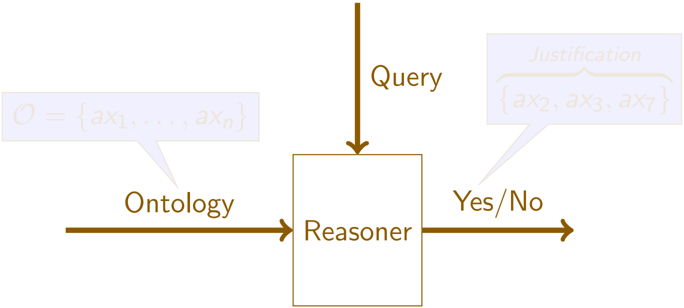
M. Fareed Arif[1], Carlos Mencía[1], Joao Marques-Silva[1,2]

[1]CASL/UCD, Ireland

[2]IST/INESC-ID, Portugal

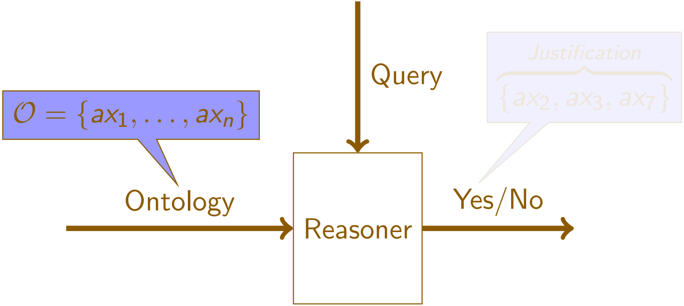
KI 2015 Dresden, Germany

Problem? - Axiom Pinpointing



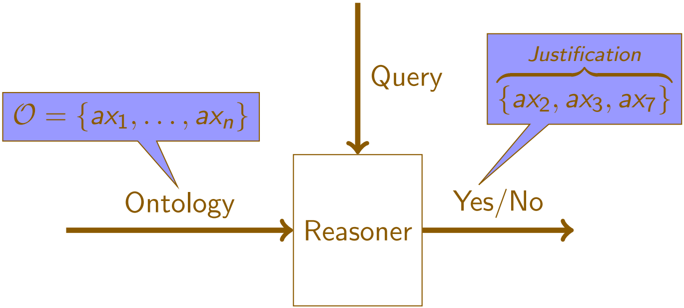
- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Problem? - Axiom Pinpointing



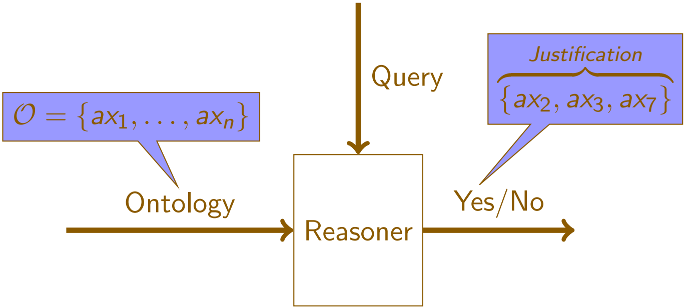
- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Problem? - Axiom Pinpointing



- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Problem? - Axiom Pinpointing



- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Preliminaries / Notations - Why Lightweight DLs?

Why $\mathcal{EL}^+$: $\mathcal{EL}^+$ is tractable and has been used for representing medical sciences ontologies , including the well-known SNOMED-CT.

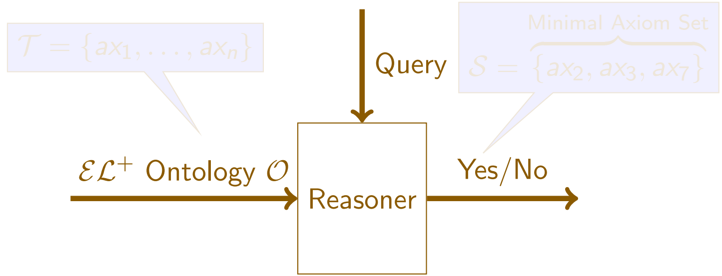
$\mathcal{EL}^+$ Syntax	$\mathcal{EL}^+$ Semantics
N_C, N_R	concept name set and role name set
$\Delta^\mathcal{I}$	non-empty domain set
$.^\mathcal{I}$	the mapping function
$\mathcal{I}$	a pair $(\Delta^\mathcal{I}, .^\mathcal{I})$
r	$r^\mathcal{I} = \Delta^\mathcal{I} \times \Delta^\mathcal{I}$
$\top$	$\top^\mathcal{I} = \Delta^\mathcal{I}$
A	$A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ for each $A \in N_C$
$C \sqcap D$	$(C \sqcap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$
$\exists r.C$	$(\exists r.C)^\mathcal{I} = \{x \in \Delta^\mathcal{I} \| y \in \Delta^\mathcal{I} : (x, y) \in r^\mathcal{I} \wedge y \in C^\mathcal{I}\}$
$GCI \{ \ C \sqsubseteq D$	$C^\mathcal{I} \subseteq D^\mathcal{I}$
$RI \{ \ r_1 \circ \ldots r_n \sqsubseteq s$	$r_1^\mathcal{I} \circ \cdots \circ r_n^\mathcal{I} \subseteq s^\mathcal{I}$
$\mathcal{T}$	set of GCIs and RIs
$C \sqsubseteq_\mathcal{T} D$	$C^\mathcal{I} \subseteq D^\mathcal{I}$ for every model $\mathcal{I}$ of $\mathcal{T}$
$\mathrm{PC}_\mathcal{T}$	$\top$ and all concept names used in $\mathcal{T}$
$\mathrm{PR}_\mathcal{T}$	all role names used in $\mathcal{T}$

We assume standard set theoretic semantics.

The main inference problem is concept subsumption and we are interested in the problem of Axiom Pinpointing (exists since mid 90s).

- Find a minimal axiom set (MinA) responsible for any subsumption relation in any given TBox $\mathcal{T}$.

Problem? - Axiom Pinpointing ($C \sqsubseteq_{\mathcal{S}} D$)

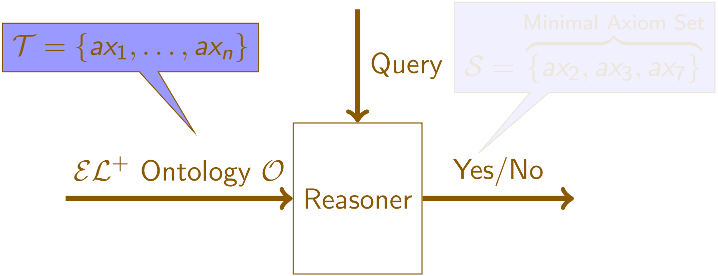


Definition (MinA)

Let $\mathcal{T}$ be an $\mathcal{EL}^+$ TBox, and let $C, D \in \mathrm{PC}_{\mathcal{T}}$ be primitive concept names, with $C \sqsubseteq_{\mathcal{T}} D$. Let $\mathcal{S}$ be a subset of $\mathcal{T}$ be such that $C \sqsubseteq_{\mathcal{S}} D$. If $\mathcal{S}$ is such that $C \sqsubseteq_{\mathcal{S}} D$ and $C \not\sqsubseteq_{\mathcal{S}'} D$ for $\mathcal{S}' \subset \mathcal{S}$, then $\mathcal{S}$ is a minimal axiom set (MinA) w.r.t. $C \sqsubseteq_{\mathcal{T}} D$.

- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Problem? - Axiom Pinpointing ($C \sqsubseteq_{\mathcal{S}} D$)

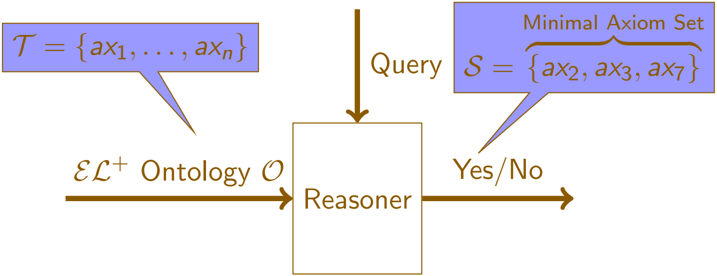


Definition (MinA)

Let $\mathcal{T}$ be an $\mathcal{EL}^+$ TBox, and let $C, D \in \mathrm{PC}_{\mathcal{T}}$ be primitive concept names, with $C \sqsubseteq_{\mathcal{T}} D$. Let $\mathcal{S}$ be a subset of $\mathcal{T}$ be such that $C \sqsubseteq_{\mathcal{S}} D$. If $\mathcal{S}$ is such that $C \sqsubseteq_{\mathcal{S}} D$ and $C \not\sqsubseteq_{\mathcal{S}'} D$ for $\mathcal{S}' \subset \mathcal{S}$, then $\mathcal{S}$ is a <u>minimal axiom set</u> (MinA) w.r.t. $C \sqsubseteq_{\mathcal{T}} D$.

- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Problem? - Axiom Pinpointing ($C \sqsubseteq_{\mathcal{S}} D$)



Definition (MinA)

Let $\mathcal{T}$ be an $\mathcal{EL}^+$ TBox, and let $C, D \in \mathrm{PC}_{\mathcal{T}}$ be primitive concept names, with $C \sqsubseteq_{\mathcal{T}} D$. Let $\mathcal{S}$ be a subset of $\mathcal{T}$ be such that $C \sqsubseteq_{\mathcal{S}} D$. If $\mathcal{S}$ is such that $C \sqsubseteq_{\mathcal{S}} D$ and $C \not\sqsubseteq_{\mathcal{S}'} D$ for $\mathcal{S}' \subset \mathcal{S}$, then $\mathcal{S}$ is a minimal axiom set (MinA) w.r.t. $C \sqsubseteq_{\mathcal{T}} D$.

- Ontology Debugging and Revision, Error-tolerant Reasoning, and Context-based Reasoning etc.

Preliminaries / Notations - Propositional Logic

Formulas are represented by $\mathcal{F}$, $\mathcal{M}$, $\mathcal{M}'$, $\mathcal{C}$ and $\mathcal{C}'$, but also by φ, ϕ, ψ and Φ.

Definition (MUS)

$\mathcal{M} \subseteq \mathcal{F}$ is a Minimal Unsatisfiable Subformula (MUS) of $\mathcal{F}$ iff $\mathcal{M}$ is unsatisfiable and $\forall_{\mathcal{M}' \subsetneq \mathcal{M}} \mathcal{M}'$ is satisfiable.

Definition (MCS)

$\mathcal{C} \subseteq \mathcal{F}$ is a Minimal Correction Subformula (MCS) of $\mathcal{F}$ iff $\mathcal{F} \setminus \mathcal{C}$ is satisfiable and $\forall_{\mathcal{C}' \subsetneq \mathcal{C}} \mathcal{F} \setminus \mathcal{C}'$ is unsatisfiable.

Theorem (Hitting Set Duality)

Each MCS of an unsatisfiable formula $\mathcal{F}$ is a minimal hitting set of the MUSes of $\mathcal{F}$ and vice-versa.

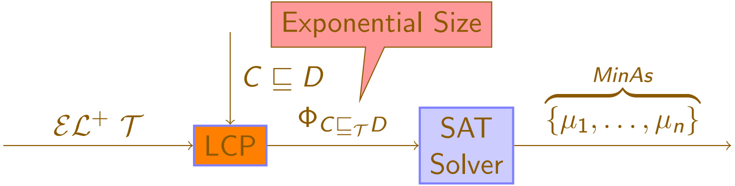
Definition (Partial MaxSAT)

A *partial MaxSAT* formula φ consists of a set of hard (φ_H) and soft (φ_S) clauses, i.e. $\varphi = \{\varphi_H, \varphi_S\}$. Hard clauses must be satisfied while soft clauses can be relaxed.

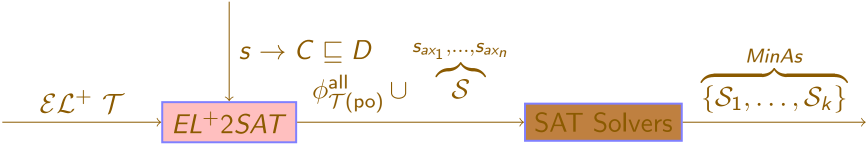
MUSes/MCSes - Example

[TODO]

Previous Work (Labeled-Based Completion Procedure)

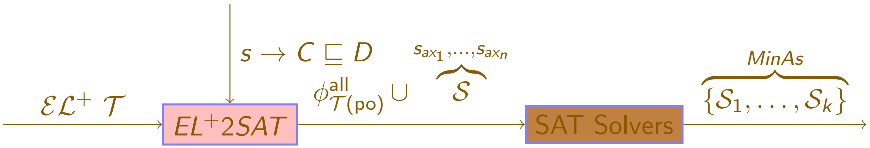


Previous Work (SAT and all-SMT based Procedure)



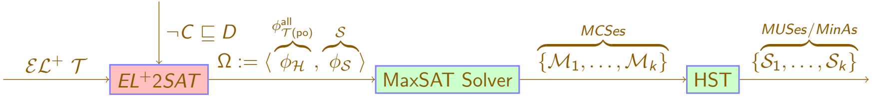
- Polynomial size Horn Encoding,
- Succinct modularity using Cone-of-Influence.

Previous Work (SAT and all-SMT based Procedure)



- Polynomial size Horn Encoding,
- Succinct modularity using Cone-of-Influence.

Solution (MaxSAT-based MCSes Enumeration)



MUSes a.k.a MinAs

- For axiom pinpointing, we can exploit state of the art MUSes and MCSes extraction algorithms and their minimal hitting set relationship.
- We can reduce SAT call by using clause set refinement, redundancy removal and model rotation.

Medical Ontology Example

$$
\begin{aligned}
&\mathsf{Endocarditis} \sqsubseteq \mathsf{Inflammation} \sqcap \exists \mathsf{hasLoc}.\mathsf{Endocardium},\\
&\mathsf{Inflammation} \sqsubseteq \mathsf{Disease} \sqcap \exists \mathsf{actsOn}.\mathsf{Tissue},\\
&\mathsf{Endocardium} \sqsubseteq \mathsf{Tissue} \sqcap \exists \mathsf{contIn}.\mathsf{HeartValve},\\
&\mathsf{HeartValve} \sqsubseteq \exists \mathsf{contIn}.\mathsf{Heart},\\
&\mathsf{HeartDisease} \equiv \mathsf{Disease} \sqcap \exists \mathsf{hasLoc}.\mathsf{Heart},\\
&\mathsf{contIn} \circ \mathsf{contIn} \sqsubseteq \mathsf{contIn},\\
&\mathsf{hasLoc} \circ \mathsf{contIn} \sqsubseteq \mathsf{hasLoc}
\end{aligned}
$$

$$\mathsf{Endocarditis} \sqsubseteq \mathsf{HeartDisease}$$

Mapping:

$s_1 \rightarrow \text{Endocarditis} \sqsubseteq \text{Inflammation}$
$s_2 \rightarrow \text{Inflammation} \sqsubseteq \text{Disease}$
$s_3 \rightarrow \text{Endocardium} \sqsubseteq \text{Tissue}$
$s_4 \rightarrow \text{HeartDisease} \sqsubseteq \text{Disease}$
$s_5 \rightarrow \text{Endocarditis} \sqsubseteq \exists\text{hasLoc.Endocardium}$
$s_6 \rightarrow \text{Inflammation} \sqsubseteq \exists\text{actsOn.Tissue}$
$s_7 \rightarrow \text{Endocardium} \sqsubseteq \exists\text{contIn.HeartValve}$
$s_8 \rightarrow \text{HeartValve} \sqsubseteq \exists\text{contIn.Heart}$
$s_9 \rightarrow \text{HeartDisease} \sqsubseteq \exists\text{hasLoc.Heart}$
$s_{10} \rightarrow \text{Disease} \sqcap \text{N} \sqsubseteq \text{HeartDisease}$
$s_{11} \rightarrow \exists\text{hasLocHeart} \sqsubseteq \text{N}$
$s_{12} \rightarrow \text{contIn} \circ \text{contIn} \sqsubseteq \text{contIn}$
$s_{13} \rightarrow \text{hasLoc} \circ \text{contIn} \sqsubseteq \text{hasLoc}$
$s_{14} \rightarrow \text{Endocarditis} \sqsubseteq \text{Disease}$
$s_{15} \rightarrow \text{Endocarditis} \sqsubseteq \exists\text{actsOn.Tissue}$
$s_{16} \rightarrow \text{Endocarditis} \sqsubseteq \exists\text{hasLoc.HeartValve}$
$s_{17} \rightarrow \text{Endocardium} \sqsubseteq \exists\text{contIn.Heart}$
$s_{18} \rightarrow \text{HeartDisease} \sqsubseteq \text{N}$
$s_{19} \rightarrow \text{Endocarditis} \sqsubseteq \exists\text{hasLocHeart} \leftarrow \text{N}$
$s_{20} \rightarrow \text{Endocarditis} \sqsubseteq \text{N}$
$s_{21} \rightarrow \textbf{Endocarditis} \sqsubseteq \textbf{HeartDisease}$

$\phi^{all}_{\mathcal{T}_{\text{MG}}(po)}$

$$
\begin{aligned}
& s_{a_1} \wedge s_{a_2} \rightarrow s_{a_4}, \\
& s_{a_6} \wedge s_{a_1} \rightarrow s_{a_{15}}, \\
& s_{a_{13}} \wedge s_{a_7} \wedge s_{a_5} \rightarrow s_{a_{16}}, \\
& s_{12} \wedge s_8 \wedge s_7 \rightarrow s_{17}, \\
& s_{11} \wedge s_9 \rightarrow s_{18}, \\
& s_{13} \wedge s_{16} \wedge s_8 \rightarrow s_{19}, \\
& s_{13} \wedge s_{17} \wedge s_5 \rightarrow s_{19}, \\
& s_{11} \wedge s_{19} \rightarrow s_{20}, \\
& s_{10} \wedge s_{14} \wedge s_{20} \rightarrow s_{21}, \\
& s_4 \wedge s_{21} \rightarrow s_{14}, \\
& s_9 \wedge s_{21} \rightarrow s_{19}
\end{aligned}
$$

$$
\begin{aligned}
& \phi_{\mathcal{H}} := \{\phi^{all}_{\mathcal{T}_{\text{MG}}(po)}\} \cup \{\neg s_{21}\} \\
& \phi_{\mathcal{S}} := \{s_1, s_2, \ldots, s_{13}\} \\
& \Omega := \langle \phi_{\mathcal{H}}, \phi_{\mathcal{S}} \rangle \\
& \textbf{MUS} := \{s_1, s_5, s_8, s_{10}, s_{11}, s_{13}\}
\end{aligned}
$$

Axiom pinpointing Tool $\Rightarrow$ EL2MCS

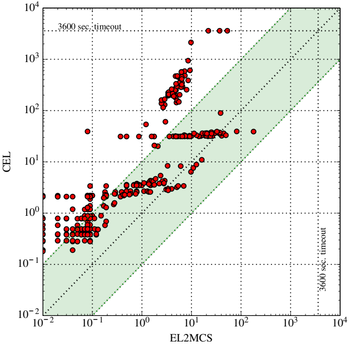
Tool Design:

Generate Horn formula (EL2SAT) → Partial MaxSAT encoding) → Compute all MCSes (CAMUS2) → Compute all MUSes (CAMUS)

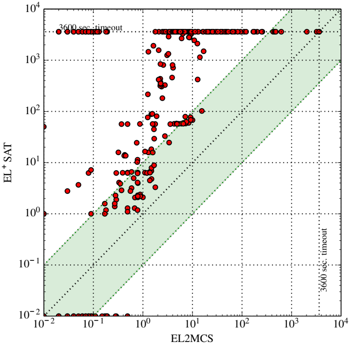
SHOW DEMO

Experimental Results (Plots)

EL2MCS vs CEL

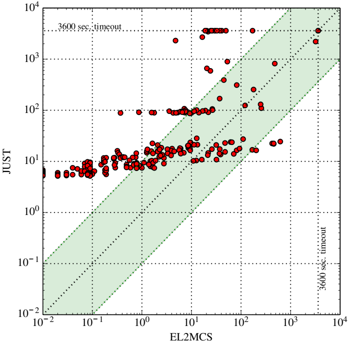


EL2MCS vs EL$^+$SAT

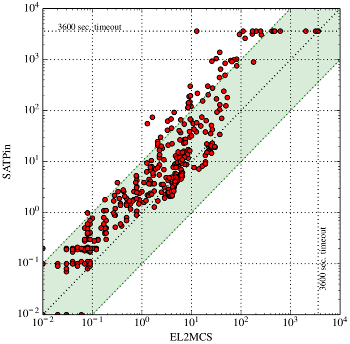


Experimental Results (Plots)

EL2MCS vs JUST



EL2MCS vs SATPin



Future Work

[TODO]