Efficient Axiom Pinpointing with EL2MCS

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

[1]CASL/UCD, Ireland

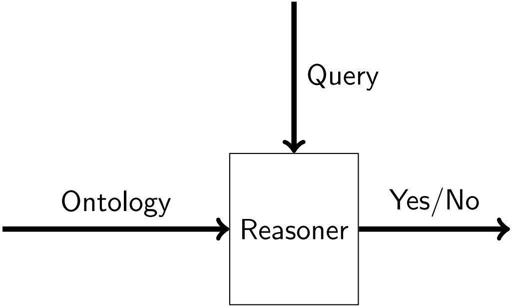
[2]IST/INESC-ID, Portugal

KI 2015

Outline

Outline

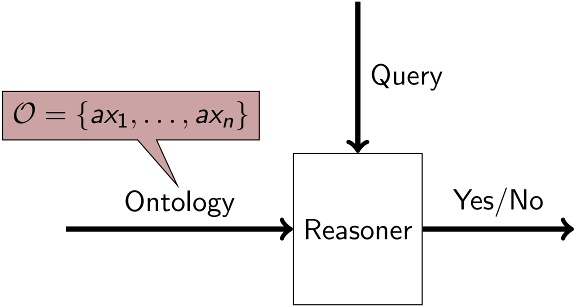
Problem? - Axiom Pinpointing



Applications:

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

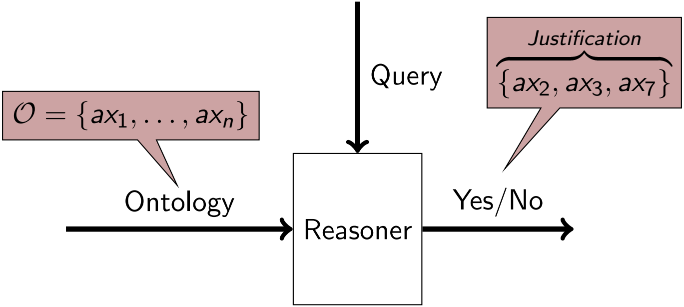
Problem? - Axiom Pinpointing



Applications:

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

Problem? - Axiom Pinpointing



Applications:

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

Outline

Outline

Lightweight Description Logic $\mathcal{EL}^+$

An $\mathcal{EL}^+$ ontology is defined using the following constructs and we assume standard set theoretic semantics.

$\mathcal{EL}^+$ Syntax	$\mathcal{EL}^+$ Semantics
N_C, N_R	concept name set and role name set
$\Delta^\mathcal{I}$	non-empty domain set
$\top$	$\Delta^\mathcal{I}$
$C \sqcap D$	$C^\mathcal{I} \cap D^\mathcal{I}$
$\exists r.C$	$\{x \in \Delta^\mathcal{I} \mid y \in \Delta^\mathcal{I} : (x, y) \in r^\mathcal{I} \wedge y \in C^\mathcal{I}\}$
GCI $\{$ $C \sqsubseteq D$	$C^\mathcal{I} \subseteq D^\mathcal{I}$
RI $\{$ $r_1 \circ \cdots \circ r_n \sqsubseteq s$	$r_1^\mathcal{I} \circ \cdots \circ r_n^\mathcal{I} \subseteq s^\mathcal{I}$

An $\mathcal{EL}^+$ TBox $\mathcal{T}$ is a finite set of GCIs and RIs.

$\mathcal{EL}^+$ Syntax	$\mathcal{EL}^+$ Semantics
$\mathrm{PC}_\mathcal{T}$	$\top$ and all concept names used in $\mathcal{T}$
$\mathrm{PR}_\mathcal{T}$	all role names used in $\mathcal{T}$
$C \sqsubseteq_\mathcal{T} D$	$C^\mathcal{I} \subseteq D^\mathcal{I}$ for every model $\mathcal{I}$ of $\mathcal{T}$

Medical Ontology Example

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

Lightweight Description Logic $\mathcal{EL}^+$

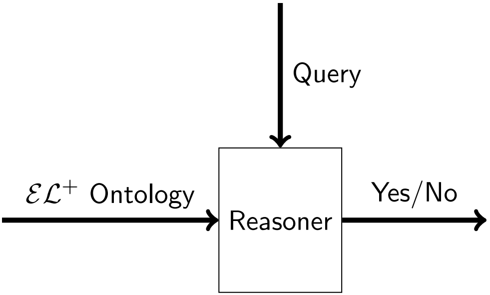
Why Lightweight DLs?

Lightweight description logic $\mathcal{EL}^+$ is tractable and used to represent many interesting medical sciences ontologies including well-known SNOMED CT.

The inference problems are:

- Concept Subsumption and Classification
- Axiom Pinpointing (exists since mid 90s): Find a Minimal Axiom Set (MinA) responsible for a subsumption relation between concepts in any given TBox $\mathcal{T}$.

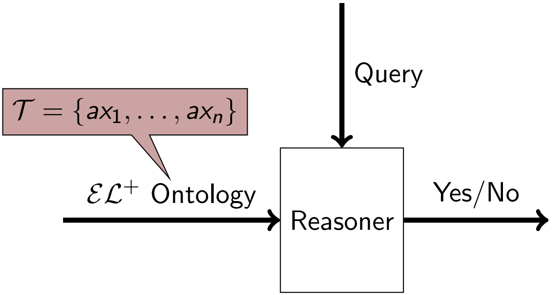
Axiom Pinpointing in Lightweight DLs ($C \sqsubseteq_{\mathcal{S}} D$)



Minimal Axiom Set (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. $C \sqsubseteq_{\mathcal{T}} D$.

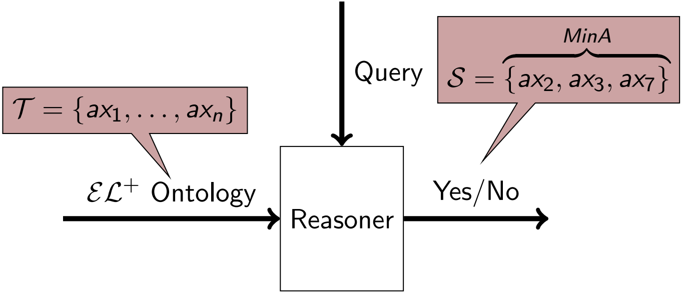
Axiom Pinpointing in Lightweight DLs ($C \sqsubseteq_{\mathcal{S}} D$)



Minimal Axiom Set (MinA):

Let $\mathcal{T}$ be an $\mathcal{EL}^+$ TBox, and let $C, D \in \mathsf{PC}_{\mathcal{T}}$ 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. $C \sqsubseteq_{\mathcal{T}} D$.

Axiom Pinpointing in Lightweight DLs ($C \sqsubseteq_{\mathcal{S}} D$)



Minimal Axiom Set (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. $C \sqsubseteq_{\mathcal{T}} D$.

Medical Ontology Example

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

$$
\overbrace{\mathsf{Endocarditis} \sqsubseteq \mathsf{HeartDisease}}^{MinAs}
$$

Outline

Minimal Unsatisfiability in Propositional Logic

Formulas are represented by $\mathcal{F}$, $\mathcal{M}$, $\mathcal{M}'$, $\mathcal{C}$ and $\mathcal{C}'$.

Minimal Unsatisfiable Subformula:

$\mathcal{M} \subseteq \mathcal{F}$ 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.

Minimal Correction Subset:

$\mathcal{C} \subseteq \mathcal{F}$ is a Minimal Correction Subset (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.

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.

Partial MaxSAT:

A *partial MaxSAT* formula Ω consists of a set of hard (φ_H) and soft (φ_S) clauses, i.e. $\Omega = \{\varphi_H, \varphi_S\}$.

Example

$$\phi_{\mathcal{H}} = \begin{pmatrix} \neg x_2 \vee \neg x_1 \vee x_{14} & \neg x_6 \vee \neg x_1 \vee x_{15} & \neg x_{13} \vee \neg x_7 \vee \neg x_5 \vee x_{16} \\ \neg x_{12} \vee \neg x_8 \vee \neg x_7 \vee x_{17} & \neg x_{11} \vee \neg x_9 \vee \neg x_{18} & \neg x_{13} \vee \neg x_{16} \vee \neg x_8 \vee x_{19} \\ \neg x_{13} \vee \neg x_{17} \vee \neg x_5 \vee x_{19} & \neg x_{11} \vee \neg x_{19} \vee x_{20} & \neg x_{10} \vee \neg x_{14} \vee \neg x_{20} \vee x_{21} \\ \neg x_4 \vee \neg x_{21} \vee x_{14} & \neg x_9 \vee \neg x_{21} \vee x_{19} & \neg x_{21} \end{pmatrix}$$

$$\phi_{\mathcal{S}} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} \end{pmatrix}$$

$$\Omega := \langle \varphi_{\mathcal{H}}, \varphi_{\mathcal{S}} \rangle$$

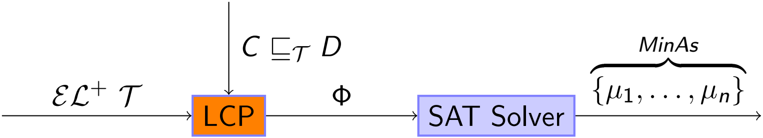
$$\text{MCSes} = \begin{pmatrix} \mathcal{C}_1 = \{x_1\} & \mathcal{C}_2 = \{x_5\} & \mathcal{C}_3 = \{x_8\} \\ \mathcal{C}_4 = \{x_{10}\} & \mathcal{C}_5 = \{x_{11}\} & \mathcal{C}_6 = \{x_{13}\} \end{pmatrix}$$

$$\text{MUSes} = \begin{pmatrix} \mathcal{M}_1 = \{x_1, x_5, x_8, x_{10}, x_{11}, x_{13}\} \end{pmatrix}$$

Outline

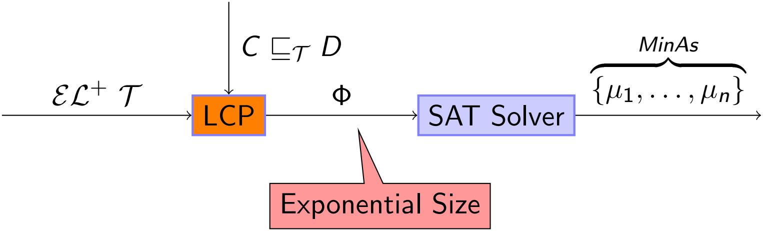
Labeled-based Completion Procedure

Labeled-based completion procedure generates a pinpointing formula Φ such that its minimal satisfying assignments of Φ corresponds to MinAs



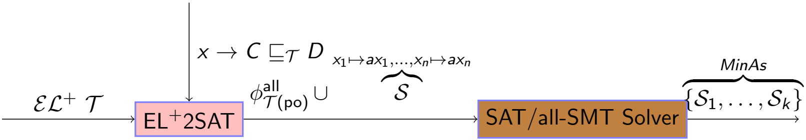
Labeled-based Completion Procedure

Labeled-based completion procedure generates a pinpointing formula Φ such that its minimal satisfying assignments of Φ corresponds to MinAs



SAT/all-SMT-based Approach

$\mathrm{EL}^{+}\mathrm{SAT}$ method encodes the classification of TBox as a Horn propositional formula $\phi^{\mathrm{all}}_{\mathcal{T}(\mathrm{po})}$ and enumerates over its satisfying models to obtain MinAs using SAT/all-SMT based algorithms



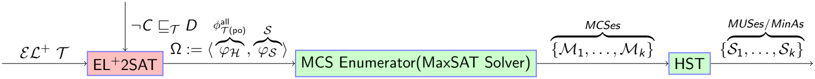
$$\phi^{\mathrm{all}}_{\mathcal{T}(\mathrm{po})} \Rightarrow x_1 \wedge x_2 \rightarrow x_4 \begin{cases} x_1 \rightarrow \text{Endocarditis} \sqsubseteq \text{Inflammation} \\ x_2 \rightarrow \text{Inflammation} \sqsubseteq \text{Disease} \\ x_4 \rightarrow \text{Endocarditis} \sqsubseteq \text{Disease} \end{cases}$$

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

Outline

EL2MCS

EL2MCS uses MaxSAT-based MCSes enumeration and hitting set duality to find MinAs



MUSes as MinAs

- Exploit state of the art MUSes and MCSes extraction algorithms and their minimal hitting set relationship
- Efficiently find all MinAs (if possible)

Medical Ontology Example

Mapping:

$x_1 \rightarrow \mathsf{Endocarditis} \sqsubseteq \mathsf{Inflammation}$
$x_2 \rightarrow \mathsf{Inflammation} \sqsubseteq \mathsf{Disease}$
$x_3 \rightarrow \mathsf{Endocardium} \sqsubseteq \mathsf{Tissue}$
$x_4 \rightarrow \mathsf{Endocarditis} \sqsubseteq \mathsf{Disease}$
$x_5 \rightarrow \mathsf{Endocarditis} \sqsubseteq \exists \mathsf{hasLoc.Endocardium}$
$x_6 \rightarrow \mathsf{Inflammation} \sqsubseteq \exists \mathsf{actsOn.Tissue}$
$x_7 \rightarrow \mathsf{Endocardium} \sqsubseteq \exists \mathsf{contIn.HeartValve}$
$x_8 \rightarrow \mathsf{HeartValve} \sqsubseteq \exists \mathsf{contIn.Heart}$
$x_9 \rightarrow \mathsf{HeartDisease} \sqsubseteq \exists \mathsf{hasLoc.Heart}$
$x_{10} \rightarrow \mathsf{Disease} \sqcap \mathsf{N} \sqsubseteq \mathsf{HeartDisease}$
$x_{11} \rightarrow \exists \mathsf{hasLoc.Heart} \sqsubseteq \mathsf{N}$
$x_{12} \rightarrow \mathsf{contIn} \circ \mathsf{contIn} \sqsubseteq \mathsf{contIn}$
$x_{13} \rightarrow \mathsf{hasLoc} \circ \mathsf{contIn} \sqsubseteq \mathsf{hasLoc}$
$x_{14} \rightarrow \mathsf{Endocarditis} \sqsubseteq \mathsf{Disease}$
$x_{15} \rightarrow \mathsf{Endocarditis} \sqsubseteq \exists \mathsf{actsOn.Tissue}$
$x_{16} \rightarrow \mathsf{Endocarditis} \sqsubseteq \exists \mathsf{hasLoc.HeartValve}$
$x_{17} \rightarrow \mathsf{Endocardium} \sqsubseteq \exists \mathsf{contIn.Heart}$
$x_{18} \rightarrow \mathsf{HeartDisease} \sqsubseteq \mathsf{N}$
$x_{19} \rightarrow \mathsf{Endocarditis} \sqsubseteq \exists \mathsf{hasLoc.Heart} \leftarrow \mathsf{N}$
$x_{20} \rightarrow \mathsf{Endocarditis} \sqsubseteq \mathsf{N}$
$x_{21} \rightarrow \mathbf{Endocarditis} \sqsubseteq \mathbf{HeartDisease}$

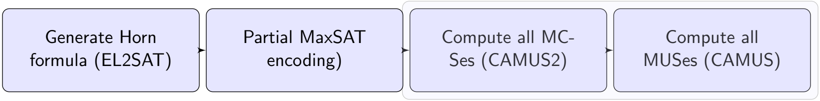
$\phi^{\mathsf{all}}_{\mathcal{T}(\mathsf{po})}$:

$$x_1 \wedge x_2 \rightarrow x_4$$
$$x_6 \wedge x_1 \rightarrow x_{15}$$
$$x_{13} \wedge x_7 \wedge x_5 \rightarrow x_{16}$$
$$x_{12} \wedge x_8 \wedge x_7 \rightarrow x_{17}$$
$$x_{11} \wedge x_9 \rightarrow x_{18}$$
$$x_{13} \wedge x_{16} \wedge x_8 \rightarrow x_{19}$$
$$x_{13} \wedge x_{17} \wedge x_5 \rightarrow x_{19}$$
$$x_{11} \wedge x_{19} \rightarrow x_{20}$$
$$x_{10} \wedge x_{14} \wedge x_{20} \rightarrow x_{21}$$
$$x_4 \wedge x_{21} \rightarrow x_{14}$$
$$x_9 \wedge x_{21} \rightarrow x_{19}$$

$$\varphi_{\mathcal{H}} := \{\phi^{\mathsf{all}}_{\mathcal{T}(\mathsf{po})}\} \cup \{\neg x_{21}\}$$
$$\varphi_{\mathcal{S}} := \{x_1, x_2, \ldots, x_{13}\}$$
$$\Omega := \langle \varphi_{\mathcal{H}}, \varphi_{\mathcal{S}} \rangle$$
$$\mathbf{MUS} := \{x_1, x_5, x_8, x_{10}, x_{11}, x_{13}\}$$

EL2MCS Tool

EL2MCS integrates CAMUS2 tool with CAMUS hitting set duality algorithm



Outline

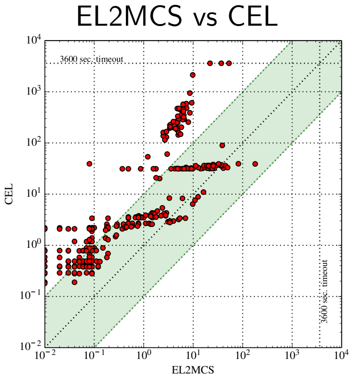
Experimentation & Results

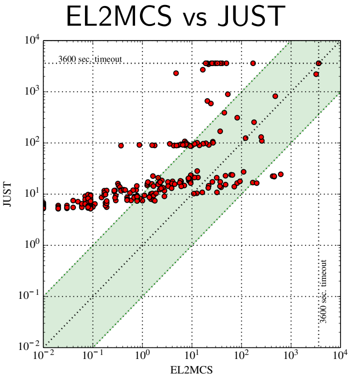
Experiment Setup:

- Tools: EL2MCS, SATPin, JUST, EL$^+$SAT and CEL
- Ontologies: GALEN, Gene, NCI and SNOMED CT
- 500 Query instances
- Timeout: 1 hour (3600 sec)

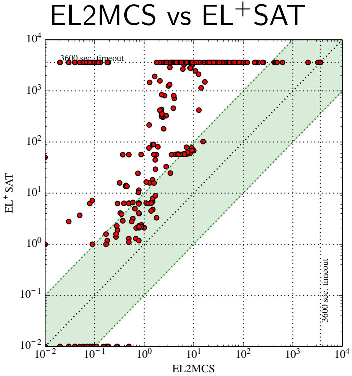
EL2MCS	vs **EL$^+$SAT**	vs **SATPin**	vs **CEL**	vs **JUST**
#Wins / #Losses	359 / 106	353 / 114	379 / 18	236 / 28
%Wins / %Losses	71.8% / 21.2%	70.6% / 22.8%	96.2% / 4.5%	80.8% / 9.6%

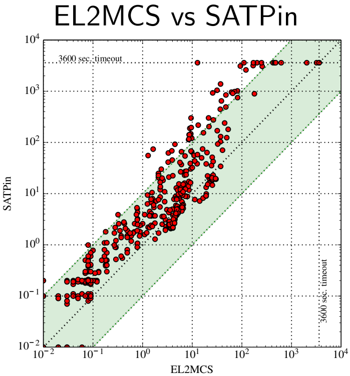
Experimental Results





Experimental Results





Outline

Conclusion

- EL2MCS exploits the relationship between MUSes a.k.a. MinAs and MCSes
- EL2MCS efficiently computes all MinAs for the terminating instances
- EL2MCS has been a stepping stone for our most recent work (HgMUS)
- HgMUS is an efficient group-MUS enumerator for Horn formulas

Questions?