Axiomatizing $\mathcal{EL}^{\bot}_{\text{gfp}}$-General Concept Inclusions in the Presence of Untrusted Examples

Daniel Borchmann

TU Dresden

July 24, 2013

Goal

Use description logic ontologies to represent knowledge of certain domains

Goal

Use description logic ontologies to represent knowledge of certain domains

Problem

How to obtain these ontologies?

Goal

Use description logic ontologies to represent knowledge of certain domains

Problem

How to obtain these ontologies?

Approach

Learn ontologies from domain data

Goal

Use description logic ontologies to represent knowledge of certain domains

Problem

How to obtain these ontologies?

Approach

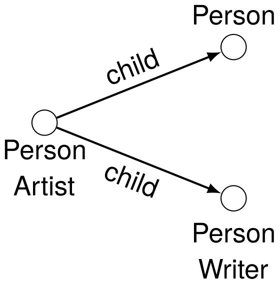
Learn first versions of ontologies from domain data

Goal

Extract terminological knowledge from factual knowledge.

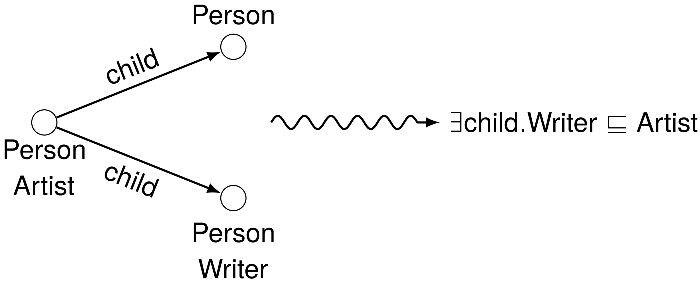
Goal

Extract terminological knowledge from factual knowledge.



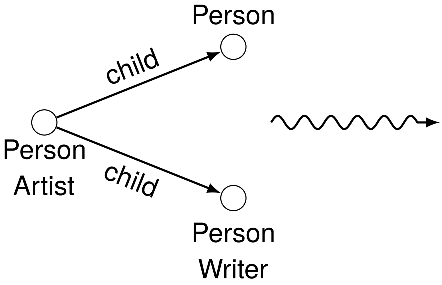
Goal

Extract terminological knowledge from factual knowledge.



Goal

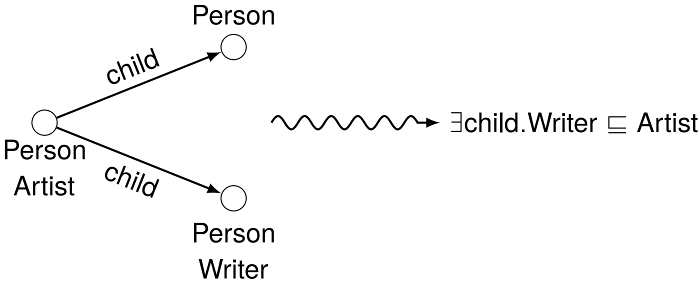
Extract terminological knowledge from interpretations.

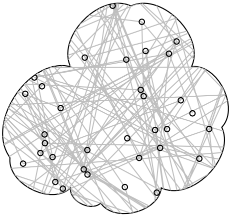


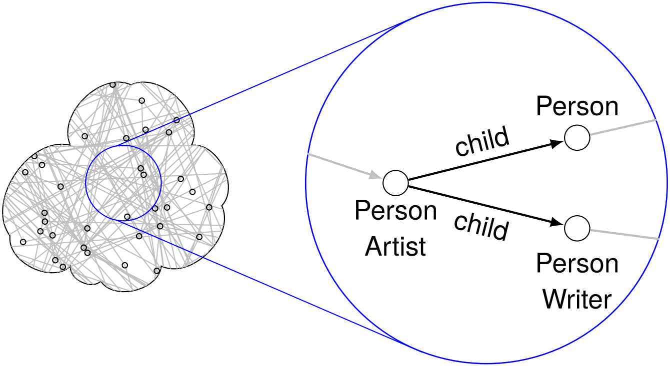
$\exists$child.Writer $\sqsubseteq$ Artist

Goal

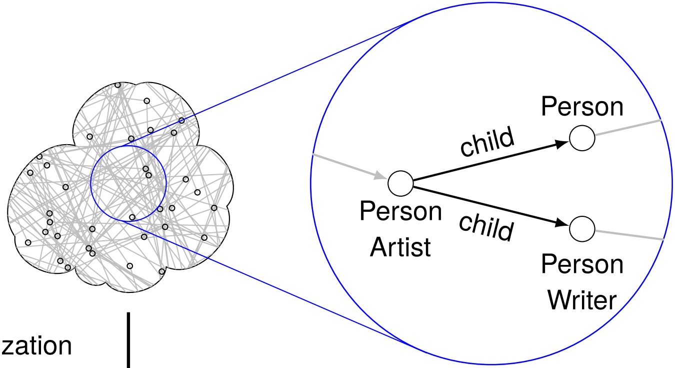
Extract finite bases of GCIs from interpretations.





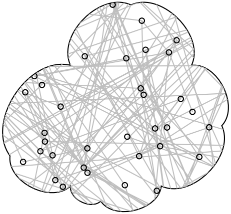


Person
child
Person
Artist
child
Person
Writer



Axiomatization
(Base of valid GCIs)

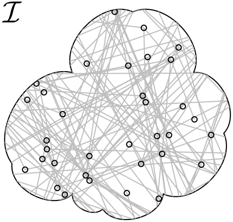
$\{\ldots, \exists \text{child}.\text{Writer} \sqsubseteq \text{Artist}, \ldots\}$



Axiomatization
(Base of valid GCIs)

$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$

$\mathcal{I}$

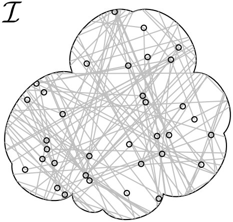


$\mathbb{K}_{\mathcal{I}}$	$M_{\mathcal{I}}$		
$\Delta^{\mathcal{I}}$	×	×	.
	×	.	×
	.	.	×

Axiomatization
(Base of valid GCIs)

$\{\ldots, \exists\text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$

$\mathcal{I}$



→

$\mathbb{K}_{\mathcal{I}}$	$M_{\mathcal{I}}$		
$\Delta^{\mathcal{I}}$	×	×	.
	×	.	×
	.	.	×

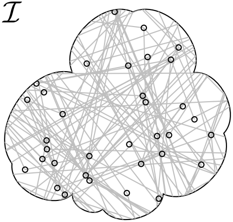
Axiomatization (Base of valid GCIs) ↓

$\{\ldots, \exists \text{child}.\text{Writer} \sqsubseteq \text{Artist}, \ldots\}$

Axiomatization (Base of valid Implications) ↓

$\{\, U \to V \mid \ldots \}$

$\mathcal{I}$



→

$\mathbb{K}_{\mathcal{I}}$	$M_{\mathcal{I}}$		
$\Delta^{\mathcal{I}}$	×	×	.
	×	.	×
	.	.	×

Axiomatization (Base of valid GCIs) ↓

Axiomatization (Base of valid Implications) ↓

$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$ ← $\{U \to V \mid \ldots\}$

Experiment

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$

$$\text{Criminal} \sqcap \exists \text{child}.\text{Politician} \sqsubseteq \bot$$

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$
$$\text{Criminal} \sqcap \exists \text{child.Politician} \sqsubseteq \bot$$
$$\text{Person} \sqcap \exists \text{child.Criminal} \sqsubseteq \text{Criminal}$$

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$
$$\text{Criminal} \sqcap \exists\text{child.Politician} \sqsubseteq \bot$$
$$\text{Person} \sqcap \exists\text{child.Criminal} \sqsubseteq \text{Criminal}$$

Observation

$$\exists\text{child.}\top \sqsubseteq \text{Person}$$

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$

$$\text{Criminal} \sqcap \exists \text{child.Politician} \sqsubseteq \bot$$

$$\text{Person} \sqcap \exists \text{child.Criminal} \sqsubseteq \text{Criminal}$$

Observation

$$\exists \text{child}.\top \sqsubseteq \text{Person}$$

does not hold in $\mathcal{I}_{\text{DBpedia}}$

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$
$$\text{Criminal} \sqcap \exists\text{child.Politician} \sqsubseteq \bot$$
$$\text{Person} \sqcap \exists\text{child.Criminal} \sqsubseteq \text{Criminal}$$

Observation

$$\exists\text{child}.\top \sqsubseteq \text{Person}$$

does not hold in $\mathcal{I}_{\text{DBpedia}}$, because of 4 counterexamples: `Teresa_Carpio`, `Charles_Heung`, `Adam_Cheng`, `Lydia_Shum`.

Experiment

- DBpedia, child-relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, size of base 1252

Some Results

$$\text{Criminal} \sqsubseteq \text{Person}$$
$$\text{Criminal} \sqcap \exists \text{child.Politician} \sqsubseteq \bot$$
$$\text{Person} \sqcap \exists \text{child.Criminal} \sqsubseteq \text{Criminal}$$

Observation

$$\exists \text{child}.\top \sqsubseteq \text{Person}$$

does not hold in $\mathcal{I}_{\text{DBpedia}}$, because of 4 erroneous counterexamples: `Teresa_Carpio`, `Charles_Heung`, `Adam_Cheng`, `Lydia_Shum`.

Observation

$$\mathsf{conf}_{\mathcal{I}_{\text{DBpedia}}}(\exists \mathsf{child}.\top \sqsubseteq \mathsf{Person}) = \frac{2547}{2551}$$

Observation

$$\mathsf{conf}_{\mathcal{I}_{\text{DBpedia}}}(\exists \mathsf{child}.\top \sqsubseteq \mathsf{Person}) = \frac{2547}{2551}$$

Definition

The *confidence* of $C \sqsubseteq D$ in $\mathcal{I}$ is defined as

$$\mathsf{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1 & \text{if } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \sqcap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{otherwise.} \end{cases}$$

Observation

$$\mathrm{conf}_{\mathcal{I}_{\mathrm{DBpedia}}}(\exists \mathrm{child}.\top \sqsubseteq \mathrm{Person}) = \frac{2547}{2551}$$

Definition

The *confidence* of $C \sqsubseteq D$ in $\mathcal{I}$ is defined as

$$\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1 & \text{if } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \sqcap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{otherwise.} \end{cases}$$

Let $c \in [0, 1]$. Define $\mathrm{Th}_c(\mathcal{I})$ as the set of all GCIs having confidence of at least c in $\mathcal{I}$.

Observation

$$\mathrm{conf}_{\mathcal{I}_{\mathrm{DBpedia}}}(\exists \mathrm{child}.\top \sqsubseteq \mathrm{Person}) = \frac{2547}{2551}$$

Definition

The *confidence* of $C \sqsubseteq D$ in $\mathcal{I}$ is defined as

$$\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1 & \text{if } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \sqcap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{otherwise.} \end{cases}$$

Let $c \in [0, 1]$. Define $\mathrm{Th}_c(\mathcal{I})$ as the set of all GCIs having confidence of at least c in $\mathcal{I}$.

Approach

Consider GCIs *with high confidence* in the data

Observation

$$\mathrm{conf}_{\mathcal{I}_{\mathrm{DBpedia}}}(\exists \mathrm{child}.\top \sqsubseteq \mathrm{Person}) = \frac{2547}{2551}$$

Definition

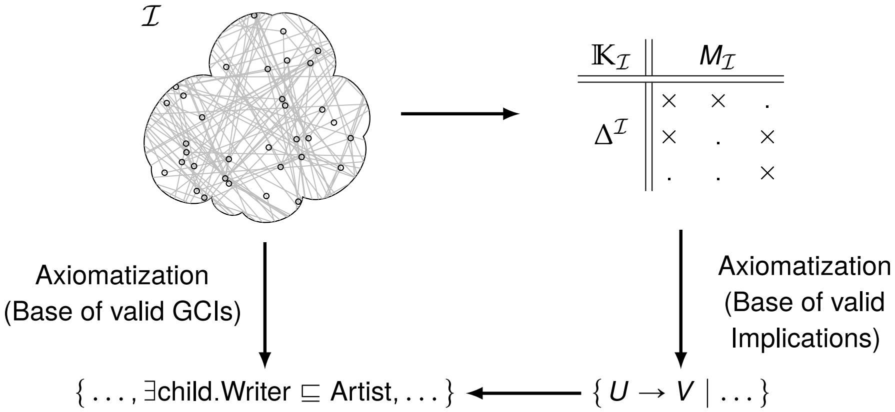
The *confidence* of $C \sqsubseteq D$ in $\mathcal{I}$ is defined as

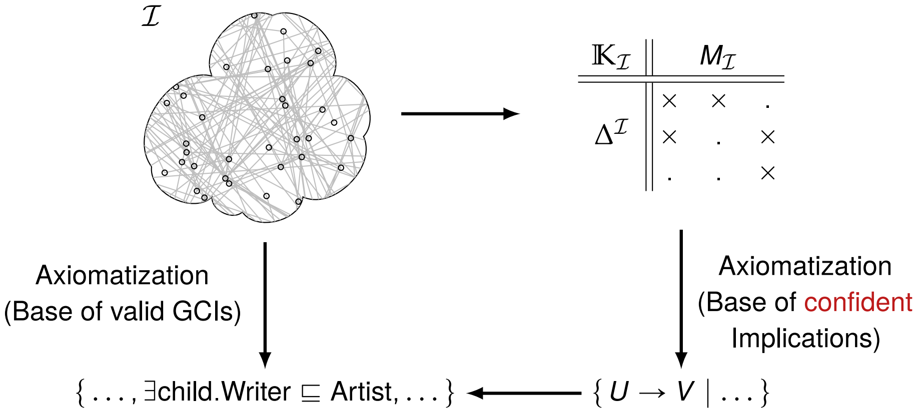
$$\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1 & \text{if } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \sqcap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{otherwise.} \end{cases}$$

Let $c \in [0, 1]$. Define $\mathrm{Th}_c(\mathcal{I})$ as the set of all GCIs having confidence of at least c in $\mathcal{I}$.

Approach

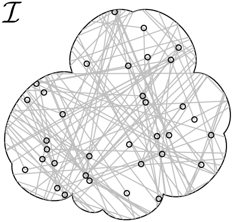
Consider GCIs *with high confidence* in the data, i. e. compute *finite bases* of $\mathrm{Th}_c(\mathcal{I})$


$\mathcal{I}$
$\mathbb{K}_{\mathcal{I}}$
$M_{\mathcal{I}}$
$\Delta^{\mathcal{I}}$
Axiomatization (Base of valid GCIs)
Axiomatization (Base of valid Implications)
$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$
$\{U \to V \mid \ldots\}$



$\mathcal{I}$
$\mathbb{K}_{\mathcal{I}}$
$M_{\mathcal{I}}$
$\Delta^{\mathcal{I}}$
Axiomatization
(Base of valid GCIs)
Axiomatization
(Base of confident
Implications)
$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$
$\{U \to V \mid \ldots\}$

$\mathcal{I}$



$\mathbb{K}_{\mathcal{I}}$	$M_{\mathcal{I}}$		
	×	×	.
$\Delta^{\mathcal{I}}$	×	.	×
	.	.	×

Axiomatization (Base of confident GCIs)

Axiomatization (Base of confident Implications)

$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\} \longleftarrow \{U \to V \mid \ldots\}$

Drawback

All individuals are suspected to be erroneous

Drawback

All individuals are suspected to be erroneous
(Problem of *rare counterexamples*)

Drawback

All individuals are suspected to be erroneous
(Problem of *rare counterexamples*)

Approach

Distinguish between *trusted* and *untrusted* individuals

Approach

Approach

- Consider untrusted individuals as *subinterpretation* $\mathcal{J}$ of $\mathcal{I}$

Approach

- Consider untrusted individuals as *subinterpretation* $\mathcal{J}$ of $\mathcal{I}$
- extract finite base of GCIs that are

Approach

- Consider untrusted individuals as *subinterpretation* $\mathcal{J}$ of $\mathcal{I}$
- extract finite base of GCIs that are
  - valid for all trusted individuals

Approach

- Consider untrusted individuals as *subinterpretation* $\mathcal{J}$ of $\mathcal{I}$
- extract finite base of GCIs that are
  - valid for all trusted individuals
  - have enough confidence within $\mathcal{J}$.

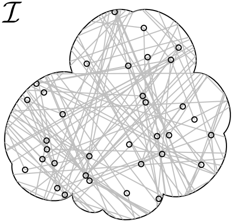
Approach

- Consider untrusted individuals as *subinterpretation* $\mathcal{J}$ of $\mathcal{I}$
- extract finite base of GCIs that are
  - valid for all trusted individuals
  - have enough confidence within $\mathcal{J}$.

Definition

$$\mathrm{Th}_c(\mathcal{I}, \mathcal{J}) := \{\, C \sqsubseteq D \mid C^{\mathcal{I}} \setminus \Delta^{\mathcal{J}} \subseteq D^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}, \\ |(C \sqcap D)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |C^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \,\}$$

$\mathcal{I}$



→

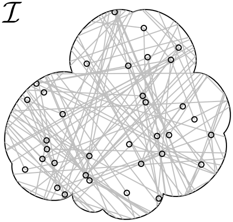
$\mathbb{K}_{\mathcal{I}}$	$M_{\mathcal{I}}$		
$\Delta^{\mathcal{I}}$	×	×	.
	×	.	×
	.	.	×

Axiomatization (Base of confident GCIs) ↓

Axiomatization (Base of confident Implications) ↓

$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\} \longleftarrow \{U \rightarrow V \mid \ldots\}$

$\mathcal{I}$



$\mathbb{K}_{\mathcal{I}}$	$M_{\mathcal{I}}$
$\Delta^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}$	trusted
$\Delta^{\mathcal{J}}$	untrusted

Axiomatization (Base of confident GCIs)

Axiomatization (Base of confident Implications)

$\{ \dots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \dots \} \longleftarrow \{ U \to V \mid \dots \}$

Theorem

Let

$$\mathsf{Conf}(\mathcal{I}, c, \mathcal{J}) = \{\, (C^{\mathcal{I}})^{\mathcal{I}} \sqsubseteq (D^{\mathcal{I}})^{\mathcal{I}} \mid (C \sqsubseteq D) \in \mathsf{Th}_c(\mathcal{I}, \mathcal{J}) \,\}.$$

Theorem

Let

$$\mathsf{Conf}(\mathcal{I}, c, \mathcal{J}) = \{ (C^{\mathcal{I}})^{\mathcal{I}} \sqsubseteq (D^{\mathcal{I}})^{\mathcal{I}} \mid (C \sqsubseteq D) \in \mathsf{Th}_c(\mathcal{I}, \mathcal{J}) \}.$$

If then $\mathcal{B}$ *is a finite base of* $\mathsf{Th}(\mathcal{I})$*, then the set*

$$\mathcal{B} \cup \mathsf{Conf}(\mathcal{I}, c, \mathcal{J})$$

is a finite base of $\mathsf{Th}_c(\mathcal{I}, \mathcal{J})$*.*

Summary

- Extended approach by Baader and Distel to setting of *trusted* and *untrusted* individuals
- Provided effective methods to computes bases in this setting

Summary

- Extended approach by Baader and Distel to setting of *trusted* and *untrusted* individuals
- Provided effective methods to computes bases in this setting

Future Work

Summary

- Extended approach by Baader and Distel to setting of *trusted* and *untrusted* individuals
- Provided effective methods to computes bases in this setting

Future Work

- Consider initial data set as untrusted, add valid counterexamples later

Summary

- Extended approach by Baader and Distel to setting of *trusted* and *untrusted* individuals
- Provided effective methods to computes bases in this setting

Future Work

- Consider initial data set as untrusted, add valid counterexamples later
- Extend approach into an algorithm with *expert interaction*

Summary

- Extended approach by Baader and Distel to setting of *trusted* and *untrusted* individuals
- Provided effective methods to computes bases in this setting

Future Work

- Consider initial data set as untrusted, add valid counterexamples later
- Extend approach into an algorithm with *expert interaction*
- *Attribute Exploration*

Thank You