# Towards an Error-Tolerant Construction of $\mathcal{EL}^{\perp}$-Ontologies from Data using Formal Concept Analysis

Daniel Borchmann

TU Dresden

May 23, 2013

## Goal

Learn Knowledge from Data.

## Goal

Learn Knowledge from Data.

## Questions

## Goal

Learn Knowledge from Data.

## Questions

- From which kind of data?

## Goal

Learn Knowledge from Data.

## Questions

- From which kind of data?
- Which knowledge?

## Goal

Learn Knowledge from Data.

## Questions

- From which kind of data?
- Which knowledge?
- How to learn?

## Goal

Learn Knowledge from Data.

## Questions

- From which kind of data? *Interpretations*
- Which knowledge?
- How to learn?

## Goal

Learn Knowledge from Data.

## Questions

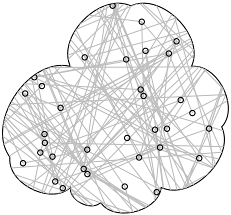
- From which kind of data? *Interpretations*
- Which knowledge? $\mathcal{EL}^{\bot}$-*General Concept Inclusions*
- How to learn?

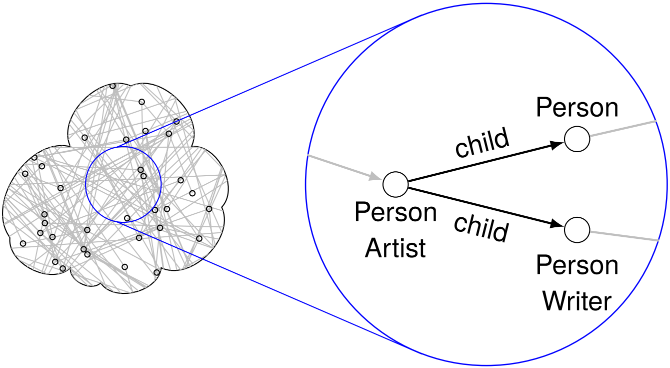
## Goal

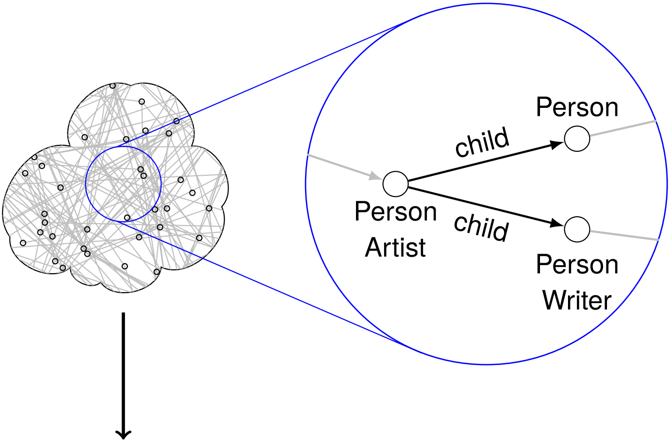
Learn Knowledge from Data.

## Questions

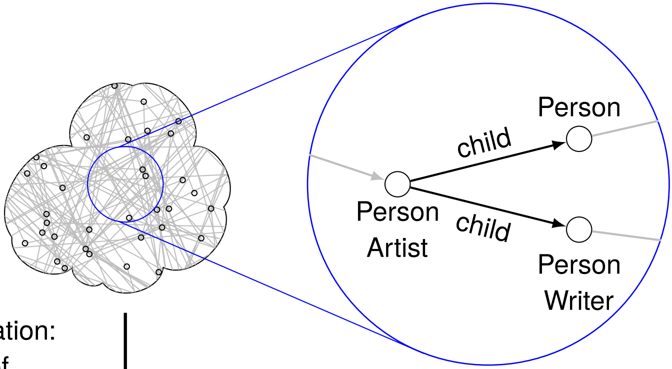
- From which kind of data? *Interpretations*
- Which knowledge? $\mathcal{EL}^{\bot}$*-General Concept Inclusions*
- How to learn? *Approach by Baader and Distel*




Person
child
Person
Artist
child
Person
Writer

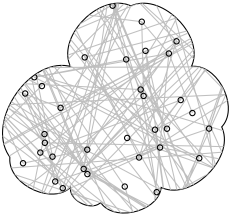



$\{\ldots, \exists \text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$



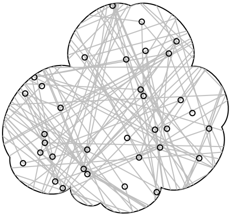

Axiomatization:
Base of
*valid* GCIs

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



Axiomatization:
Base of
*valid* GCIs

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

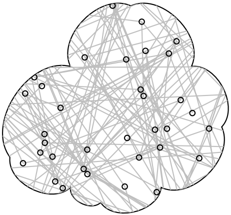


Axiomatization:
Base of
*valid* GCIs

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

**Problem**

What about *errors* in the data?



Axiomatization:
Base of
*valid* GCIs

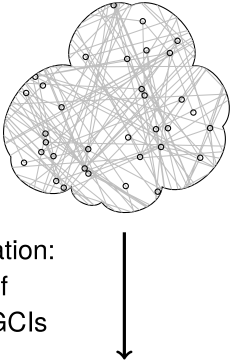
$\{\ldots, \exists \text{child}.\text{Writer} \sqsubseteq \text{Artist}, \ldots\}$

### Problem

What about *errors* in the data?

### Approach

Consider GCIs with *high confidence*



Axiomatization:
Base of
*confident* GCIs

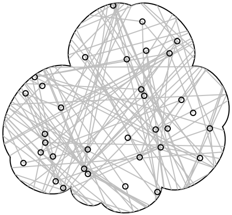
$\{\ldots, \exists\text{child.Writer} \sqsubseteq \text{Artist}, \ldots\}$

### Problem

What about *errors* in the data?

### Approach

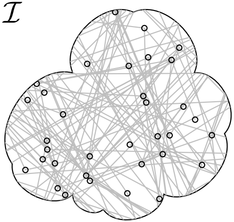
Consider GCIs with *high confidence*



Axiomatization
Base of valid GCIs

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

$\mathcal{I}$

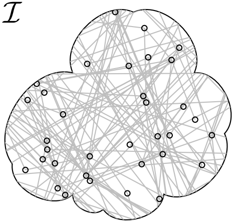


| $\mathbb{K}_{\mathcal{I}}$ | $M_{\mathcal{I}}$ |
| --- | --- |
| $\Delta^{\mathcal{I}}$ | $\nabla$ |

Axiomatization
Base of valid GCIs

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

$\mathcal{I}$



$\longrightarrow$

| $\mathbb{K}_{\mathcal{I}}$ | $M_{\mathcal{I}}$ |
|---|---|
| $\Delta^{\mathcal{I}}$ | $\nabla$ |

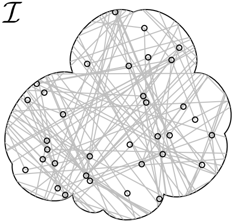
Axiomatization
Base of valid GCIs

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

Axiomatization
(Base of *valid* Implications)

$\{ U \to U'' \mid \ldots \}$

$\mathcal{I}$



$\longrightarrow$

| $\mathbb{K}_{\mathcal{I}}$ | $M_{\mathcal{I}}$ |
|---|---|
| $\Delta^{\mathcal{I}}$ | $\nabla$ |

Axiomatization
Base of valid GCIs

$\downarrow$

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

$\downarrow$

Axiomatization
(Base of *valid*
Implications)

## Interpretations

Vertex- and edge-labeled graphs (interpretations) $\mathcal{I}$ with

## Interpretations

Vertex- and edge-labeled graphs (interpretations) $\mathcal{I}$ with

- vertex set $\Delta^{\mathcal{I}}$,

## Interpretations

Vertex- and edge-labeled graphs (interpretations) $\mathcal{I}$ with

- vertex set $\Delta^{\mathcal{I}}$,
- vertex labels in $N_C$,

# Interpretations

Vertex- and edge-labeled graphs (interpretations) $\mathcal{I}$ with

- vertex set $\Delta^{\mathcal{I}}$,
- vertex labels in $N_C$,
- edge labels in $N_R$.

## Interpretations

Vertex- and edge-labeled graphs (interpretations) $\mathcal{I}$ with

- vertex set $\Delta^{\mathcal{I}}$,
- vertex labels in $N_C$,
- edge labels in $N_R$.

## Remark

Multiple labels per vertex/edge allowed

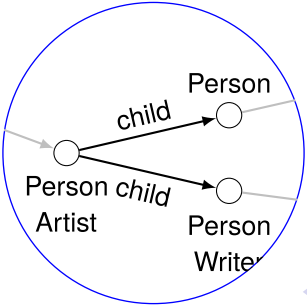
## Interpretations

Vertex- and edge-labeled graphs (interpretations) $\mathcal{I}$ with

- vertex set $\Delta^{\mathcal{I}}$,
- vertex labels in $N_C$,
- edge labels in $N_R$.

## Remark

Multiple labels per vertex/edge allowed



### Definition

Consider expressions of the form

$$C ::= A \mid C \sqcap C \mid \exists r.C \mid \bot \mid \top$$

for $A \in N_C, r \in N_R$.

**Definition**

Consider expressions of the form

$$C ::= A \mid C \sqcap C \mid \exists r.C \mid \bot \mid \top$$

for $A \in N_C, r \in N_R$. $C$ is called an $\mathcal{EL}^\perp$-*concept description*.

## Definition

Consider expressions of the form

$$C ::= A \mid C \sqcap C \mid \exists r.C \mid \bot \mid \top$$

for $A \in N_C, r \in N_R$. $C$ is called an $\mathcal{EL}^\bot$*-concept description*.

## Definition

For $A \in N_C$, $C, D$ two $\mathcal{EL}^\bot$-concept descriptions, $r \in N_R$:

- $\bot^\mathcal{I} = \emptyset$, $\top^\mathcal{I} = \Delta^\mathcal{I}$,
- $A^\mathcal{I}$ set of all vertices labeled with $A$,
- $(C \sqcap D)^\mathcal{I} := C^\mathcal{I} \cap D^\mathcal{I}$
- $(\exists r.C)^\mathcal{I} := \{\, x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I} \colon (x, y) \in r^\mathcal{I}, y \in C^\mathcal{I} \,\}$,

# General Concept Inclusions

*General Concept Inclusions* (GCIs) are of the form

$$C \sqsubseteq D$$

where $C, D$ are $\mathcal{EL}^{\bot}$-concept descriptions.

## General Concept Inclusions

*General Concept Inclusions* (GCIs) are of the form

$$C \sqsubseteq D$$

where $C, D$ are $\mathcal{EL}^{\bot}$-concept descriptions.
$C \sqsubseteq D$ *holds* in $\mathcal{I}$ if and only if

$$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$$

## General Concept Inclusions

*General Concept Inclusions* (GCIs) are of the form

$$C \sqsubseteq D$$

where $C, D$ are $\mathcal{EL}^{\bot}$-concept descriptions.
$C \sqsubseteq D$ *holds* in $\mathcal{I}$ if and only if

$$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$$

## Example

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

## General Concept Inclusions

*General Concept Inclusions* (GCIs) are of the form

$$C \sqsubseteq D$$

where $C, D$ are $\mathcal{EL}^{\bot}$-concept descriptions.
$C \sqsubseteq D$ *holds* in $\mathcal{I}$ if and only if

$$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$$

## Example

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

holds in $\mathcal{I}$ if and only if every individual having a child (-successor) which is a Writer is also an Artist.

## General Concept Inclusions

*General Concept Inclusions* (GCIs) are of the form

$$C \sqsubseteq D$$

where $C, D$ are $\mathcal{EL}^{\bot}$-concept descriptions.
$C \sqsubseteq D$ *holds* in $\mathcal{I}$ if and only if

$$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$$

## Example

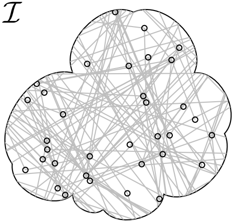
$$\exists\text{child.Writer} \sqsubseteq \text{Artist}$$

holds in $\mathcal{I}$ if and only if every individual having a child (-successor) which is a Writer is also an Artist.

## Learning Goal

Find *finite bases* of *valid* GCIs of $\mathcal{I}$.

$\mathcal{I}$



$\longrightarrow$

| $\mathbb{K}_{\mathcal{I}}$ | $M_{\mathcal{I}}$ |
|---|---|
| $\Delta^{\mathcal{I}}$ | $\nabla$ |

Axiomatization (Base of *valid* GCIs) $\downarrow$

Axiomatization (Base of *valid* Implications) $\downarrow$

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

## Outline

- Introduce *model-based most-specific concept descriptions* $X^{\mathcal{I}}$ for $X \subseteq \Delta^{\mathcal{I}}$

## Outline

- Introduce *model-based most-specific concept descriptions* $X^{\mathcal{I}}$ for $X \subseteq \Delta^{\mathcal{I}}$
- Define *induced context* $\mathbb{K}_{\mathcal{I}} = (\Delta^{\mathcal{I}}, M_{\mathcal{I}}, \nabla)$, where

$$(x, C) \in \nabla \iff x \in C^{\mathcal{I}}$$

## Outline

- Introduce *model-based most-specific concept descriptions* $X^{\mathcal{I}}$ for $X \subseteq \Delta^{\mathcal{I}}$
- Define *induced context* $\mathbb{K}_{\mathcal{I}} = (\Delta^{\mathcal{I}}, M_{\mathcal{I}}, \nabla)$, where

$$(x, C) \in \nabla \iff x \in C^{\mathcal{I}}$$

## Theorem (Baader, Distel 2008)

*If $\mathcal{L}$ is a base of $\mathbb{K}_{\mathcal{I}}$ then the set*

$$\{\, \bigsqcap U \sqsubseteq ((\bigsqcap U)^{\mathcal{I}})^{\mathcal{I}} \mid (U \to U'') \in \mathcal{L} \,\}$$

*is a finite base of $\mathcal{I}$.*

## Experiment

- DBpedia, child-relation

## Experiment

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

## Experiment

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

## Observation

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

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

### Experiment

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

### Observation

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

*does not hold* in $\mathcal{I}_{\text{DBpedia}}$, but there are only 4 *erroneous* counterexamples: `Teresa_Carpio`, `Charles_Heung`, `Adam_Cheng`, `Lydia_Shum`.

## Experiment

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

## Observation

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

*does not hold* in $\mathcal{I}_{\text{DBpedia}}$, but there are only 4 *erroneous* counterexamples: `Teresa_Carpio`, `Charles_Heung`, `Adam_Cheng`, `Lydia_Shum`.

## Observation

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

### 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 $\mathcal{I}$.

### 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 $\mathcal{I}$.

### New Goal

Axiomatize $\mathrm{Th}_c(\mathcal{I})$, i. e. find a *finite base* of $\mathrm{Th}_c(\mathcal{I})$.

## Question

How to find bases for $\mathrm{Th}_c(\mathcal{I})$?

### Question

How to find bases for $\mathrm{Th}_c(\mathcal{I})$?

### Observation

Related work by M. Luxenburger on *partial implications*

## Question

How to find bases for $\mathsf{Th}_c(\mathcal{I})$?

## Observation

Related work by M. Luxenburger on *partial implications*

## Approach (Luxenburger)

### Question

How to find bases for $\mathrm{Th}_c(\mathcal{I})$?

### Observation

Related work by M. Luxenburger on *partial implications*

### Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs

### Question

How to find bases for $\mathrm{Th}_c(\mathcal{I})$?

### Observation

Related work by M. Luxenburger on *partial implications*

### Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## The Idea in FCA

## Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## The Idea in FCA

Let $\mathcal{B}$ base of $\mathbb{K}$,

$$\mathcal{C} = \{ X'' \to Y'' \mid 1 > \operatorname{conf}_{\mathbb{K}}(X'' \to Y'') \geqslant c \}$$

and $1 > \operatorname{conf}(A \to B) \geqslant c$.

## Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## The Idea in FCA

Let $\mathcal{B}$ base of $\mathbb{K}$,

$$\mathcal{C} = \{\, X'' \to Y'' \mid 1 > \mathrm{conf}_{\mathbb{K}}(X'' \to Y'') \geqslant c \,\}$$

and $1 > \mathrm{conf}(A \to B) \geqslant c$.
Then $\mathrm{conf}_{\mathbb{K}}(A \to B) = \mathrm{conf}_{\mathbb{K}}(A'' \to B'')$

### Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

### The Idea in FCA

Let $\mathcal{B}$ base of $\mathbb{K}$,

$$\mathcal{C} = \{\, X'' \to Y'' \mid 1 > \operatorname{conf}_{\mathbb{K}}(X'' \to Y'') \geqslant c \,\}$$

and $1 > \operatorname{conf}(A \to B) \geqslant c$.
Then $\operatorname{conf}_{\mathbb{K}}(A \to B) = \operatorname{conf}_{\mathbb{K}}(A'' \to B'')$ and

$$\mathcal{B} \cup \mathcal{C} \models A \to A''$$

## Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## The Idea in FCA

Let $\mathcal{B}$ base of $\mathbb{K}$,

$$\mathcal{C} = \{ X'' \to Y'' \mid 1 > \mathrm{conf}_{\mathbb{K}}(X'' \to Y'') \geqslant c \}$$

and $1 > \mathrm{conf}(A \to B) \geqslant c$.
Then $\mathrm{conf}_{\mathbb{K}}(A \to B) = \mathrm{conf}_{\mathbb{K}}(A'' \to B'')$ and

$$\mathcal{B} \cup \mathcal{C} \models A \to A'' \to B''$$

## Approach (Luxenburger)

- Separately axiomatize *valid* GCIs and *properly confident* GCIs
- Consider concept descriptions of the form $(C^{\mathcal{I}})^{\mathcal{I}}$ only

## The Idea in FCA

Let $\mathcal{B}$ base of $\mathbb{K}$,

$$\mathcal{C} = \{ X'' \to Y'' \mid 1 > \operatorname{conf}_{\mathbb{K}}(X'' \to Y'') \geqslant c \}$$

and $1 > \operatorname{conf}(A \to B) \geqslant c$.
Then $\operatorname{conf}_{\mathbb{K}}(A \to B) = \operatorname{conf}_{\mathbb{K}}(A'' \to B'')$ and

$$\mathcal{B} \cup \mathcal{C} \models A \to A'' \to B'' \to B$$

### Definition

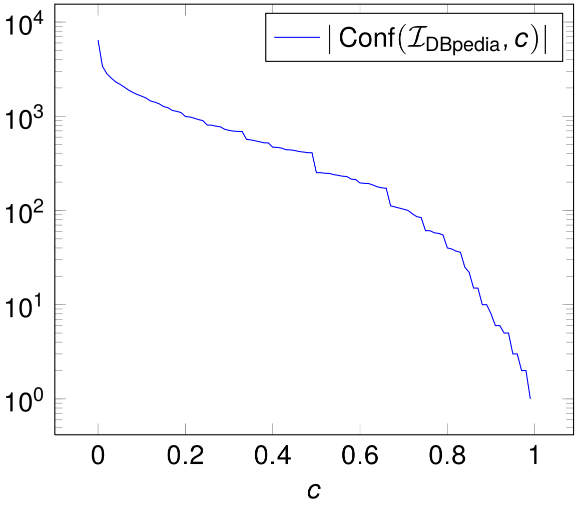
$$\mathsf{Conf}(\mathcal{I}, c) := \{\, X^{\mathcal{I}} \sqsubseteq Y^{\mathcal{I}} \mid Y \subseteq X \subseteq \Delta_{\mathcal{I}}, 1 > \mathsf{conf}_{\mathcal{I}}(X^{\mathcal{I}} \sqsubseteq Y^{\mathcal{I}}) \geqslant c \,\}.$$

**Definition**

$$\mathrm{Conf}(\mathcal{I}, c) := \{\, X^{\mathcal{I}} \sqsubseteq Y^{\mathcal{I}} \mid Y \subseteq X \subseteq \Delta_{\mathcal{I}}, 1 > \mathrm{conf}_{\mathcal{I}}(X^{\mathcal{I}} \sqsubseteq Y^{\mathcal{I}}) \geqslant c \,\}.$$

**Theorem**

*Let $\mathcal{B}$ be a finite base of $\mathcal{I}$, and $c \in [0, 1]$. Then $\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c)$ is a finite base of $\mathrm{Th}_c(\mathcal{I})$.*


$|\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, c)|$
$10^4$
$10^3$
$10^2$
$10^1$
$10^0$
0
0.2
0.4
0.6
0.8
1
$c$

## Further Results

## Further Results

- “Smaller” bases via another idea from Luxenburger

## Further Results

- “Smaller” bases via another idea from Luxenburger
- “Minimal Completion”

## Further Results

- “Smaller” bases via another idea from Luxenburger
- “Minimal Completion”
- Bases of $\mathrm{Th}_c(\mathcal{I})$ directly from bases of confident implications of $\mathbb{K}_{\mathcal{I}}$

## Further Results

- "Smaller" bases via another idea from Luxenburger
- "Minimal Completion"
- Bases of $\mathrm{Th}_c(\mathcal{I})$ directly from bases of confident implications of $\mathbb{K}_{\mathcal{I}}$

## Open Questions

## Further Results

- "Smaller" bases via another idea from Luxenburger
- "Minimal Completion"
- Bases of $\mathrm{Th}_c(\mathcal{I})$ directly from bases of confident implications of $\mathbb{K}_{\mathcal{I}}$

## Open Questions

- Minimal bases of $\mathrm{Th}_c(\mathcal{I})$?

## Further Results

- "Smaller" bases via another idea from Luxenburger
- "Minimal Completion"
- Bases of $\mathsf{Th}_c(\mathcal{I})$ directly from bases of confident implications of $\mathbb{K}_{\mathcal{I}}$

## Open Questions

- Minimal bases of $\mathsf{Th}_c(\mathcal{I})$?
- Exploration of $\mathsf{Th}_c(\mathcal{I})$?

Thank You