Learning Terminological Knowledge with High Confidence from Erroneous Data

Daniel Borchmann

daniel.borchmann@mailbox.tu-dresden.de

Research Training Group 1763 "QuantLA"

16. June 2014

TECHNISCHE
UNIVERSITÄT
DRESDEN

QuantLA

Motivation and Problem Description

Goal
Use description logic ontologies to represent knowledge
of certain domains

Motivation and Problem Description

Goal

Use description logic ontologies to represent knowledge
of certain domains

Problem

How to obtain these ontologies?

Motivation and Problem Description

Goal

Use description logic ontologies to represent knowledge
of certain domains

Problem

How to obtain these ontologies?

Approach

Learn ontologies from domain data

Motivation and Problem Description

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
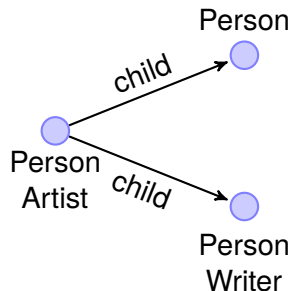
Motivation and Problem Description
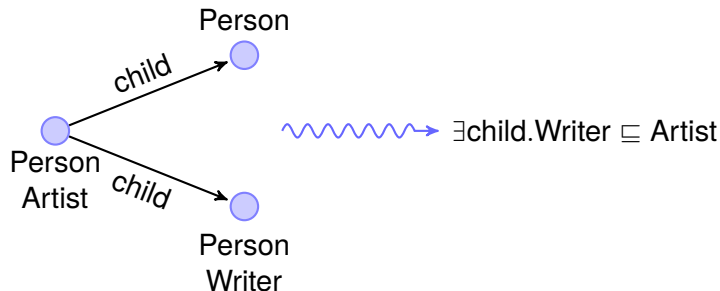
Goal

Extract terminological knowledge from factual knowledge.

Motivation and Problem Description

Goal

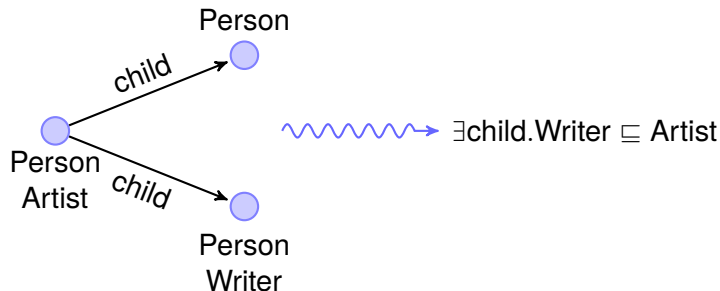Extract terminological knowledge from factual knowledge.

Motivation and Problem Description

Goal

Extract terminological knowledge from factual knowledge.
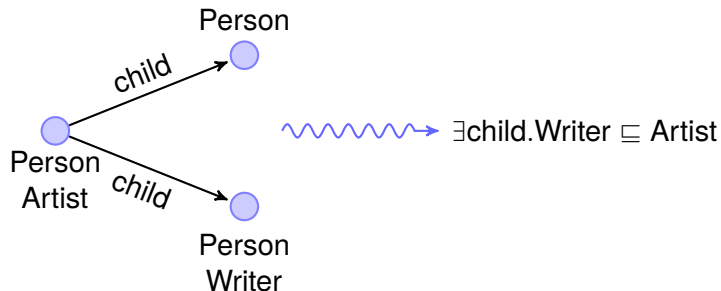
Motivation and Problem Description

Goal

Extract terminological knowledge from interpretations.

Motivation and Problem Description

Goal

Extract finite bases of GCIs from interpretations.

Experiment

Extracting Knowledge from DBpedia

Experiment

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

Extracting Knowledge from DBpedia

Experiment

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

Extracting Knowledge from DBpedia

Experiment

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

Some Results

Extracting Knowledge from DBpedia

Experiment

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

Some Results

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

Extracting Knowledge from DBpedia

Experiment

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

Some Results

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

Extracting Knowledge from DBpedia

Experiment

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

Some Results

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

Extracting Knowledge from DBpedia

Experiment

- DBpedia, child-relation $\leadsto \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, 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

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

Extracting Knowledge from DBpedia

Experiment

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

Some Results

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

Observation

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

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

Extracting Knowledge from 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`.

Extracting Knowledge from DBpedia

Experiment

- DBpedia, child-relation $\leadsto \mathcal{I}_{\text{DBpedia}}$
- $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$, 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`.

Confidence

Observation

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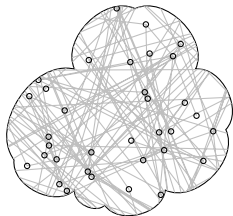
Confidence

Observation

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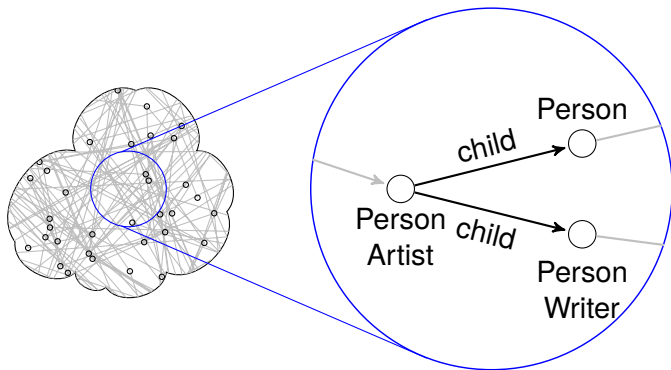
Approach

Consider GCIs with *high confidence* in $\mathcal{I}_{\text{DBpedia}}$ as well.
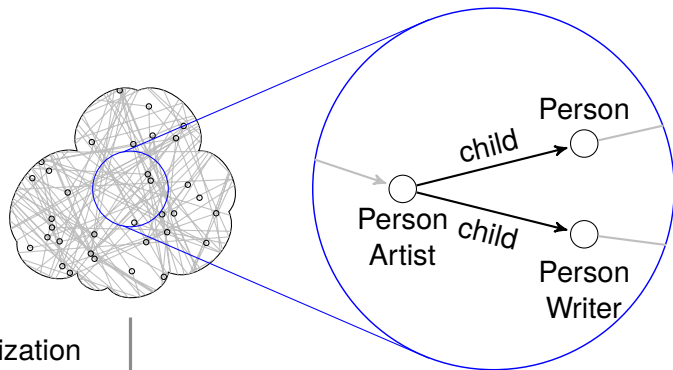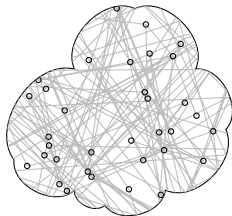
Method Overview

Method Overview

Axiomatization
(Base of valid GCIs)

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

Method Overview

Axiomatization
(Base of valid GCIs)

$\{ \exists child.Writer \sqsubseteq Artist, \dots \}$

Method Overview
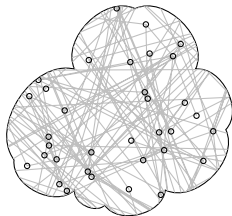
Description Logics

Formal Concept Analysis

Axiomatization
(Base of valid GCIs)

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

Method Overview

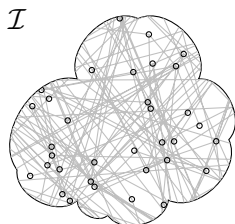*Description Logics*
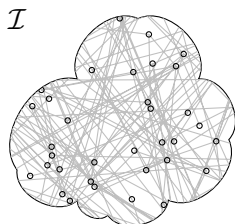
$\mathcal{I}$

Axiomatization
(Base of valid GCIs)

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

Formal Concept Analysis

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

Method Overview

Description Logics

$\mathcal{I}$

Axiomatization
(Base of valid GCIs)

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

Formal Concept Analysis

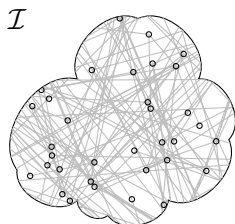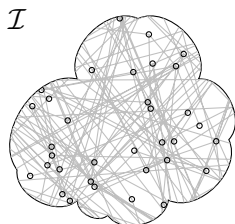$$\begin{array}{c|ccc} \mathbb{K}_{\mathcal{I}} & & M_{\mathcal{I}} & \\ \hline & \times & \times & . \\ \Delta^{\mathcal{I}} & \times & . & \times \\ & . & . & \times \end{array}$$

Axiomatization
(Base of valid
Implications)

$\{ U \rightarrow V \mid \dots \}$

Method Overview
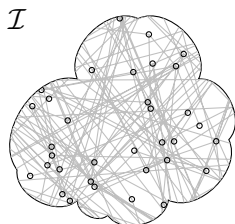
Description Logics | *Formal Concept Analysis*

$\mathcal{I}$

$$\begin{array}{c|ccc} \mathbb{K}_{\mathcal{I}} & & M_{\mathcal{I}} & \\ \hline & \times & \times & . \\ \Delta^{\mathcal{I}} & \times & . & \times \\ & . & . & \times \end{array}$$

Axiomatization
(Base of valid GCIs)

Axiomatization
(Base of valid
Implications)

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

Method Overview

Description Logics   *Formal Concept Analysis*

$\mathcal{I}$

$\mathbb{K}_{\mathcal{I}}$ | $M_{\mathcal{I}}$

	$\times$	$\times$	$.$
$\Delta^{\mathcal{I}}$	$\times$	$.$	$\times$
	$.$	$.$	$\times$

Axiomatization
(Base of confident GCIs)

Axiomatization
(Base of valid Implications)

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

Method Overview

Description Logics

$\mathcal{I}$

Formal Concept Analysis

$\mathbb{K}_{\mathcal{I}} \quad \quad M_{\mathcal{I}}$

$\Delta^{\mathcal{I}}$
$\times \quad \times \quad .$
$\times \quad . \quad \times$
$. \quad . \quad \times$

Axiomatization
(Base of confident GCIs)

Axiomatization
(Base of confident Implications)

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