

Lernen Terminologischen Wissens mit hoher Konfidenz aus fehlerhaften Daten

Daniel Borchmann

9. September 2014

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar

Beispiel (RDF-Triple aus DBpedia/Wikidata)

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar

Beispiel (RDF-Triple aus DBpedia/Wikidata)

```
<http://dbpedia.org/resource/Aldous_Huxley>  
  <http://dbpedia.org/ontology/notableWork>  
<http://dbpedia.org/resource/Brave_New_World> .
```

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar

Beispiel (RDF-Triple aus DBpedia/Wikidata)

```
<http://dbpedia.org/resource/Aldous_Huxley>  
  <http://dbpedia.org/ontology/notableWork>  
<http://dbpedia.org/resource/Brave_New_World> .
```

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar
- ▶ terminologisches (begriffliches) Wissen schwer extrahierbar

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar
- ▶ terminologisches (begriffliches) Wissen schwer extrahierbar

Beispiel

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar
- ▶ terminologisches (begriffliches) Wissen schwer extrahierbar

Beispiel

- ▶ Jede Katze ist ein Säugetier.

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar
- ▶ terminologisches (begriffliches) Wissen schwer extrahierbar

Beispiel

- ▶ Jede Katze ist ein Säugetier.
- ▶ Hunde sind keine Katzen.

Ziel

Wissen für maschinelle Bearbeitung aus Daten extrahieren

Beobachtung

- ▶ faktisches Wissen leicht extrahierbar
- ▶ terminologisches (begriffliches) Wissen schwer extrahierbar

Beispiel

- ▶ Jede Katze ist ein Säugetier.
- ▶ Hunde sind keine Katzen.
- ▶ Jeder Mensch, der ein Kind hat, ist ein Elternteil.

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten?

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten?
- ▶ Wie Wissen darstellen?

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten?
- ▶ Wie Wissen darstellen?
- ▶ Wie Wissen extrahieren?

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten?
- ▶ Wie Wissen darstellen?
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen?
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Beispiel (\mathcal{EL}^\perp , General Concept Inclusions (GCIs))

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Beispiel (\mathcal{EL}^\perp , General Concept Inclusions (GCIs))

- ▶ Katze \sqsubseteq Säugetier

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Beispiel (\mathcal{EL}^\perp , General Concept Inclusions (GCIs))

- ▶ Katze \sqsubseteq Säugetier
- ▶ Hund \sqcap Katze $\sqsubseteq \perp$

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren?
- ▶ Welches Wissen extrahieren?

Beispiel (\mathcal{EL}^\perp , General Concept Inclusions (GCIs))

- ▶ Katze \sqsubseteq Säugetier
- ▶ Hund \sqcap Katze $\sqsubseteq \perp$
- ▶ $\exists \text{hatKind.}\top \sqsubseteq \text{Elternteil}$

Ziel

Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

Grundlegende Fragen

- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren? \Rightarrow *Formale Begriffsanalyse*
- ▶ Welches Wissen extrahieren?

Beispiel (\mathcal{EL}^\perp , General Concept Inclusions (GCIs))

- ▶ Katze \sqsubseteq Säugetier
- ▶ Hund \sqcap Katze $\sqsubseteq \perp$
- ▶ $\exists \text{hatKind.}\top \sqsubseteq \text{Elternteil}$

Ziel

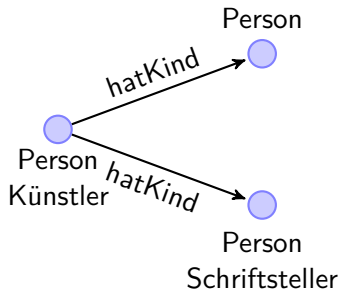
Terminologisches Wissen für maschinelle Bearbeitung aus Daten extrahieren

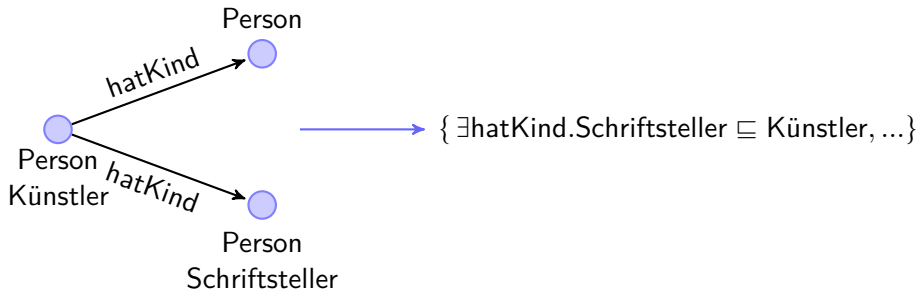
Grundlegende Fragen

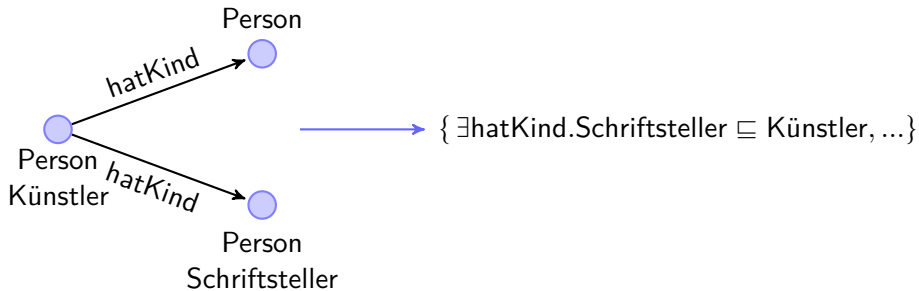
- ▶ Form der Daten? \Rightarrow *Relationale Daten*
- ▶ Wie Wissen darstellen? \Rightarrow *Beschreibungslogiken*
- ▶ Wie Wissen extrahieren? \Rightarrow *Formale Begriffsanalyse*
- ▶ Welches Wissen extrahieren? \Rightarrow „*interessantes*“

Beispiel (\mathcal{EL}^\perp , General Concept Inclusions (GCIs))

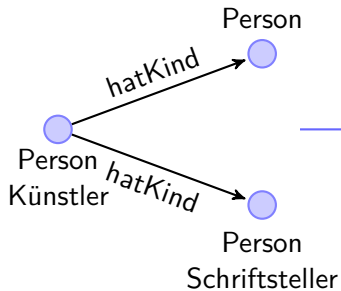
- ▶ Katze \sqsubseteq Säugetier
- ▶ Hund \sqcap Katze $\sqsubseteq \perp$
- ▶ $\exists \text{hatKind.}\top \sqsubseteq \text{Elternteil}$







Interpretation



→ { $\exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots$ }

Interpretation

Basis von GCI

Fixiere disjunkte Mengen N_C (Konzeptnamen) und N_R (Rollennamen).

Fixiere disjunkte Mengen N_C (Konzeptnamen) und N_R (Rollennamen).

Definition

\mathcal{EL}^\perp -Konzeptbeschreibungen C sind von der Form

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

für $A \in N_C, r \in N_R$.

Fixiere disjunkte Mengen N_C (Konzeptnamen) und N_R (Rollennamen).

Definition

\mathcal{EL}^\perp -Konzeptbeschreibungen C sind von der Form

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

für $A \in N_C, r \in N_R$.

Beispiel

Katze \sqcap Hund, \exists hatKind.Schriftsteller, \top, \perp

Definition

Eine *Interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ besteht aus

Definition

Eine *Interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ besteht aus

- ▶ einer nicht-leeren Menge $\Delta^{\mathcal{I}}$ von *Elementen*,

Definition

Eine *Interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ besteht aus

- ▶ einer nicht-leeren Menge $\Delta^{\mathcal{I}}$ von *Elementen*,
- ▶ einer Abbildung $\cdot^{\mathcal{I}}$ mit

$$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$$

$$r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$$

für $A \in N_C, r \in N_R$.

Definition

Eine *Interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ besteht aus

- ▶ einer nicht-leeren Menge $\Delta^{\mathcal{I}}$ von *Elementen*,
- ▶ einer Abbildung $\cdot^{\mathcal{I}}$ mit

$$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$$

$$r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$$

für $A \in N_C, r \in N_R$.

Definition

Für $A \in N_C$, C, D zwei \mathcal{EL}^{\perp} -Konzeptbeschreibungen und $r \in N_R$ sei

- ▶ $\perp^{\mathcal{I}} = \emptyset$, $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$,
- ▶ $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$,
- ▶ $(\exists r.C)^{\mathcal{I}} := \{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}}, y \in C^{\mathcal{I}}\}$.

Definition

Sind C, D zwei \mathcal{EL}^\perp -Konzeptbeschreibungen, so heißt

$$C \sqsubseteq D$$

Allgemeine Konzeptinklusion (General Concept Inclusion, GCI).

Definition

Sind C, D zwei \mathcal{EL}^\perp -Konzeptbeschreibungen, so heißt

$$C \sqsubseteq D$$

Allgemeine Konzeptinklusion (General Concept Inclusion, GCI).

$C \sqsubseteq D$ gilt in \mathcal{I} , $\mathcal{I} \models (C \sqsubseteq D)$, falls

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

Definition

$\mathbb{K} = (G, M, I)$ heißt *formaler Kontext*, falls G, M Mengen und $I \subseteq G \times M$.

Definition

$\mathbb{K} = (G, M, I)$ heißt *formaler Kontext*, falls G, M Mengen und $I \subseteq G \times M$.

Für $B \subseteq M$ sei

Definition

$\mathbb{K} = (G, M, I)$ heißt *formaler Kontext*, falls G, M Mengen und $I \subseteq G \times M$.

Für $B \subseteq M$ sei

$$B' = \{g \in G \mid \forall m \in B: (g, m) \in I\}.$$

Definition

$\mathbb{K} = (G, M, I)$ heißt *formaler Kontext*, falls G, M Mengen und $I \subseteq G \times M$.

Für $B \subseteq M$ sei

$$B' = \{g \in G \mid \forall m \in B: (g, m) \in I\}.$$

Definition

$X \rightarrow Y$ heißt *Implikation* in \mathbb{K} , falls $X, Y \subseteq M$.

Definition

$\mathbb{K} = (G, M, I)$ heißt *formaler Kontext*, falls G, M Mengen und $I \subseteq G \times M$.

Für $B \subseteq M$ sei

$$B' = \{g \in G \mid \forall m \in B: (g, m) \in I\}.$$

Definition

$X \rightarrow Y$ heißt *Implikation* in \mathbb{K} , falls $X, Y \subseteq M$. $X \rightarrow Y$ heißt *gültig* in \mathbb{K} , falls

$$X' \subseteq Y'.$$

Offene Frage

Was ist „interessantes“ Wissen?

Offene Frage

Was ist „interessantes“ Wissen?

Erste Idee [Baader, Distel 2009]

Offene Frage

Was ist „interessantes“ Wissen?

Erste Idee [Baader, Distel 2009]

Betrachte alle *gültigen* GCIs

Offene Frage

Was ist „interessantes“ Wissen?

Erste Idee [Baader, Distel 2009]

Betrachte alle *gültigen* GCIs \implies berechne *Basis* aller gültigen GCIs.

Offene Frage

Was ist „interessantes“ Wissen?

Erste Idee [Baader, Distel 2009]

Betrachte alle *gültigen* GCIs \implies berechne *Basis* aller gültigen GCIs.

Beobachtung

Offene Frage

Was ist „interessantes“ Wissen?

Erste Idee [Baader, Distel 2009]

Betrachte alle *gültigen* GCIs \implies berechne *Basis* aller gültigen GCIs.

Beobachtung

- ▶ GCIs sind Implikationen sehr ähnlich

Offene Frage

Was ist „interessantes“ Wissen?

Erste Idee [Baader, Distel 2009]

Betrachte alle *gültigen* GCIs \implies berechne *Basis* aller gültigen GCIs.

Beobachtung

- ▶ GCIs sind Implikationen sehr ähnlich
- ▶ FBA bietet Möglichkeiten, Implikationen aus Daten zu berechnen

Offene Frage

Was ist „interessantes“ Wissen?

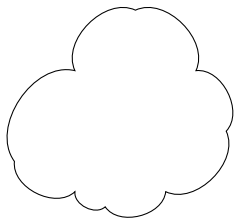
Erste Idee [Baader, Distel 2009]

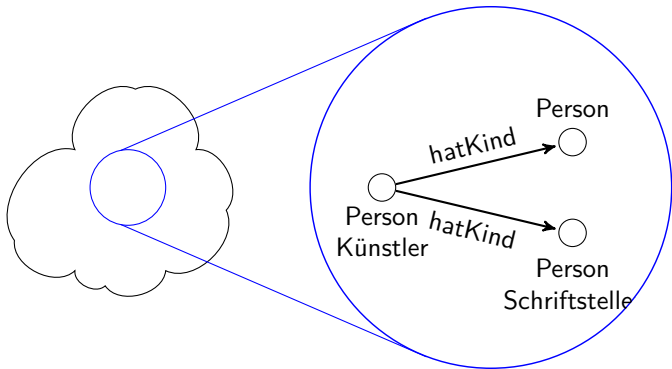
Betrachte alle *gültigen* GCIs \implies berechne *Basis* aller gültigen GCIs.

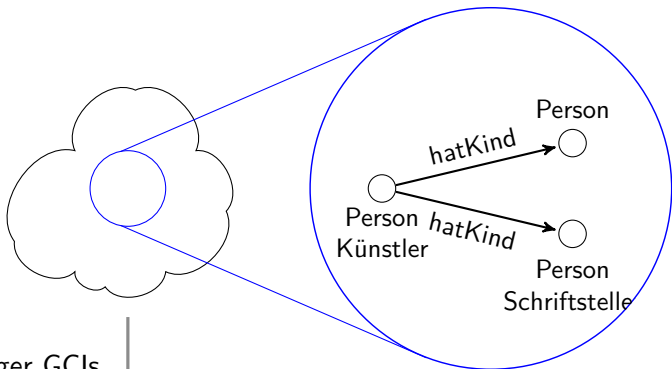
Beobachtung

- ▶ GCIs sind Implikationen sehr ähnlich
- ▶ FBA bietet Möglichkeiten, Implikationen aus Daten zu berechnen

\implies FBA-Methoden verallgemeinern



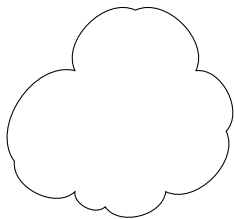




Basis gültiger GCIs



$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$



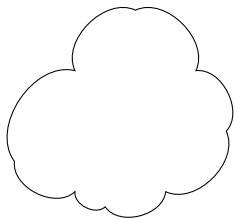
Basis gültiger GCIs



{ \exists hatKind.Schriftsteller \sqsubseteq Künstler, ... }

Beschreibungslogik

Formale Begriffsanalyse



Basis gültiger GCIs

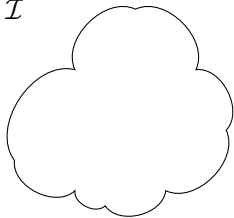


$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$

Beschreibungslogik

Formale Begriffsanalyse

\mathcal{I}



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

Basis gültiger GCIs

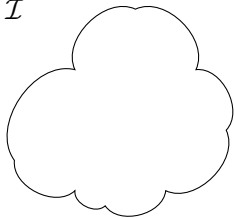


$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$

Beschreibungslogik

Formale Begriffsanalyse

\mathcal{I}



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

Basis gültiger GCIs



$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$

Basis gültiger Implikationen

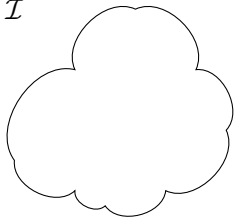


$\{ U \rightarrow V \mid \dots \}$

Beschreibungslogik

Formale Begriffsanalyse

\mathcal{I}



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

Basis gültiger GCIs



$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$

Basis gültiger Implikationen



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



Experiment [Borchmann, Distel 2011]

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

MemberOfParliament \sqsubseteq Person \sqcap Politician

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

MemberOfParliament \sqsubseteq Person \sqcap Politician
 \exists child.Person \sqsubseteq Person

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

MemberOfParliament \sqsubseteq Person \sqcap Politician

\exists child.Person \sqsubseteq Person

FictionalCharacter \sqcap \exists child.Person \sqsubseteq \exists child.FictionalCharacter

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

MemberOfParliament \sqsubseteq Person \sqcap Politician

\exists child.Person \sqsubseteq Person

FictionalCharacter \sqcap \exists child.Person \sqsubseteq \exists child.FictionalCharacter

Fragwürdige Ergebnisse

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

$\text{MemberOfParliament} \sqsubseteq \text{Person} \sqcap \text{Politician}$

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

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

Fragwürdige Ergebnisse

$\text{Person} \sqcap \exists \text{child. Book} \sqsubseteq \text{FictionalCharacter}$

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

$\text{MemberOfParliament} \sqsubseteq \text{Person} \sqcap \text{Politician}$

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

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

Fragwürdige Ergebnisse

$\text{Person} \sqcap \exists \text{child. Book} \sqsubseteq \text{FictionalCharacter}$

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

Experiment [Borchmann, Distel 2011]

- ▶ DBpedia: aus der Wikipedia halbautomatisch gewonnenes Wissen
- ▶ betrachte nur child-Relation $\rightsquigarrow \mathcal{I}_{\text{DBpedia}}$
- ▶ $\Delta^{\mathcal{I}_{\text{DBpedia}}} = 5626$

Einige Ergebnisse

$\text{MemberOfParliament} \sqsubseteq \text{Person} \sqcap \text{Politician}$

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

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

Fragwürdige Ergebnisse

$\text{Person} \sqcap \exists \text{child. Book} \sqsubseteq \text{FictionalCharacter}$

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

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

Beobachtung

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

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$, denn es gibt 4 Gegenbeispiele, d. h.

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$, denn es gibt 4 Gegenbeispiele, d. h.

$$|(\exists \text{child.T})^{\mathcal{I}_{\text{DBpedia}}} \setminus \text{Person}^{\mathcal{I}_{\text{DBpedia}}}| = 4.$$

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$, denn es gibt 4 Gegenbeispiele, d. h.

$$|(\exists \text{child.T})^{\mathcal{I}_{\text{DBpedia}}} \setminus \text{Person}^{\mathcal{I}_{\text{DBpedia}}}| = 4.$$

Andererseits: 2547 Elemente in $\mathcal{I}_{\text{DBpedia}}$ erfüllen $\exists \text{child.T} \sqsubseteq \text{Person}$, d. h.

$$|(\exists \text{child.T} \sqcap \text{Person})^{\mathcal{I}_{\text{DBpedia}}}| = 2547.$$

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$, denn es gibt 4 Gegenbeispiele, d. h.

$$|(\exists \text{child.T})^{\mathcal{I}_{\text{DBpedia}}} \setminus \text{Person}^{\mathcal{I}_{\text{DBpedia}}}| = 4.$$

Andererseits: 2547 Elemente in $\mathcal{I}_{\text{DBpedia}}$ erfüllen $\exists \text{child.T} \sqsubseteq \text{Person}$, d. h.

$$|(\exists \text{child.T} \sqcap \text{Person})^{\mathcal{I}_{\text{DBpedia}}}| = 2547.$$

$\exists \text{child.T} \sqsubseteq \text{Person}$ ist also „fast“ richtig.

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$, denn es gibt 4 **falsche Gegenbeispiele**:

Andererseits: 2547 Elemente in $\mathcal{I}_{\text{DBpedia}}$ erfüllen $\exists \text{child.T} \sqsubseteq \text{Person}$, d. h.

$$|(\exists \text{child.T} \sqcap \text{Person})^{\mathcal{I}_{\text{DBpedia}}}| = 2547.$$

$\exists \text{child.T} \sqsubseteq \text{Person}$ ist also „**fast**“ richtig.

Beobachtung

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

gilt nicht in $\mathcal{I}_{\text{DBpedia}}$, denn es gibt 4 **falsche Gegenbeispiele**:

Teresa_Carpio, Charles_Heung, Adam_Cheng, Lydia_Shum.

Andererseits: 2547 Elemente in $\mathcal{I}_{\text{DBpedia}}$ erfüllen $\exists \text{child.T} \sqsubseteq \text{Person}$, d. h.

$$|(\exists \text{child.T} \sqcap \text{Person})^{\mathcal{I}_{\text{DBpedia}}}| = 2547.$$

$\exists \text{child.T} \sqsubseteq \text{Person}$ ist also **„fast“** richtig.

Intuition

Betrachte auch GCIs, die „fast“ richtig sind.

Intuition

Betrachte auch GCIs, die „fast“ richtig sind.

Definition

Die *Konfidenz* von $C \sqsubseteq D$ in \mathcal{I} ist gegeben durch

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

Intuition

Betrachte auch GCIs, die „fast“ richtig sind.

Definition

Die *Konfidenz* von $C \sqsubseteq D$ in \mathcal{I} ist gegeben durch

$$\text{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1, & \text{falls } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \cap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{sonst.} \end{cases}$$

Sei $c \in [0, 1]$. Dann

$$\text{Th}_c(\mathcal{I}) := \{ C \sqsubseteq D \mid \text{conf}_{\mathcal{I}}(C \sqsubseteq D) \geq c \}.$$

Intuition

Betrachte auch GCIs, die „fast“ richtig sind.

Definition

Die *Konfidenz* von $C \sqsubseteq D$ in \mathcal{I} ist gegeben durch

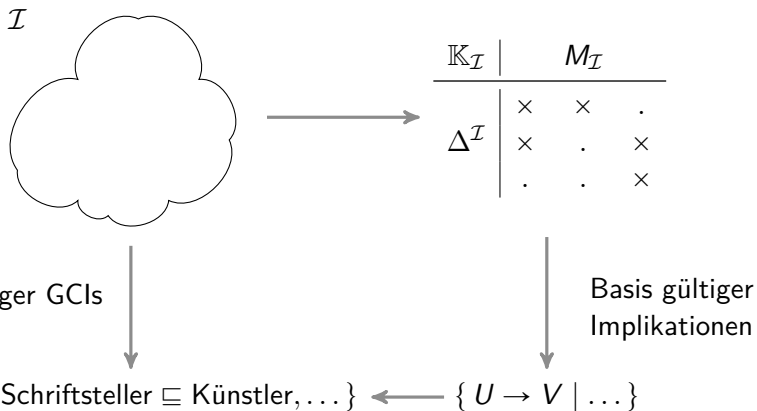
$$\text{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1, & \text{falls } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \cap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{sonst.} \end{cases}$$

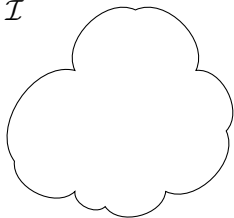
Sei $c \in [0, 1]$. Dann

$$\text{Th}_c(\mathcal{I}) := \{ C \sqsubseteq D \mid \text{conf}_{\mathcal{I}}(C \sqsubseteq D) \geq c \}.$$

Ansatz [Borchmann 2012]

Betrachte $\text{Th}_c(\mathcal{I})$ als „interessantes“ Wissen in \mathcal{I} .



\mathcal{I} 

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

Basis von GCIs
mit hoher Konfidenz



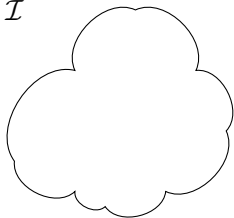
$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$

Basis gültiger
Implikationen



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



\mathcal{I} 

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

Basis von GCIs
mit hoher Konfidenz



$\{ \exists \text{hatKind.Schriftsteller} \sqsubseteq \text{Künstler}, \dots \}$

Basis von
Implikationen mit
hoher Konfidenz



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



Parallelen zwischen FBA und BL

Formale Begriffsanalyse

Beschreibungslogik

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G Merkmale M	Elemente $\Delta^{\mathcal{I}}$ Konzeptbeschreibungen

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\sqcap A)^{\mathcal{I}}$

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\prod A)^{\mathcal{I}}$
$B', B \subseteq G$?

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\prod A)^{\mathcal{I}}$
$B', B \subseteq G$?

Definition [Baader, Distel 2009]

Sei $X \subseteq \Delta^{\mathcal{I}}$.

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\prod A)^{\mathcal{I}}$
$B', B \subseteq G$?

Definition [Baader, Distel 2009]

Sei $X \subseteq \Delta^{\mathcal{I}}$. Eine Konzeptbeschreibung C heißt *model-based most-specific concept description* von X in \mathcal{I} falls

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\prod A)^{\mathcal{I}}$
$B', B \subseteq G$?

Definition [Baader, Distel 2009]

Sei $X \subseteq \Delta^{\mathcal{I}}$. Eine Konzeptbeschreibung C heißt *model-based most-specific concept description* von X in \mathcal{I} falls

- ▶ $X \subseteq C^{\mathcal{I}}$ und

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\prod A)^{\mathcal{I}}$
$B', B \subseteq G$?

Definition [Baader, Distel 2009]

Sei $X \subseteq \Delta^{\mathcal{I}}$. Eine Konzeptbeschreibung C heißt *model-based most-specific concept description* von X in \mathcal{I} falls

- ▶ $X \subseteq C^{\mathcal{I}}$ und
- ▶ für jedes D mit $X \subseteq D^{\mathcal{I}}$ ist C *spezieller als* D .

Parallelen zwischen FBA und BL

Formale Begriffsanalyse	Beschreibungslogik
Gegenstände G	Elemente $\Delta^{\mathcal{I}}$
Merkmale M	Konzeptbeschreibungen
Formale Kontexte \mathbb{K}	Interpretationen \mathcal{I}
Implikationen	GCI's
$A', A \subseteq M$	$(\prod A)^{\mathcal{I}}$
$B', B \subseteq G$?

Definition [Baader, Distel 2009]

Sei $X \subseteq \Delta^{\mathcal{I}}$. Eine Konzeptbeschreibung C heißt *model-based most-specific concept description* von X in \mathcal{I} falls

- ▶ $X \subseteq C^{\mathcal{I}}$ und
- ▶ für jedes D mit $X \subseteq D^{\mathcal{I}}$ ist C *spezieller als* D .

Bezeichnung: $C = X^{\mathcal{I}}$

Definition

$$M_I = N_C \cup \{\perp\} \cup \{\exists r.X^I \mid X \subseteq \Delta^I, X \neq \emptyset\}.$$

Definition

$$M_I = N_C \cup \{\perp\} \cup \{\exists r.X^I \mid X \subseteq \Delta^I, X \neq \emptyset\}.$$

Definition

Setze $\mathbb{K}_I = (\Delta^I, M_I, \nabla)$ mit

$$(x, C) \in \nabla \iff x \in C^I.$$

Definition

$$M_{\mathcal{I}} = N_C \cup \{\perp\} \cup \{\exists r.X^{\mathcal{I}} \mid X \subseteq \Delta^{\mathcal{I}}, X \neq \emptyset\}.$$

Definition

Setze $\mathbb{K}_{\mathcal{I}} = (\Delta^{\mathcal{I}}, M_{\mathcal{I}}, \nabla)$ mit

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

Satz [Baader, Distel 2009]

Ist \mathcal{L} eine Basis von $\mathbb{K}_{\mathcal{I}}$, dann ist

$$\{\bigcap U \subseteq (\bigcap U)^{\mathcal{I}\mathcal{I}} \mid (U \rightarrow V) \in \mathcal{L}\}$$

eine Basis von \mathcal{I} .

Problem

Wie Basen von $\text{Th}_c(\mathcal{I})$ finden?

Problem

Wie Basen von $\text{Th}_c(\mathcal{I})$ finden?

Ansatz [Luxenburger 1993, Borchmann 2012]

Problem

Wie Basen von $\text{Th}_c(\mathcal{I})$ finden?

Ansatz [Luxenburger 1993, Borchmann 2012]

- ▶ Finde Basen von $\text{Th}_c(\mathcal{I}) \setminus \text{Th}(\mathcal{I})$.

Problem

Wie Basen von $\text{Th}_c(\mathcal{I})$ finden?

Ansatz [Luxemburger 1993, Borchmann 2012]

- ▶ Finde Basen von $\text{Th}_c(\mathcal{I}) \setminus \text{Th}(\mathcal{I})$.
- ▶ Betrachte nur Konzeptbeschreibungen der Form $X^{\mathcal{I}}$ für $X \subseteq \Delta^{\mathcal{I}}$.

Problem

Wie Basen von $\text{Th}_c(\mathcal{I})$ finden?

Ansatz [Luxenburger 1993, Borchmann 2012]

- ▶ Finde Basen von $\text{Th}_c(\mathcal{I}) \setminus \text{Th}(\mathcal{I})$.
- ▶ Betrachte nur Konzeptbeschreibungen der Form $X^{\mathcal{I}}$ für $X \subseteq \Delta^{\mathcal{I}}$.

Definition

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

Problem

Wie Basen von $\text{Th}_c(\mathcal{I})$ finden?

Ansatz [Luxenburger 1993, Borchmann 2012]

- ▶ Finde Basen von $\text{Th}_c(\mathcal{I}) \setminus \text{Th}(\mathcal{I})$.
- ▶ Betrachte nur Konzeptbeschreibungen der Form $X^{\mathcal{I}}$ für $X \subseteq \Delta^{\mathcal{I}}$.

Definition

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

Satz [Borchmann 2012]

Es sei \mathcal{B} eine Basis von \mathcal{I} , und $c \in [0, 1]$. Dann ist

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

eine Basis $\text{Th}_c(\mathcal{I})$.

Definition

$$\text{conf}_{\mathbb{K}}(X \rightarrow Y) := \begin{cases} 1 & X' = \emptyset, \\ \frac{|(X \cup Y)'|}{|X'|} & \text{sonst.} \end{cases}$$

Definition

$$\text{conf}_{\mathbb{K}}(X \rightarrow Y) := \begin{cases} 1 & X' = \emptyset, \\ \frac{|(X \cup Y)'|}{|X'|} & \text{sonst.} \end{cases}$$

Für $c \in [0, 1]$ ist

$$\text{Th}_c(\mathbb{K}) := \{ X \rightarrow Y \mid \text{conf}_{\mathbb{K}}(X \rightarrow Y) \geq c \}.$$

Definition

$$\text{conf}_{\mathbb{K}}(X \rightarrow Y) := \begin{cases} 1 & X' = \emptyset, \\ \frac{|(X \cup Y)'|}{|X'|} & \text{sonst.} \end{cases}$$

Für $c \in [0, 1]$ ist

$$\text{Th}_c(\mathbb{K}) := \{X \rightarrow Y \mid \text{conf}_{\mathbb{K}}(X \rightarrow Y) \geq c\}.$$

Satz [Borchmann 2012]

Ist \mathcal{L} eine Basis von $\text{Th}_c(\mathbb{K}_{\mathcal{I}})$, so ist

$$\bigcap \mathcal{L} := \{ \bigcap U \sqsubseteq \bigcap V \mid (U \rightarrow V) \in \mathcal{L} \}$$

eine Basis von $\text{Th}_c(\mathcal{I})$.

Definition

$$\text{conf}_{\mathbb{K}}(X \rightarrow Y) := \begin{cases} 1 & X' = \emptyset, \\ \frac{|(X \cup Y)'|}{|X'|} & \text{sonst.} \end{cases}$$

Für $c \in [0, 1]$ ist

$$\text{Th}_c(\mathbb{K}) := \{X \rightarrow Y \mid \text{conf}_{\mathbb{K}}(X \rightarrow Y) \geq c\}.$$

Satz [Borchmann 2012]

Ist \mathcal{L} eine Basis von $\text{Th}_c(\mathbb{K}_{\mathcal{I}})$, so ist

$$\bigcap \mathcal{L} := \{ \bigcap U \sqsubseteq \bigcap V \mid (U \rightarrow V) \in \mathcal{L} \}$$

eine Basis von $\text{Th}_c(\mathcal{I})$.

Ist $\mathcal{L} \subseteq \text{Th}_c(\mathbb{K}_{\mathcal{I}})$, so ist auch $\bigcap \mathcal{L} \subseteq \text{Th}_c(\mathcal{I})$.

Experiment

Experiment

$$\text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) =$$

Experiment

$$\begin{aligned} \text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ & \exists \text{child}.\top \sqsubseteq \text{Person}, \\ & \text{Place} \sqsubseteq \text{PopulatedPlace}, \\ & \exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder} \\ & \sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \} \end{aligned}$$

Experiment

$$\begin{aligned} \text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ & \exists \text{child}.\top \sqsubseteq \text{Person}, \quad \checkmark \\ & \text{Place} \sqsubseteq \text{PopulatedPlace}, \\ & \exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder} \\ & \sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \} \end{aligned}$$

Experiment

$$\text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ \exists \text{child}.\top \sqsubseteq \text{Person}, \quad \checkmark \\ \text{Place} \sqsubseteq \text{PopulatedPlace}, \\ \exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder} \\ \sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \}$$

Gegenbeispiel zu zweiter GCI

Greenwich_Village

Experiment

$$\text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ \exists \text{child}.\top \sqsubseteq \text{Person}, \quad \checkmark$$
$$\text{Place} \sqsubseteq \text{PopulatedPlace}, \quad \checkmark$$
$$\exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder}$$
$$\sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \}$$

Gegenbeispiel zu zweiter GCI

Greenwich_Village

Experiment

$$\begin{aligned} \text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ & \exists \text{child}.\top \sqsubseteq \text{Person}, && \checkmark \\ & \text{Place} \sqsubseteq \text{PopulatedPlace}, && \checkmark \\ & \exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder} \\ & \sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \} \end{aligned}$$

Experiment

$$\begin{aligned} \text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ & \exists \text{child}.\top \sqsubseteq \text{Person}, \quad \checkmark \\ & \text{Place} \sqsubseteq \text{PopulatedPlace}, \quad \checkmark \\ & \exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder} \\ & \sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \} \end{aligned}$$

Gegenbeispiel zu dritter GCI

Pierre_Samuel_du_Pont_de_Nemours

Experiment

$$\text{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95) = \{ \begin{array}{l} \exists \text{child}.\top \sqsubseteq \text{Person}, \quad \checkmark \\ \text{Place} \sqsubseteq \text{PopulatedPlace}, \quad \checkmark \\ \exists \text{child}.\exists \text{child}.\top \sqcap \exists \text{child}.\text{OfficeHolder} \quad \times \\ \sqsubseteq \exists \text{child}.\text{(OfficeHolder} \sqcap \exists \text{child}.\top) \end{array} \}$$

Gegenbeispiel zu dritter GCI

Pierre_Samuel_du_Pont_de_Nemours

Problem

Ansatz kann *fehlerhafte Daten* von *seltenen Gegenbeispielen* nicht unterscheiden.

Problem

Ansatz kann *fehlerhafte Daten* von *seltenen Gegenbeispielen* nicht unterscheiden.

Beispiel

Die Aussage *Alle Säugetiere gebären lebend*,

Säugetier \sqsubseteq LebendGebärend,

Problem

Ansatz kann *fehlerhafte Daten* von *seltenen Gegenbeispielen* nicht unterscheiden.

Beispiel

Die Aussage *Alle Säugetiere gebären lebend*,

Säugetier \sqsubseteq LebendGebärend,

hat Gegenbeispiel *Schnabeltier*.



<http://www.sealifeconservation.org.au/adopt-an-animal/>

Problem

Ansatz kann *fehlerhafte Daten* von *seltenen Gegenbeispielen* nicht unterscheiden.

Beispiel

Die Aussage *Alle Säugetiere gebären lebend*,

Säugetier \sqsubseteq LebendGebärend,

hat Gegenbeispiel *Schnabeltier*.



<http://www.sealifeconservation.org.au/adopt-an-animal/>

Ausweg

Verwendung externer *Experten*

Problem

Ansatz kann *fehlerhafte Daten* von *seltenen Gegenbeispielen* nicht unterscheiden.

Beispiel

Die Aussage *Alle Säugetiere gebären lebend*,

$\text{Säugetier} \sqsubseteq \text{LebendGebärend}$,

hat Gegenbeispiel *Schnabeltier*.



<http://www.sealifeconservation.org.au/adopt-an-animal/>

Ausweg

Verwendung externer *Experten* \implies Merkmalexploration mit Implikationen/GCIs mit hoher Konfidenz [Borchmann 2013a, Borchmann 2013b]

Zusammenfassung

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*
- ▶ Experimentelle Auswertung dieses Ansatzes

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*
- ▶ Experimentelle Auswertung dieses Ansatzes
- ▶ Ausarbeitung einer Merkmalexploration von Implikationen/GCIs mit hoher Konfidenz

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*
- ▶ Experimentelle Auswertung dieses Ansatzes
- ▶ Ausarbeitung einer Merkmalexploration von Implikationen/GCIs mit hoher Konfidenz

Ausblick

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*
- ▶ Experimentelle Auswertung dieses Ansatzes
- ▶ Ausarbeitung einer Merkmalexploration von Implikationen/GCIs mit hoher Konfidenz

Ausblick

- ▶ Anwendung auf praktische Probleme (medizinische Daten)

Zusammenfassung

- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*
- ▶ Experimentelle Auswertung dieses Ansatzes
- ▶ Ausarbeitung einer Merkmalexploration von Implikationen/GCIs mit hoher Konfidenz

Ausblick

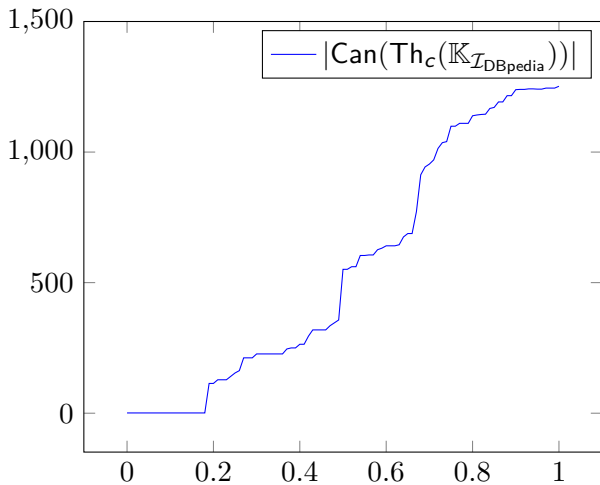
- ▶ Anwendung auf praktische Probleme (medizinische Daten)
- ▶ Adaption und Optimierung der entwickelten Algorithmen

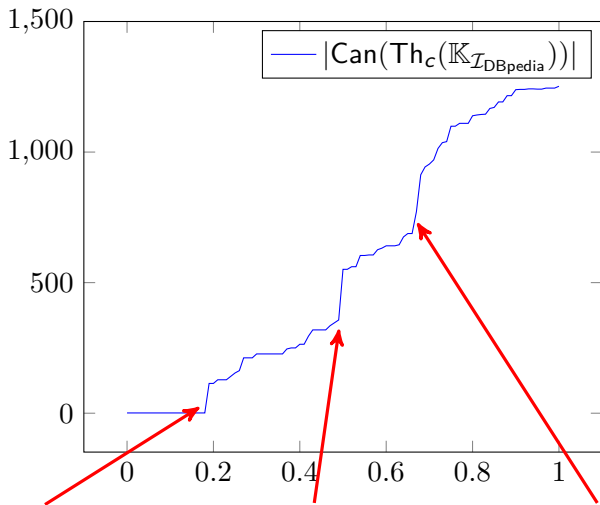
Zusammenfassung

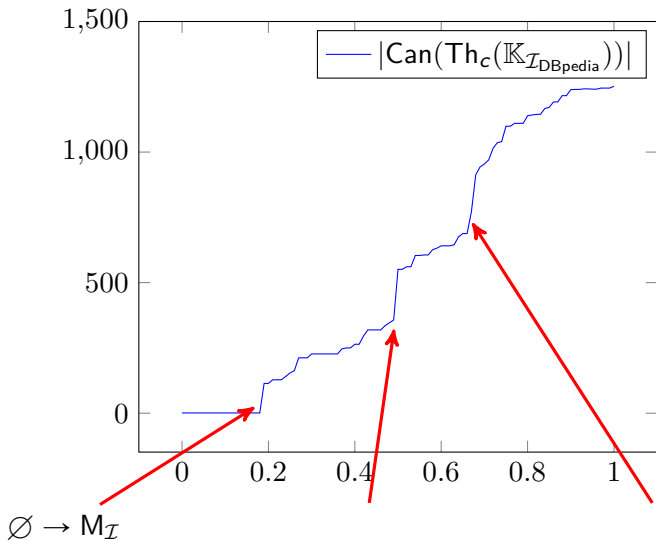
- ▶ Experimentelle Untersuchung der Resultate von Baader und Distel
- ▶ Erweiterung der Resultate um den Begriff der *Konfidenz*
- ▶ Experimentelle Auswertung dieses Ansatzes
- ▶ Ausarbeitung einer Merkmalexploration von Implikationen/GCIs mit hoher Konfidenz

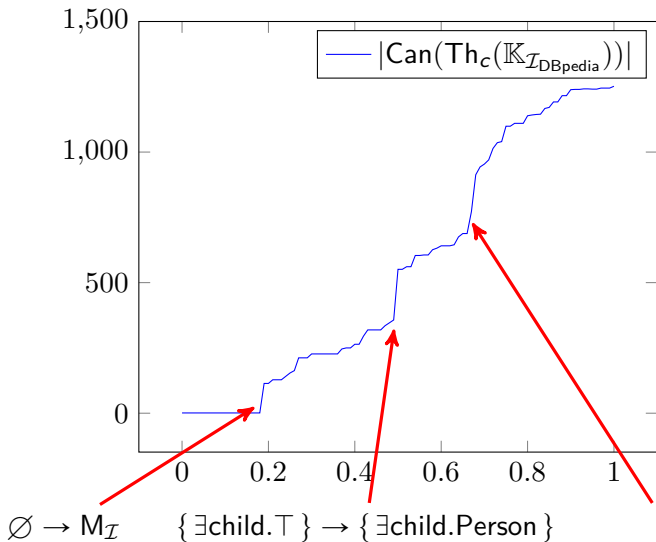
Ausblick

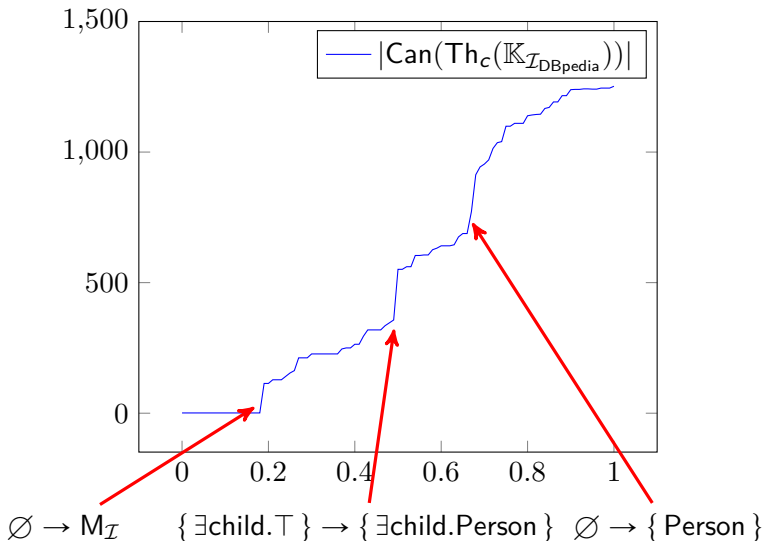
- ▶ Anwendung auf praktische Probleme (medizinische Daten)
- ▶ Adaption und Optimierung der entwickelten Algorithmen
- ▶ Beschränkte Rollentiefe [Distel 2012]











Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

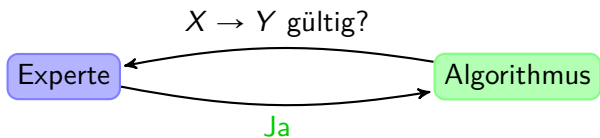
$$\mathcal{S} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

$X \rightarrow Y$ gültig?

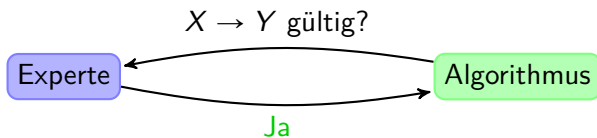
Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1,$
g_1		\dots		\dots
\vdots				$A_\ell \rightarrow B_\ell \}$
g_k		\dots		



\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1,$
g_1		\dots		\dots
\vdots				$A_\ell \rightarrow B_\ell \}$
g_k		\dots		



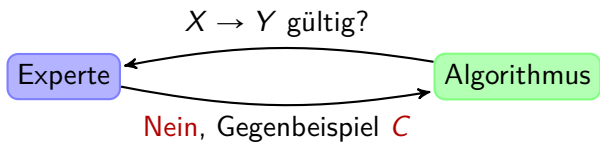
\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1,$
g_1		\dots		\dots
\vdots				$A_\ell \rightarrow B_\ell,$
g_k		\dots		$X \rightarrow Y \}$

$X \rightarrow Y$ gültig?

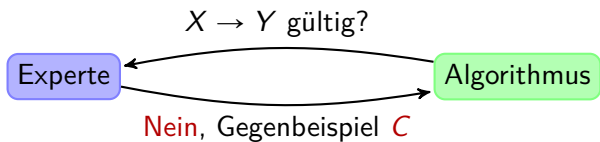
Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1,$
g_1		\dots		\dots
\vdots				$A_\ell \rightarrow B_\ell \}$
g_k		\dots		



\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1,$
g_1		\dots		\dots
\vdots				$A_\ell \rightarrow B_\ell \}$
g_k		\dots		



\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1,$
g_1		\dots		\dots
\vdots				$A_\ell \rightarrow B_\ell \}$
g_k		\dots		
g_{k+1}		C		

Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{S} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n
g_1		\dots	
\vdots			
g_k		\dots	

$$\mathcal{S} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$$

Bemerkung

Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n	$\mathcal{S} = \{ A_1 \rightarrow B_1, \\ \dots \\ A_\ell \rightarrow B_\ell \}$
g_1		\dots		
\vdots				
g_k		\dots		

Bemerkung

- ▶ Nach Abschluss der Exploration gilt $\text{Cn}(\mathcal{S}) = \text{Th}(\mathbb{K})$ und \mathcal{S} ist eine Basis der vom Experten bestätigten Implikationen.

Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n	$\mathcal{S} = \{ A_1 \rightarrow B_1, \dots, A_\ell \rightarrow B_\ell \}$
g_1		\dots		
\vdots				
g_k		\dots		

Bemerkung

- ▶ Nach Abschluss der Exploration gilt $\text{Cn}(\mathcal{S}) = \text{Th}(\mathbb{K})$ und \mathcal{S} ist eine Basis der vom Experten bestätigten Implikationen.
- ▶ Die Anzahl der *bestätigten* Implikationen ist kleinst-möglich.

Experte

Algorithmus

\mathbb{K}	m_1	\dots	m_n	$S = \{ A_1 \rightarrow B_1, \dots, A_\ell \rightarrow B_\ell \}$
g_1		\dots		
\vdots				
g_k		\dots		

Bemerkung

- ▶ Nach Abschluss der Exploration gilt $Cn(S) = Th(\mathbb{K})$ und S ist eine Basis der vom Experten bestätigten Implikationen.
- ▶ Die Anzahl der *bestätigten* Implikationen ist kleinst-möglich.

Resultat [Distel 2008]

Merkmalexploration für gültige GCIs einer endlichen Interpretation möglich

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ursache

Operator conf nicht monoton:

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ursache

Operator conf nicht monoton:

$$\text{conf}(A \rightarrow B) < c$$

$$\text{conf}(C \rightarrow B) \geq c$$

für $C \subseteq A$ möglich.

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

- ▶ Mehr Implikationen abfragen

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

- ▶ Mehr Implikationen abfragen

$$A \rightarrow A'' \quad (\text{für abgeschlossene } A)$$

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

- ▶ Mehr Implikationen abfragen

$$A \rightarrow A'' \quad (\text{für abgeschlossene } A)$$
$$B'' \rightarrow \{m\} \quad (\text{falls } \text{conf}(B'' \rightarrow \{m\}) \geq c)$$

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

- ▶ Mehr Implikationen abfragen

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

- ▶ Mehr Implikationen abfragen
- ▶ Bei GCIs: mehr Fragen an den Experten

Ziel

Exploration auch für GCIs/Implikationen mit hoher Konfidenz

Problem

Erzeugung von Implikationen schwieriger

Ausweg

- ▶ Mehr Implikationen abfragen
- ▶ Bei GCIs: mehr Fragen an den Experten

Resultat [Borchmann 2013a, Borchmann 2013b]

Merkmalexploration auch mit Implikationen/GCIs mit hoher Konfidenz möglich