Semantic Integration
–
Ontologiedesign und -repräsentation

Jan Ortmann

11. Mai 2006

Gliederung

Ontologien – Begriffsbestimmung

Eine Definition

„[...][E]ine Ontologie ist ein Verzeichnis der Typen von den Dingen, von denen angenommen wird, dass sie in einer Domäne von Interesse $\mathfrak{D}$ aus der Perspektive einer Person existieren, die die Sprache $\mathfrak{L}$ benutzt, um über $\mathfrak{D}$ zu reden. Die Typen der Ontologie stellen die Prädikate, die Wortbedeutungen oder die Konzept- und Beziehungstypen der Sprache $\mathfrak{L}$ dar, wenn sie dazu benutzt wird, über die Domäne $\mathfrak{D}$ zu reden. [...] Eine formale Ontologie wird durch eine Reihe von Namen für Konzept- und Beziehungstypen spezifiziert, die durch die Typ-Subtyp-Relation in einer partiellen Ordnung organisiert sind" Sowa (2000) (S. 492).

Ontologien – Repräsentation

Bestandteile

Definition (Ontologie)

Eine Ontologie ist ein Tupel $\mathfrak{Ont} =_{def} (\mathfrak{O}, \mathfrak{I}, \mathfrak{L}_{\mathfrak{O}}, \mathfrak{L}_{\mathfrak{I}})$, bestehend aus

- $\mathfrak{O}$ einer terminologische Struktur,
- $\mathfrak{I}$ einer assertorische Struktur,
- $\mathfrak{L}_{\mathfrak{O}}$ einem terminologischen Lexikon und
- $\mathfrak{L}_{\mathfrak{I}}$ einem assertorischen Lexikon.

Ontologien – Repräsentation

Terminologisches Wissen

Definition (Terminologische Struktur)

Eine *terminologische Struktur* ist ein Tupel $\mathfrak{T} =_{def} (\mathfrak{C}, \mathfrak{R}, \preceq_{\mathfrak{C}}, \preceq_{\mathfrak{R}}, \mathfrak{ar}, \mathfrak{Ax}, \top)$, wobei

- $\mathfrak{C}$ eine Menge von Konzeptsymbolen ist,
- $\mathfrak{R}$ eine Menge von Relationssymbolen genannt *Rollen* ist,
- $\preceq_{\mathfrak{C}} \subseteq \mathfrak{C} \times \mathfrak{C}$ eine partielle Ordnung auf $\mathfrak{C}$ ist, die *Konzepthierarchie* genannt wird,
- $\preceq_{\mathfrak{R}} \subseteq \mathfrak{R} \times \mathfrak{R}$ eine partielle Ordnung auf $\mathfrak{R}$ ist, die *Rollenhierarchie* genannt wird,
- $\mathfrak{ar} : \mathfrak{R} \rightarrow \mathfrak{C} \times \mathfrak{C}$ die Arität (Stelligkeit) der Rollen angibt,
- $\mathfrak{Ax}$ eine Menge von logischen Axiomen in einer beliebigen Logik ist und
- $\top$ das Wurzelkonzeptsymbol der Ontologie ist, so dass $\forall \mathfrak{c} \in \mathfrak{C}. \top \preceq^*_{\mathfrak{C}} \mathfrak{c}$ gilt.

Ontologien – Repräsentation

Assertorisches Wissen

Definition (Assertorische Struktur, Gültigkeit)

Eine *assertorische Struktur* zu einer terminologischen Struktur $\mathfrak{T}$ ist ein Tupel $\mathfrak{A}_{\mathfrak{T}} =_{def} (\mathfrak{Ci}, \cdot^{\mathfrak{I}})$, wobei

- $\mathfrak{Ci}$ eine Menge von Konzeptinstanzsymbolen ist und
- $\cdot^{\mathfrak{I}}$ eine Abbildung ist, die
  - $\cdot^{\mathfrak{I}}|_{\mathfrak{C}} : \mathfrak{C} \to 2^{\mathfrak{Ci}}$ jedem Konzept eine Menge von Instanzen zuweist, so dass, $\forall \mathfrak{c}_1, \mathfrak{c}_2 \in \mathfrak{C}. \mathfrak{c}_1 \leq_{\mathfrak{C}} \mathfrak{c}_2 \implies \mathfrak{c}_1{}^{\mathfrak{I}} \subseteq \mathfrak{c}_2{}^{\mathfrak{I}}$,
  - $\cdot^{\mathfrak{I}}|_{\mathfrak{R}} : \mathfrak{R} \to 2^{\mathfrak{Ci} \times \mathfrak{Ci}}$ jeder Rolle eine Menge von binären Relationen über den Instanzen zuweist, so dass $\forall \mathfrak{r}_1, \mathfrak{r}_2 \in \mathfrak{R}. \mathfrak{r}_1 \leq_{\mathfrak{R}} \mathfrak{r}_2 \implies \mathfrak{r}_1{}^{\mathfrak{I}} \subseteq \mathfrak{r}_2{}^{\mathfrak{I}}$ und für die
  - $\top^{\mathfrak{I}} = \mathfrak{Ci}$ ist.

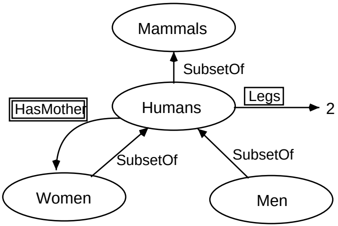
Eine assertorische Struktur $\mathfrak{A}$ *genügt* einer terminologischen Struktur $\mathfrak{T}$, wenn sie der Menge der Axiome $\mathfrak{Ax}$ in $\mathfrak{T}$ genügt. Die Menge aller assertorischen Strukturen zu einer taxonomischen Struktur $\mathfrak{T}$ wird als $\mathfrak{Ass}(\mathfrak{T})$ notiert.

Formalismen
Übersicht

- Semantische Netze
- Konzeptrahmen (Frames) /Frame-Logik
- Konzeptgraphen

Formalismen

Semantische Netzwerke

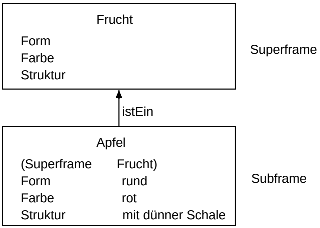


- Graphen mit gerichteten beschrifteten Kanten
- Knoten stellen sprachliche Einheiten dar, Kanten Relationen
- Sub-/Superklassenkanten und Eigenschaftskanten
- Vorgänger der meisten anderen Formalismen

Formalismen

Frames (Konzeptrahmen)

- Objektzentrierter Ansatz
- Frame gibt ein allgemeines Bild eine Konzeptes
- Slots eines Frames spezifizieren seines Eigenschaften
- Subframes eines Frames konkretisieren die Eigenschaften.
- Zunächst keine Instanzen, nur Subklassen



Formalismen

Frames (Verwendung in OKBC)

- Verwendung in der OKBC Schnittstelle
- OKBC unterscheidet Wissensbasen, Klassen, Individuen, Slots und Facets. Hinzu kommen eine Reihe von Konstanten primitiver Datentypen wie Zeichenketten und Zahlen.
- Frames repräsentieren Individuen und Klassen und besitzen eine Menge von Slots, ähnlich den Attributen eines Objektes.
- Facets stellen Bedingungen dar, die für die Slots eines Frames eingehalten werden müssen.
- Neben dem Modell stellt die OKBC auch eine Menge von Operationen auf Wissensbasen zur Verfügung, wie z.B. getOwnSlotValue und setOwnSlotValue zum Zugriff auf den Inhalt der Slots eines Frames.

Formalismen

Frame-Logik (F-Logik)

- Erweiterung der Prädikatenlogik um objektorientierte Konzepte, daher unentscheidbare Probleme
- Basis von Ontobroker, Flora und WSML
- Frame-Logik bedient sich Methodenausdrücken, die bei bestimmten Parametern bestimmte Werte zurückgeben.
- Signaturmethoden erlauben die Definition von Signaturen. Sie geben die Rückgabewerte einer Methode an.
- Besteht aus Konstantensymbolen, Funktionssymbolen und Variablensymbolen sowie den Symbolen $\rightarrow, \twoheadrightarrow, \bullet\!\!\rightarrow, \bullet\!\!\twoheadrightarrow, \Rightarrow, \Rrightarrow$, : und :: und den logischen Junktoren und Quantoren.

Formalismen

Frame-Logik (Forts.)

ID-Terme werden aus den Konstanten-, Funktions- und Variablensymbolen in der üblichen Art gebildet. Die einfachste Art von Formeln werden molekulare F-Formeln (F-Moleküle) genannt und sind wie folgt aufgebaut:

- $C :: D$ oder $o : C$, wobei C, D und o ID-Terme sind. Ersteres bezeichnet die Subtypbeziehung währende letzteres die Instanzbeziehung darstellt.
- ein Objektmolekül der Form o[eine ';'-separierte Liste von Methodenausdrücken], wobei o ein ID-Term ist, der ein Objekt bezeichnet. Ein Methodenausdruck ist entweder ein nicht-vererbbarer Datenausdruck, ein vererbbarer Datenausdruck oder ein Signaturausdruck.

Formalismen

Frame-Logik (Forts.)

Datenausdrücke:

- ein nicht-vererbbarer Datenausdruck hat eine der beiden Formen
  - ein nicht-vererbbarer Skalarausdruck ($k \geq 0$): $ScalarMethod@Q_1, \dots, Q_k \rightarrow T$;
  - ein nicht-vererbbarer Mengenausdruck ($l, m \geq 0$): $SetMethod@R_1, \dots, R_l \twoheadrightarrow \{S_1, \dots, S_m\}$.
- vererbbare Skalarausdrücke und Mengenausdrücke sind wie nicht-vererbbare Ausdrücke, nur dass $\rightarrow$ gegen $\bullet\!\!\rightarrow$ ausgetauscht wird und $\twoheadrightarrow$ gegen $\bullet\!\!\twoheadrightarrow$,

Signaturausdrücke haben ebenfalls zwei Formen

- ein Skalarsignaturausdruck ($n, r \geq 0$): $ScalarMethod@V_1, \dots, V_n \Rightarrow (A_1, \dots, A_r)$;
- ein Mengensignaturausdruck ($s, t \geq 0$): $SetMethod@W_1, \dots, W_s \Rrightarrow (B_1, \dots, B_t)$.

Formalismen

Frame-Logik (Beispiel)

```
woman::person
man::person
person[father=>man].
person[mother=>woman].
person[daugther=>woman].
person[son=>man].

FORALL X,Y X[son->>Y] <- Y:man[father->X].
FORALL X,Y X[son->>Y] <- Y:man[mother->X].
FORALL X,Y X[daughter->>Y] <- Y:woman[father->X].
FORALL X,Y X[daughter->>Y] <- Y:woman[mother->X].

abraham:man.
sarah:woman.
isaac:man[father->abraham; mother->sarah].
ishmael:man[father->abraham; mother->hagar:woman].
jacob:man[father->isaac; mother->rebekah:woman].
esau:man[father->isaac; mother->rebekah].
```

Formalismen

Frame-Logik (Beispiel Forts.)

- Schlussfolgerungsmechanismen erlauben Anfragen
- Sämtliche ID-Terme können durch Variablen dargestellt werden

Dies ermöglicht Anfragen der folgenden Art:

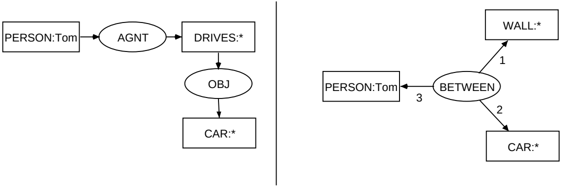
```
FORALL X,Y <- X:woman[son->>Y[father->abraham]].

FORALL X,Y <- isaac[X->>Y].
```

Formalismen

Konzeptgraphen

- Bipartiter Graph deren Knotenmenge sich in Konzeptknoten und Konzeptbeziehungsknoten aufteilt
- Erlaubt volle Mächtigkeit der Prädikatenlogik, Unentscheidbarkeit von Problemen
- Formalisierung der Beziehungen durch Prädikatenlogik
- Schlussfolgerungsmechanismen entscheiden, ob ein Graph einen anderen subsumiert, d.h. die entsprechenden Formeln sich implizieren, oder ob ein Graph gültig ist, d.h. die entsprechende Formel gültig ist.
- generischen Konzeptknoten sind mit generischem Marker belegt, individuellen Konzeptknoten sind mit Objekt des Konzepttyps belegt.
- Getypte Knoten und Subtypbeziehung.

Formalismen
Konzeptgraphen (Beispiel)



Formalismen

Beschreibungslogiken vs. Frame-Logik

- Beschreibungslogiken lassen sich in der vollen Variante von Frame-Logik darstellen
- Volle Variante nicht implementiert, daher nicht alles darstellbar.
- Horn-Fragment von DL in Horn-Fragment von Frame-Logik darstellbar.
- Frame-Logik ist berechnungsvollständig (Turing-mächtig)
- Beschreibungslogiken erlauben die Darstellung von Existenz und Disjunktion. Ersteres erlaubt das intensionale Schließen.
- die explizite Unterscheidung zwischen Instanzen und Konzepten ist in Frame-Logik nur durch die explizite Einführung von Sorten möglich.

Formalismen

Beschreibungslogiken vs. Frame-Logik (Forts.)

Unterschied zwischen Frame-Logik und Beschreibungslogik:

$$\mathrm{Bebaubar} \equiv \mathrm{Feld} \sqcap \exists \mathrm{nachbarFeld}.\mathrm{EigenesFeld}$$

Das Hinzufügen von $\mathrm{nachbarFeld}(\mathrm{feld1}, \mathrm{eigenesFeld2})$:

- Beschreibungslogiken: $\mathrm{feld1}$ bekommt Typ $\mathrm{Bebaubar}$
- Frame-Logik: Ausgabe eine Typfehlers.

Ontologiebegriff

- Lehre vom Sein
- „Essenz des Seienden“
- Abgrenzung zur Erkenntnistheorie
- Wesentliche Kategorien:
    - Das vom Menschen Erschaffene
    - Das a priori Existente (ggf. mit Bedeutung belegt)
    - Das Soziale / Sozialwahrnehmende

Moderne Ontologie

- Analytische Ontologie
  Vollständige Konzeptualisierung
- „beschreibende Ontologie“ (Smith, 2003)
  Teilbereiche unterschiedlicherer Granularität
- Pragmatische Verwendung in der Informatik

Erkenntnistheorie

Kritischer Rationalismus

- Allgemeinen Konventionen prägen DesignNiemann (2004)
- Aussage: „Alle Schwäne sind weiß“
- Falsifizierbar / Nicht beweisbar
- Würde ein schwarzer Schwan die Aussage falsifizieren?

Erkenntnistheorie

Kritischer Rationalismus

- Sozialer und kultureller Hintergrund wichtig
- Konzepte stellen Konventionen dar
- Bsp.: *ÄlterePerson*

Ontologiedesign

Meta-Eigenschaften

- Unterstützung eines guten Designs
- Unterscheidung verschiedener Konzeptarten

Ontologiedesign
Meta-Eigenschaften – Identität

Definition (Identitätsbedingung)

Eine Identitätsbedingung IC ist eine Identitätsformel (sameness formula) Σ, die mit Ausnahme trivialer Fälle entweder (1) oder (2) erfüllt, wobei $E(x,t)$ das Prädikat der tatsächlichen Existenz von x zum Zeitpunkt t darstellt:

$$\Box(E(x,t) \land \phi(x,t) \land E(y,t') \land \phi(y,t') \land x = y \implies \Sigma(x,y,t,t')) \quad (1)$$

$$\Box(E(x,t) \land \phi(x,t) \land E(y,t') \land \phi(y,t') \land \Sigma(x,y,t,t') \implies x = y) \quad (2)$$

Ontologiedesign

Meta-Eigenschaften – Rigidität

Definition (essentiell, rigide, arigide, antirigide)

Eine Eigenschaft eines Objektes ist nach Welty und Guarino (2001) *essentiell*, wenn sie notwenig für dieses Objekt ist, d.h. das Objekt besitzt diese Eigenschaft immer und in jeder möglichen Welt.

Eine Eigenschaft ϕ ist *rigide*, wenn sie essentiell für alle ihre Instanzen ist, d.h. wenn
$\Box(\forall x, t.\phi(x, t) \implies \Box\forall t'\phi(x, t'))$.

Eine Eigenschaft ϕ ist *arigide*, wenn sie nicht essentiell für einige ihrer Instanzen ist, d.h. wenn $\Diamond(\exists x, t.\phi(x, t) \wedge \Diamond\exists t'\neg\phi(x, t'))$.

Eine Eigenschaft ϕ ist *antirigide*, wenn sie nicht essentiell für alle ihre Instanzen ist, d.h. wenn $\Box(\forall x, t.\phi(x, t) \implies \Diamond\exists t'\neg\phi(x, t'))$.

Ontologiedesign

Meta-Eigenschaften – Abhängigkeit

Definition (abhängig)

Eine Eigenschaft ϕ ist *extern abhängig* von einer Eigenschaft φ, wenn für jede ihrer Instanzen x eine Instanz von φ existieren muss, die nicht Bestandteil von x ist.

+O	**+I**	**+R**	**+D** **−D**	Type	Sortal
−O	**+I**	**+R**	**+D** **−D**	Quasi-Type	Sortal
−O	**+I**	**∼ R**	**+D**	Material Row	Sortal
−O	**+I**	**∼ R**	**−D**	Phased Sortal	Sortal
−O	**+I**	**¬R**	**+D** **−D**	Mixin	Sortal
−O	**−I**	**+R**	**+D** **−D**	Category	Non-Sortal
−O	**−I**	**∼ R**	**+D**	Formal Role	Non-Sortal
−O	**−I**	**∼ R**	**−D**	Attribution	Non-Sortal
−O	**−I**	**¬R**	**+D** **−D**	Attribution	
+O	**−I**	*	*	Incoherent	
+O	**+I**	**∼ R** **¬R**	* *	Incoherent	

Ontologiedesign

Meta-Eigenschaften – Regeln

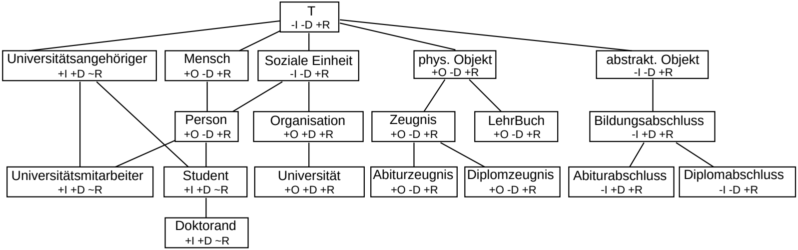
Laut Welty und Guarino (2001) gelten nun die folgenden Bedingungen:

$$\phi^{\sim\mathbf{R}} \text{ muss } \phi^{\sim\mathbf{R}} \text{ subsumieren} \tag{3}$$

$$\phi^{+\mathbf{I}} \text{ muss } \phi^{+\mathbf{I}} \text{ subsumieren} \tag{4}$$

$$\phi^{+\mathbf{D}} \text{ muss } \phi^{+\mathbf{D}} \text{ subsumieren} \tag{5}$$

Zudem müssen Eigenschaften mit inkompatiblen Identitätsbedingungen disjunkt sein.


T
-I -D +R
Universitätsangehöriger
+I +D ~R
Mensch
+O -D +R
Soziale Einheit
-I -D +R
phys. Objekt
+O -D +R
abstrakt. Objekt
-I -D +R
Person
+O -D +R
Organisation
+O +D +R
Zeugnis
+O -D +R
LehrBuch
+O -D +R
Bildungsabschluss
-I +D +R
Universitätsmitarbeiter
+I +D ~R
Student
+I +D ~R
Universität
+O +D +R
Abiturzeugnis
+O -D +R
Diplomzeugnis
+O -D +R
Abiturabschluss
-I +D +R
Diplomabschluss
-I -D +R
Doktorand
+I +D ~R

Dankeschön.

[Niemann 2004] Niemann, Hans-Joachim: *Lexikon des Kritischen Rationalismus*. Tübingen : Mohr Siebeck, 2004

[Smith 2003] Smith, Barry: Ontology. In: Floridi, Luciano (Hrsg.): *Blackwell Guide to the Philosophy of Computing and Information*. Oxford : Blackwell, 2003, S. 155–166

[Sowa 2000] Sowa, John F.: *Knowledge Representation: logical, philosophical, and computational foundations*. Pacific Grove : Brooks/Cole, 2000

[Welty und Guarino 2001] Welty, Christopher A. ; Guarino, Nicola: Supporting ontological analysis of taxonomic relationships. In: *Data Knowl. Eng.* 39 (2001), Nr. 1, S. 51–74