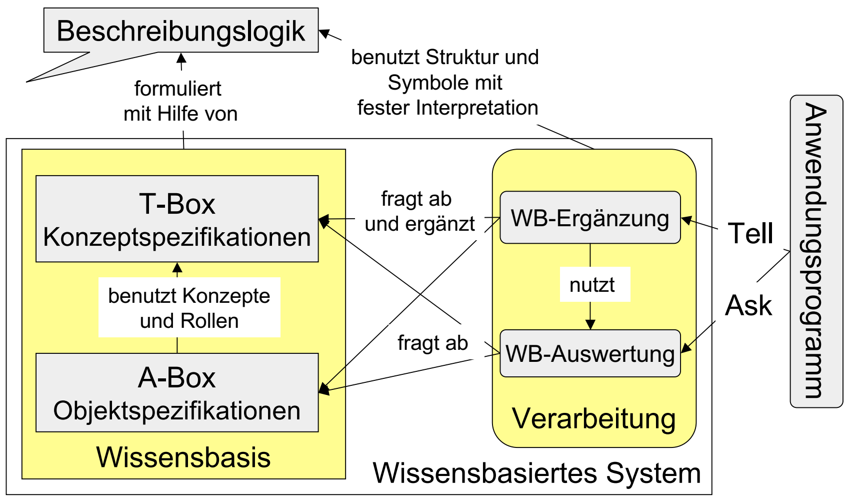
Wissensrepräsentation
–
Christopher Habel, Hedda Schmidtke
Sommersemester 2004

Sitzung 9: Beschreibungslogiken (3)
- Weltmodell (A-Box)
- Ergänzung der Wissensbasis
- Auswertung: Tableau-Verfahren

Wissensbasiertes System mit Beschreibungslogiken



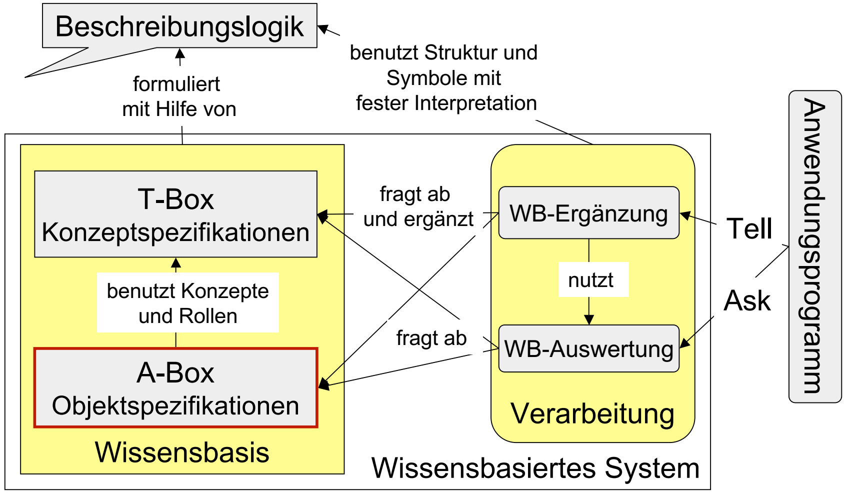
Wo stehen wir ?

Beschreibungslogiksprachen
- atomare Konzepte und Rollen
- komplexe Beschreibungen:
  - Konzeptbildungsoperatoren und Rollenbildungsoperatoren
  - Standardnotation & Syntax
  - Interpretation komplexer Beschreibungen auf Basis
    - der Interpretation der atomaren Konzepte und Rollen und
    - der festgelegten Interpretation der Konzept- und Rollenbildungsoperatoren

Wo stehen wir ?

T-Box-Struktur
- primitive vs. definierte (atomare) Konzepte
- Formelbildungsoperatoren
- definitorische T-Box
  - linke Seite der Formeln atomar, nur Äquivalenzen
  - jedes atomare Konzept max. einmal links
  - zyklenfreie T-Boxen
  - Basisinterpretation (primitive Konzepte) bestimmt Interpretation aller Konzepte
- Zyklische Definitionen
  - Fixpunktsemantik + Basisinterpretation bestimmt Interpretation aller Konzepte

Wissensbasiertes System mit Beschreibungslogiken



A-Box und Schnittstellen zur T-Box

A-Box (Assertionen, Weltmodell)

- enthält Formeln der Form C(a) (bei L&B, 2001: (a → C)) und R(a, b), wobei a und b Konstanten, C ein Konzept und R eine Rolle ist.

Interpretation < $\mathcal{I}$, $\mathcal{D}$ >

- jede Konstante wird durch $\mathcal{I}$ auf ein Objekt in $\mathcal{D}$ abgebildet.

$\mathcal{I}$(C(a)) = wahr, gdw. $\mathcal{I}$(a) ∈ $\mathcal{I}$(C)

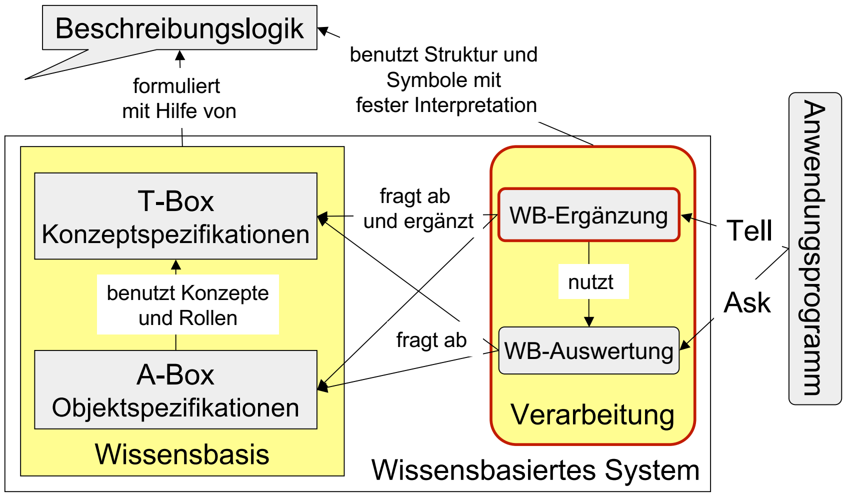
$\mathcal{I}$(R(a, b)) = wahr, gdw. ($\mathcal{I}$(a), $\mathcal{I}$(b)) ∈ $\mathcal{I}$(R)

Konzeptbildungsoperatoren mit Zugriff auf A-Box (?)

‚one of' : $\mathcal{I}(\{a_1, a_2, \ldots, a_n\}) = \{\mathcal{I}(a_1), \mathcal{I}(a_2), \ldots, \mathcal{I}(a_n)\}$

‚fills': $\mathcal{I}(R{:}a) = \{b \in \mathcal{D} \mid (b, \mathcal{I}(a)) \in \mathcal{I}(R)\}$

Wissensbasiertes System mit Beschreibungslogiken



Ergänzung der Wissensbasis: Taxonomien und Klassifikation

Taxonomie

- Subsumptionsstruktur (partielle Ordnung) von atomaren Konzepten, Gespeichert als zyklenfreier, gerichteter Graph
- Verbindungen nur zu den direkten Nachbarn

Klassifikation

- Bestimmung der Position eines neuen Konzeptes
  - 'unterhalb' aller subsumierenden Knoten
  - 'oberhalb' aller subsumierten Knoten
- Rahmenannahme
  - Der Graph wird sukzessive aufgebaut, beginnend mit dem einzigen Knoten für ⊤ bzw. Thing
  - Die relative Position der Knoten wird nur durch das Einfügen **neuer** Konzepte verändert. (keine Zyklen)

Berechnung der Klassifikation

Aufgabe: Ergänzung um ($C_{new} \doteq D$)

1) Bestimmung der Menge **S** der spezifischsten Knoten, die D bzgl. KB subsumieren.
2) Bestimmung der Menge **G** der allgemeinsten Knoten, die von D bzgl KB. subsumiert werden.
3) Wenn $\mathbf{S} \cap \mathbf{G} \neq \emptyset$, dann ist C_{new} schon durch einen Knoten vertreten und wir ergänzen dort das Symbol C_{new}.
4) Alle direkten Verbindungen zwischen Knoten von **G** und Knoten von **S** werden gelöscht.
5) Ein Knoten für C_{new} wird mit direkten (korrekt gerichteten) Verbindungen zu den Knoten von **S** und **G** eingehängt.
6) Alle Konstanten a, die unterhalb von allen Knoten von **S** aber keinem Knoten von **G** hängen, werden bestimmt und geprüft, ob KB $\models$ D(a) und die Konstanten ggf. geeignet umgehängt.

Berechnung der Klassifikation (2)

Aufgabe: Ergänzung um ($C_{new} \sqsubseteq D$)

- In diesem Fall muss nur die Menge **S** bestimmt werden, da kein altes Konzept von C_{new} subsumiert werden kann (die hinreichenden Bedingungen sind nicht bekannt.)

Aufgabe: Ergänzung um $D(a_{new})$

- In diesem Fall muss ebenfalls nur die Menge **S** bestimmt und a_{new} darunter eingehängt werden.

Bestimmung der Menge S: Vorgehen

Ziel

- **S** ist die Menge der spezifischsten Knoten, die D bzgl. KB subsumieren

Vorgehen

- $\mathbf{S_0} := \{\top\}$; i := 0;
- Solange es in $\mathbf{S_i}$ einen Knoten C_i gibt mit einem direkten Nachfolger, der D bzgl. KB subsumiert, mach folgendes:
  - $\mathbf{N_i}$ sei die Menge aller direkten Nachfolger von C_i, die D subsumieren
  - $\mathbf{S_{i+1}} := (\mathbf{S_i} \setminus \{C_i\}) \cup \mathbf{N_i}$
  - i := i + 1
- $\mathbf{S} := \mathbf{S_i}$

Bestimmung der Menge G: Vorgehen

Ziel

- **G** ist die Menge der allgemeinsten Knoten, die von D bzgl. KB subsumiert werden

Vorgehen

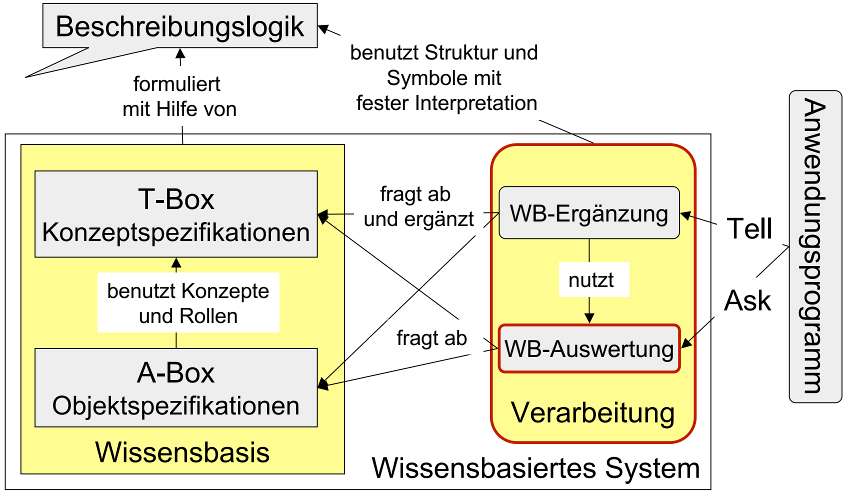
- $\mathbf{G_0} := \mathbf{S}$; i := 0;
- Solange es in $\mathbf{G_i}$ einen Knoten C_i gibt, der nicht von D bzgl. KB subsumiert wird, mach folgendes:
  - $\mathbf{N_i}$ sei die Menge aller direkten Nachfolger von C_i
  - $\mathbf{G_{i+1}} := (\mathbf{G_i} \setminus \{C_i\}) \cup \mathbf{N_i}$
  - i := i + 1
- $\mathbf{G} := \mathbf{G_i}$

Die Bestimmung von S und G

Beide Verfahren

- sind korrekt, wenn die Subsumptionsbeziehung berechnet werden kann
- sind durch das top-down-Vorgehen vergleichsweise effizient
  - wenn Subsumption effizient berechnet werden kann,
  - da Teilstrukturen der Taxonomie mit ihrem obersten Knoten von der Betrachtung ausgeschlossen werden können
  - da eine Orientierung an den direkten Nachfolgern erfolgen kann

Wissensbasiertes System mit Beschreibungslogiken



Auswertungsaufgaben

T-Box

- Erfüllbarkeit (gibt es ein Modell)
- Äquivalenz (von zwei T-Boxen)

Konzepte (basierend auf einer T-Box)

- Erfüllbarkeit vs. Widersprüchlichkeit
- Subsumption (zwischen zwei Beschreibungen)
- Äquivalenz (von zwei Beschreibungen)
- Exklusivität (Diskretheit)

A-Box (basierend auf einer T-Box)

- Erfüllbarkeit / Konsistenz (gibt es ein Modell)

Übersetzung in Prädikatenlogik

Beschreibungen

- entsprechen Formeln mit einer freien Variable

Übersetzungsfunktion ϕ: $\mathcal{Var} \times \mathcal{ALUENC} \rightarrow \mathcal{PL}$

- bildet eine Variable und eine Beschreibung auf eine Prädikatenlogische Formel ab, in der höchstens die genannte Variable frei vorkommt

Induktive Definition von ϕ (Anfang)

$\phi(x, A) = A(x)$

$\phi(x, \top) = \top(x)$

$\phi(x, C \sqcap D) = \phi(x, C) \wedge \phi(x, D)$

$\phi(x, C \sqcup D) = \phi(x, C) \vee \phi(x, D)$

$\phi(x, \neg C) = \neg\phi(x, C)$

$\phi(x, \bot) = \bot(x)$

Übersetzung in Prädikatenlogik (Forts.)

Induktive Definition von ϕ (Forts)

Hier wird davon ausgegangen, dass nur atomare Rollen vorliegen.

Es sei y eine Variable, die von x verschieden ist

$\phi(x, \forall R.C) = \forall y\, [R(x, y) \Rightarrow \phi(y, C)]$

$\phi(x, \exists R.\top) = \exists y\, [R(x, y)]$

$\phi(x, \exists R.D) = \exists y\, [R(x, y) \wedge \phi(y, D)]$

Es seien y_i Variable, die von x und voneinander verschieden sind

$\phi(x,(\geq n\ R)) = \exists y_1 \ldots y_n\, [R(x, y_1) \wedge \ldots \wedge R(x, y_n) \wedge y_1 \neq y_2 \wedge y_1 \neq y_3 \wedge y_2 \neq y_3 \wedge \ldots \wedge y_{n-1} \neq y_n]$

$\phi(x,(\leq n\ R)) = \forall y_1 \ldots y_{n+1}\, [R(x, y_1) \wedge \ldots \wedge R(x, y_{n+1}) \Rightarrow y_1 = y_2 \vee y_1 = y_3 \vee y_2 = y_3 \vee \ldots \vee y_n = y_{n+1}]$

Übersetzung in Prädikatenlogik: Beobachtungen

Die Übersetzung von $\mathcal{ALUEC}$

- benötigt die Identitätssymbole = und ≠ nicht
- kommt insgesamt mit zwei Variablen aus

➢ zwei-Variablen Fragment der Prädikatenlogik: Entscheidbar

Die Übersetzung von $\mathcal{ALUENC}$

- Einführung zählender Quantoren in $\mathcal{PL}$ mit zwei Variablen ergibt ebenfalls ein entscheidbares Fragment von $\mathcal{PL}$ (ohne Identitätssymbol).

Übersetzung in aussagenlogische Modallogik

- besonders fruchtbar für zyklische T-Boxen

Beschreibungen

- entsprechen aussagenlogischen Formeln

Rollen

- werden über Modalitäten repräsentiert

Übersetzungsfunktion ϕ: $\mathcal{ALUENC} \rightarrow \mathcal{ML}$

Induktive Definition von ϕ (Anfang)

$\phi(A) = A$

$\phi(\top) = \top$

$\phi(C \sqcap D) = \phi(C) \wedge \phi(D)$

$\phi(C \sqcup D) = \phi(C) \vee \phi(D)$

$\phi(\neg C) = \neg\phi(C)$

$\phi(\bot) = \bot$

Übersetzung in aussagenlogische Modallogik (Forts.)

Induktive Definition von ϕ (Forts)

Hier wird davon ausgegangen, dass nur atomare Rollen vorliegen.

- Für jede (atomare) Rolle R werden zwei duale Modal-Operatoren [R] und ⟨R⟩ eingeführt.

$\phi(\forall R.C) = [R]\, \phi(C)$

$\phi(\exists R.\top) = \langle R\rangle\, \top$

$\phi(\exists R.D) = \langle R\rangle\, \phi(D)$

- Beschreibungen sind genau dann erfüllbar, wenn ihre modallogische Übersetzung $\mathcal{K}$-erfüllbar ist. (➜ LOS)
- Das modallogische System $\mathcal{K}$ stellt keine Einschränkungen an die Sichtbarkeitsrelation.

Konsequenzen der Übersetzbarkeit

Verfahren
- die für Prädikatenlogik mit zwei Variablen oder
- $\mathcal{K}$-Modallogik mit mehreren Modalitäten entwickelt worden sind,
- sind für die Beschreibungslogiken einsetzbar,
- aber nicht unbedingt effizient

Beispiel
- Tableau-Verfahren

Strukturelle Verarbeitungsverfahren
- sind für Disjunktion, volle Negation und volle existentielle Restriktion nicht vollständig

Auswertungsaufgaben: Beziehungen

T-Box
- Erfüllbarkeit (gibt es ein Modell)
  - Rückführbar auf die **Erfüllbarkeit der A-Box** $\{\top(a_0)\}$ basierend auf der T-Box
  - jede definitorische T-Box ist erfüllbar
- Äquivalenz (von zwei T-Boxen)
  - Rückführbar auf die Prüfung der Gültigkeit der einzelnen Formeln in der jeweils anderen T-Box

Auswertungsaufgaben: Beziehungen

Konzepte (basierend auf einer T-Box)
- Erfüllbarkeit vs. Widersprüchlichkeit einer Beschreibung D
  - Rückführbar auf **Erfüllbarkeit der A-Box** $\{D(a_0)\}$ basierend auf der T-Box
- Subsumption (zwischen zwei Beschreibungen C und D)
  - D subsummiert C genau dann (bzgl. T-Box), wenn die **A-Box** $\{(C \sqcap \neg D)(a_0)\}$ (bzgl. T-Box) **nicht erfüllbar** ist
- Äquivalenz (von zwei Beschreibungen)
  - Rückführbar auf **Subsumption**
- Exklusivität (Diskretheit)
  - D und C sind genau dann exklusiv (bzgl. T-Box), wenn die **A-Box** $\{(C \sqcap D)(a_0)\}$ (bzgl. T-Box) **nicht erfüllbar** ist

Auswertungsaufgaben: Beziehungen

A-Box (basierend auf einer T-Box)
➢ Erfüllbarkeit / Konsistenz (gibt es ein Modell)

Wenn die T-Box definitorisch und azyklisch ist,
- können definierte Konzepte in den A-Box-Beschreibungen systematisch durch die Definitionen ersetzt werden.
- Dies ist eine Äquivalenzumformung

Die Erfüllbarkeit einer A-Box,
- in der keine definierten Konzepte vorkommen, bzgl. einer definitorischen T-Box (keine echten Inklusionen),
- kann ohne Berücksichtigung der T-Box (also mit leerer T-Box) geprüft werden.

Wissensbasiertes System mit Beschreibungslogiken

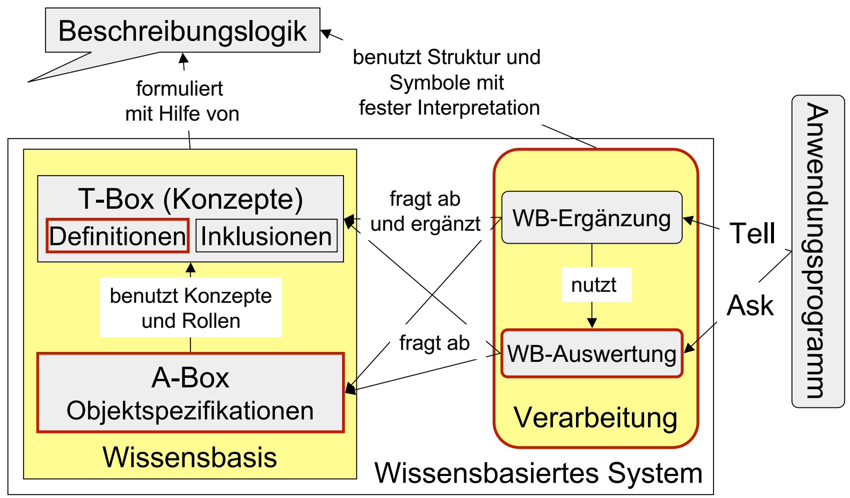


Tableau-Verfahren für Beschreibungslogiken

Tableau-Verfahren

- stellen (Un)Erfüllbarkeit fest
  - hier: Erfüllbarkeit von Beschreibungen
- konstruieren Modelle
  - hier: konsistente A-Box
- bilden disjunktive Normalformen
  - hier: alternative Modelle ➜ Model-Checking

Voraussetzung

- Übersetzung der eigentlichen Frage in eine Erfüllbarkeitsfrage (Beschreibung mit leerer T-Box)
- Herstellung der Negations-Normalform:
  - Negation tritt nur an den Konzeptsymbolen auf
- Bildung einer Instanz: $C(a_0)$

Beispiel: $\mathcal{ALUENC}$

Ist $C_0 = (\exists R.A) \sqcap (\exists R.B) \sqcap \neg(\exists R.A \sqcap B)$ **konsistent?**

NNF: $(\exists R.A) \sqcap (\exists R.B) \sqcap (\forall R.\neg A \sqcup \neg B)$

A-Box: $A_0 = \{((\exists R.A) \sqcap (\exists R.B) \sqcap (\forall R.\neg A \sqcup \neg B))(a_0)\}$

$\sqcap$: $A_1 = A_0 \cup \{(\exists R.A)(a_0), (\exists R.B)(a_0), (\forall R.\neg A \sqcup \neg B)(a_0)\}$

$\exists$: $A_2 = A_1 \cup \{R(a_0, a_1), A(a_1)\}$

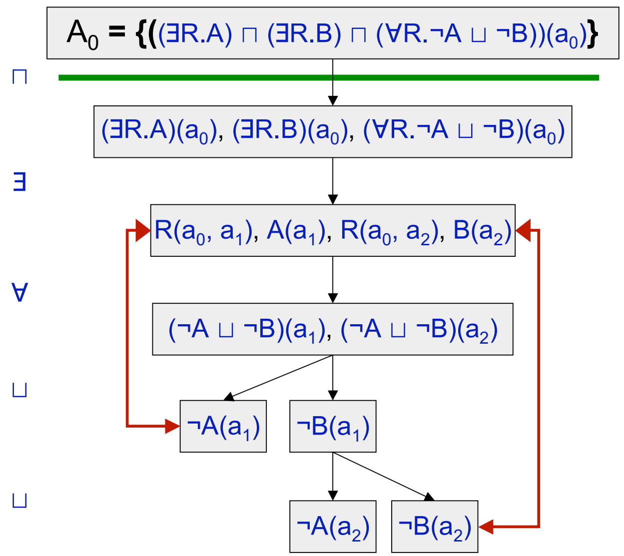
$\exists$: $A_3 = A_2 \cup \{R(a_0, a_2), B(a_2)\}$

$\forall$: $A_4 = A_3 \cup \{(\neg A \sqcup \neg B)(a_1), (\neg A \sqcup \neg B)(a_2)\}$

$\sqcup$: $A_5 = A_4 \cup \{\neg A(a_1)\}$ (Widerspruch), $A_6 = A_4 \cup \{\neg B(a_1)\}$

$\sqcup$: $A_7 = A_6 \cup \{\neg A(a_2)\}$, $A_8 = A_6 \cup \{\neg B(a_2)\}$ (Widerspruch)

A_7 ist maximal expandiert und widerspruchsfrei



Beispiel: $\mathcal{ALUENC}$

Ist $C_0 = (\exists R.A) \sqcap (\exists R.B) \sqcap \neg(\exists R.A \sqcap B) \sqcap \leq 1\ R$ **konsistent?**

NNF: $(\exists R.A) \sqcap (\exists R.B) \sqcap (\forall R.\neg A \sqcup \neg B) \sqcap \leq 1\ R$

A-Box: $A_0 = \{((\exists R.A) \sqcap (\exists R.B) \sqcap (\forall R.\neg A \sqcup \neg B) \sqcap \leq 1\ R)(a_0)\}$

$\sqcap$: $A_1 = A_0 \cup \{(\exists R.A)(a_0), (\exists R.B)(a_0), (\forall R.\neg A \sqcup \neg B)(a_0), \leq 1\ R(a_0)\}$

$\exists$: $A_2 = A_1 \cup \{R(a_0, a_1), A(a_1)\}$

$\exists$: $A_3 = A_2 \cup \{R(a_0, a_2), B(a_2)\}$

≤ 1: $A_4 = A_2 \cup \{B(a_1)\}$

$\forall$: $A_5 = A_4 \cup \{(\neg A \sqcup \neg B)(a_1)\}$

$\sqcup$: $A_6 = A_5 \cup \{\neg A(a_1)\}$ (Widerspruch)

$A_7 = A_5 \cup \{\neg B(a_1)\}$ (Widerspruch)

Alle Alternativen führen in einen Widerspruch

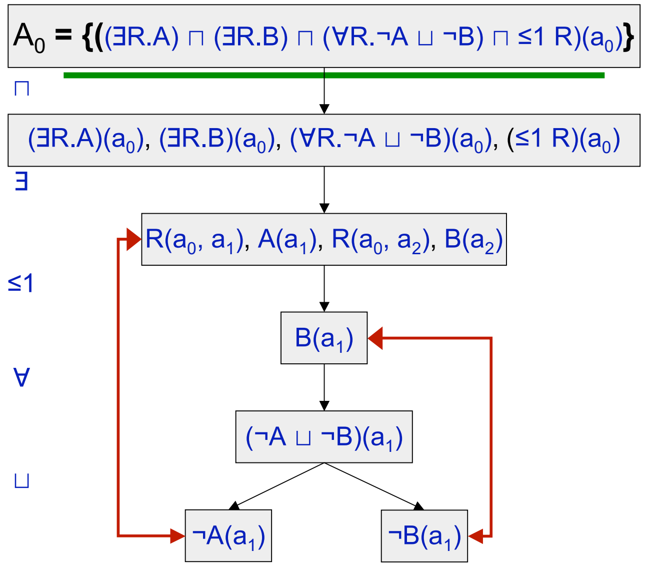


Tableau-Regeln für $\mathcal{ALUENC}$

$\sqcap$	$(C \sqcap D)(x)$	$A' = A \cup \{C(x), D(x)\}$	
$\sqcup$	$(C \sqcup D)(x)$	$A' = A \cup \{C(x)\}$ $A'' = A \cup \{D(x)\}$	
$\exists$	$(\exists R.C)(x)$	$A' = A \cup \{R(x, a), C(a)\}$	a ist neu
$\forall$	$(\forall R.C)(x)$, $R(x, y)$	$A' = A \cup \{C(y)\}$	
$\geq$	$(\geq n\ R)(x)$	$A' = A \cup \{R(x, a_i) \mid 1 \leq i \leq n\}$ $\cup \{a_i \neq a_j \mid 1 \leq i < j \leq n\}$	a_i sind neu
$\leq$	$(\leq n\ R)(x)$, $R(x, y_1)$, …, $R(x, y_{n+1})$	$A_{i,j} = A[y_j / y_i]$	$1 \leq i < j \leq n+1$; n*(n+1)/2 Zweige; Substitution von y_j durch y_i

Abschlussbedingungen für Tableau-Zweige

Abgeschlossene Tableau-Zweige enthalten Inkonsistenzen

- $\{D(x), \neg D(x)\} \subseteq A$
- $\{\bot(x)\} \subseteq A$
- $\{x \neq x\} \subseteq A$

Nicht-abgeschlossene voll-expandierte Tableau-Zweige

- repräsentieren Modelle des Konzeptes
- signalisieren Erfüllbarkeit des Konzeptes

Tableau-Verfahren für $\mathcal{ALUENC}$: Korrektheit

Korrektheit (Soundness)

- jede Expansionsregel führt zu einer Menge von A-Boxen, deren Disjunktion erfüllbarkeitsäquivalent zur Ausgangs-A-Box ist
- Die ursprüngliche A-Box ist genau dann erfüllbar, wenn die Disjunktion der resultierenden A-Boxen erfüllbar ist.
- Die ursprüngliche A-Box ist genau dann erfüllbar, wenn (mind.) eine der resultierenden A-Boxen erfüllbar ist.

Tableau-Verfahren für $\mathcal{ALUENC}$: Termination

Termination

- Ausgehend von einer A-Box {$C(a_0)$} können höchstens endlich viele (echte) Expansionsschritte vorgenommen werden.
- Das Tableau-Verfahren terminiert mit einer Menge von Zweigen, die abgeschlossen oder maximal expandiert sind.

Begründung

- In jeder Formel D(b), die in eine A-Box gelangt, ist D eine Teilbeschreibung von C. Für jedes b gibt es also nur endlich viele solche Einträge.
- Es werden höchstens endlich viele Konstanten eingeführt.
  - (Begründung nächste Folie)

Tableau-Verfahren für $\mathcal{ALUENC}$: Termination (2)

- Die Konstanten und Rollenzuordnungen bilden einen Baum mit der Wurzel a und c ist genau dann Nachfolger von b, wenn R(b, c) in der A-Box enthalten ist.
  - Die Konstanten und Rollenzuordnungen werden immer gemeinsam eingeführt.
  - Identifikation von Konstanten erfolgt nur, wenn einheitliche Rollenzuordnungen vorliegen.
- Jeder Knoten hat höchstens endlich viele direkte Nachfolgeknoten.
- Der Baum hat eine endliche Tiefe.
  - Beschreibungen von Nachfolgeknoten sind echte Teilbeschreibungen der Vorgängerknoten.
- Der Baum ist endlich.

Tableau-Verfahren für $\mathcal{ALUENC}$

Vollständigkeit

- Genau dann, wenn das Tableau-Verfahren mit einem nicht-abgeschlossenen Zweig terminiert, hat die ursprüngliche A-Box ein Modell.

Entscheidbarkeit

- Konzept-Erfüllbarkeit mit leerer T-Box ist für $\mathcal{ALUENC}$ entscheidbar.

Aufwand

- Das geschilderte Verfahren hat exponentielle Komplexität sowohl für zeit- als auch für Platzbedarf.
- Reduktion auf polynominellen Platzbedarf möglich.
- Aber: Konzept-Erfüllbarkeit ist für $\mathcal{ALUENC}$ ist PSpace-vollständig

Die Beschreibungslogik $\mathcal{ALUENC}$

Eine Beschreibung

- ist genau dann erfüllbar, wenn es ein endliches Modell gibt

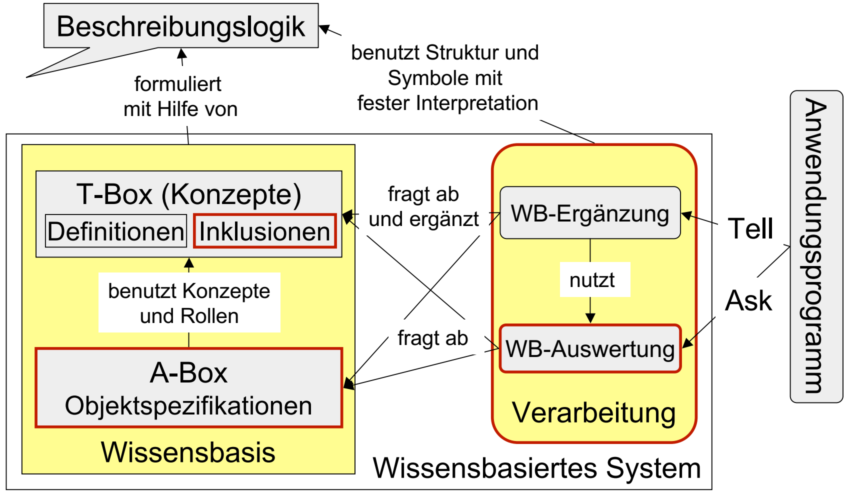
➢hat die endliche-Modell-Eigenschaft (finite model property)

Eine Beschreibung

- ist genau dann erfüllbar, wenn es ein endliches Modell gibt, das einen Baum bildet
- (jeder Knoten über genau eine Rollen-Folge erreichbar)

➢hat die endliche-Baum-Modell-Eigenschaft (tree model property)

Wissensbasiertes System mit Beschreibungslogiken



T-Box mit echten Inklusionen

echte Inklusion: Axiom der Form $(C \sqsubseteq D)$

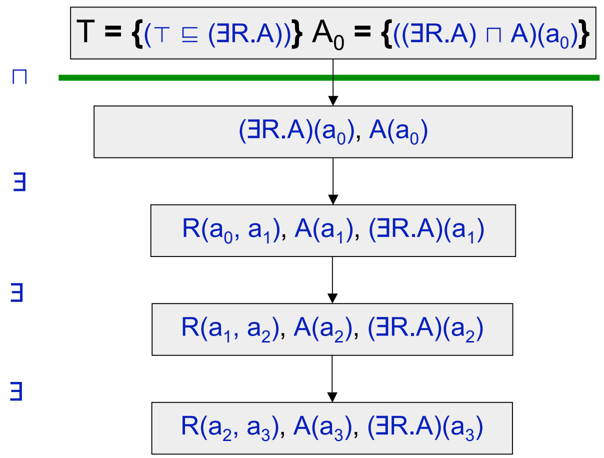
- wobei C definiert oder nicht atomar ist

Konzept-Konsistenz bzgl. T-Box

- in diesem Fall kann T-Box nicht ignoriert werden.

Vorgehen

- Fasse alle echten Inklusionen $(C_i \sqsubseteq D_i)$ der T-Box wie folgt zusammen: $\mathcal{C} = (\neg C_1 \sqcup D_1) \sqcap \ldots \sqcap (\neg C_n \sqcup D_n)$
- und verwende nur das T-Box Axiom $(\top \sqsubseteq \mathcal{C})$
- indem für jede Konstante a die Konzeptzuordnung $\mathcal{C}$(a) zugefügt wird
- Damit ist das Terminieren des Tableauverfahrens nicht mehr direkt gewährleistet.



Blockierung der Expansion

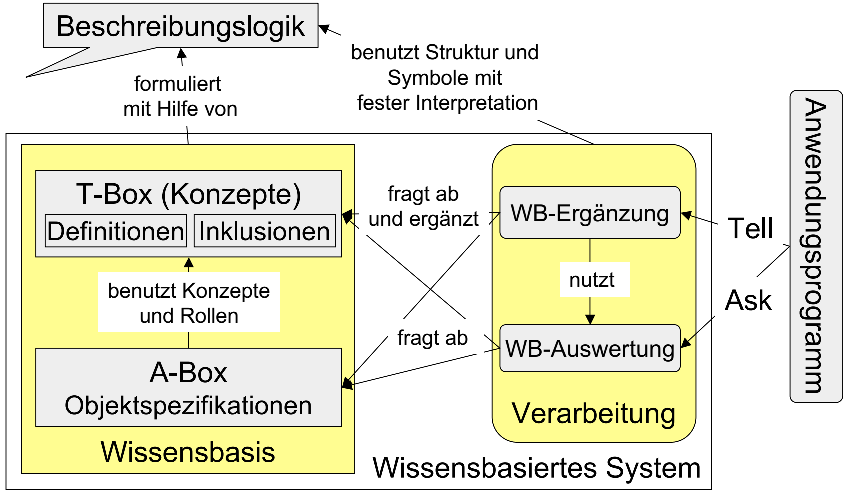
Ziel
- Erkennung, dass Abschluss unmöglich ist,
- und zwar nach endlich vielen Expansionen

Vorgehen
- Verhindern der Generierung neuer Konstanten

Bedingung
- (∃R.A)(b) ist zu expandieren,
- wobei in der A-Box eine Konstante a vorhanden ist,
- so dass alle Konzeptzuordnungen zu b entsprechend auch für a vorliegen (also auch (∃R.A)(a)) und
- für a die Expansionen schon vorgenommen wurden.

Begründung
- ein partielles Modell kann bzgl. b so erweitert werde, dass a und b denselben Wert erhalten.

Wissensbasiertes System mit Beschreibungslogiken



Aufgabe

Was ist zu tun
wenn eine Beschreibungslogik (und ein dazugehöriger Tableau-Beweiser) um einen weiteren Operator ergänzt werden soll?
- (Formelbildung, Konzeptbildung, Rollenbildung)
- Liste der Teil-Aufgaben

Beispiel
Bearbeiten Sie die Teilaufgaben für den Rollenbildungsoperator $\mathbf{R|_C}$