

TECHNISCHE UNIVERSITÄT DRESDEN

FAKULTÄT INFORMATIK

INSTITUT FÜR THEORETISCHE INFORMATIK

LEHRSTUHL FÜR AUTOMATENTHEORIE

DIPLOMARBEIT

**Existenz des Least Common Subsumers in
der Beschreibungslogik \mathcal{EL} bezüglich
genereller Terminologien**

Autor:
Benjamin Zarriß

Betreuerin:
Dr.-Ing. Anni-Yasmin Turhan

Betreuender Hochschullehrer:
Prof. Dr.-Ing. Franz Baader

31. Mai 2012

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbstständig erstellt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Dresden, den 31. Mai 2012

Benjamin Zarriß

Inhaltsverzeichnis

1	Einführung	3
1.1	Motivation	3
1.2	Bisherige Arbeiten	5
1.3	Inhaltsübersicht	6
2	Grundlagen	7
2.1	Beschreibungslogiken und Standardinferenzen	7
2.1.1	Syntax und Semantik von \mathcal{ALC}	7
2.1.2	Standardinferenzen	9
2.2	Least Common Subsumer	10
3	Existenz des \mathcal{EL}-LCS	15
3.1	Eine hinreichende Bedingung für die Existenz des LCS	15
3.2	Charakterisierung der Subsumption	16
3.3	Charakterisierung der Existenz des LCS	20
3.4	Entscheidbarkeit des Existenzproblems	26
4	Der \mathcal{EL}-LCS bezüglich einer \mathcal{ALC}-TBox	37
4.1	Normalisierung und Inferenzregeln für generelle \mathcal{ALC} -TBoxen	38
4.2	Vervollständigung von \mathcal{ALC} -TBoxen	40
4.3	Komplexität der Vervollständigung	49
5	Zusammenfassung und Ausblick	50
	Literaturverzeichnis	52

Kapitel 1

Einführung

1.1 Motivation

Durch die immer größer werdenden Mengen an Daten, die in modernen Informationssystemen verarbeitet werden müssen, werden Systeme immer wichtiger, die Möglichkeiten bieten, das vorhandene Wissen der zu Grunde liegenden Anwendungsbereiche strukturiert und formal zu repräsentieren. In diesem Zusammenhang sind wissensbasierte Systeme besonders interessant, da diese einen Formalismus zur Repräsentation von explizitem Wissen und Werkzeuge bieten, mit denen implizites Wissen abgeleitet werden kann.

In vielen dieser wissensbasierten Systeme werden Beschreibungslogiken erfolgreich angewendet. Beschreibungslogiken (BL) [BCM⁺03] sind eine Familie von Logiken zur Wissensrepräsentation. Sie bilden unter anderem die Grundlage für die standardisierte Ontologiesprache OWL (*web ontology language*) [HKP⁺09]. Die Wissensbasis eines BL-Systems besteht aus zwei Komponenten: der Terminologie, die auch als TBox bezeichnet wird, und der ABox. In der TBox können für eine Anwendungsdomäne relevante Konzepte und deren Beziehungen zueinander definiert, und in der ABox Individuen und deren Eigenschaften beschrieben werden. Zudem bietet ein BL-System eine Reihe von Werkzeugen, um, aus dem in der Terminologie explizit repräsentierten Wissen, neues implizites Wissen abzuleiten. Der praktische Nutzen von BL-Systemen liegt zum einen darin begründet, dass auf Grundlage der formalen Semantik Schlussfolgerungsprobleme definiert werden können und zum anderen darin, dass für Schlussfolgerungsverfahren, deren Korrektheit und Vollständigkeit für diese Probleme bewiesen wurden, effiziente Implementierungen existieren. Klassische Schlussfolgerungsverfahren sind zum Beispiel das Prüfen der Wissensbasis auf Konsistenz oder das Berechnen der Ober-/Unterkonzeptbeziehungen der Konzepte aus der Terminologie. Für viele Anwendungen sind jedoch auch Schlussfolgerungsverfahren wichtig, die es ermöglichen neues Wissen aus dem vorhandenen Wissen zu extrahieren, um zum Beispiel automatisch verschiedene Daten aus

der ABox in ein Konzept zu generalisieren.

Diese Arbeit beschäftigt sich mit diesen sogenannten Generalisierungsinferenzen in der Beschreibungslogik \mathcal{EL} bezüglich genereller Terminologien. Dazu zählt die Berechnung des *Least Common Subsumers* (LCS) und des *most specific concepts* (MSC). Der LCS beschreibt die speziellste Generalisierung für eine Reihe von Konzepten bezüglich einer Terminologie und die Berechnung des MSC dient zur Generalisierung von Individuen aus einer Wissensbasis in ein Konzept. Im Zusammenspiel sind diese Inferenzen vor allem für den Entwurf und die Erweiterung von Wissensbasen nützlich. So kann für Individuen jeweils das speziellste Konzept berechnet werden. Der LCS dieser speziellsten Konzepte beschreibt dann die Gemeinsamkeiten der Individuen. Durch das automatische Berechnen von Generalisierungen kann es somit auch Anwendern gelingen, die keine Experten auf dem Gebiet der Wissensrepräsentation oder Logik sind, die flexiblen Möglichkeiten zur Abstraktion, die Beschreibungslogiken bieten, zu nutzen. Berechnungsmethoden und auch weitere praktische Anwendungen für diese Inferenzen sind bereits intensiv erforscht worden [BKM99, BST07, MET11]. Jedoch wurde gezeigt, dass der LCS und das MSC in \mathcal{EL} bezüglich genereller Terminologien, die in vielen existierenden Wissensbasen verwendet werden, im Allgemeinen nicht existiert [Baa03].

Es gibt Algorithmen zur Berechnung von Approximationen dieser speziellsten Generalisierungen [PT10b], die für viele praktische Anwendungen ausreichend sind. Entscheidungsverfahren für die Probleme, ob der LCS oder das MSC bezüglich einer gegebenen Wissensbasis existieren und ob eine berechnete Generalisierung der speziellsten Generalisierung entspricht, sind aber noch nicht bekannt. Wurde zum Beispiel die Approximation des LCS berechnet, ist es jedoch interessant zu bestimmen, ob durch die Approximation Information verloren gegangen ist. Daraus ergeben sich die Fragen ob diese Approximation dem exakten LCS entspricht, ob eine speziellere Generalisierung existiert oder ob der exakte LCS bezüglich der gegebenen Terminologie für die beiden Konzepte überhaupt nicht existiert.

Das Ziel dieser Arbeit ist es, die exakten Bedingungen unter denen der LCS existiert zu charakterisieren und ein Entscheidungsverfahren für das Existenzproblem zu entwickeln. Außerdem soll die Methode zur Approximation des \mathcal{EL} -LCS bezüglich genereller \mathcal{EL} -TBoxen auf generelle \mathcal{ALC} -TBoxen erweitert werden.

Generalisierungsinferenzen und deren Anwendungen wurden insbesondere für Beschreibungslogiken der \mathcal{EL} -Familie schon intensiv erforscht. Im folgenden Abschnitt wird ein kurzer Überblick über bisherige Arbeiten zu diesem Thema vorgestellt.

1.2 Bisherige Arbeiten

Die Berechnung des LCS als Inferenz für BLen wurde erstmals in [CBH92] eingeführt, um mit dem LCS von Konzeptbeschreibungen alle Gemeinsamkeiten dieser Konzeptbeschreibungen zu berechnen. In einer ausdrucksstarken BL wie zum Beispiel \mathcal{ALC} ist der LCS von einer Menge von Konzeptbeschreibungen die Disjunktion dieser Konzeptbeschreibungen und bietet damit keine nützlichen Informationen. Daher sind die Berechnungsmethoden für den LCS meist auf ausdruckschwache BLen wie \mathcal{EL} oder \mathcal{ALE} beschränkt.

In [BKM99] wurde ein Algorithmus zur Berechnung des LCS für die Beschreibungslogiken \mathcal{EL} und \mathcal{ALE} bezüglich azyklischer, auffaltbarer TBoxen vorgestellt, und in [Baa03] wurde das Problem für \mathcal{EL} und zyklische TBoxen untersucht. Es wurde gezeigt, dass bereits bezüglich zyklischer TBoxen der \mathcal{EL} -LCS im Allgemeinen nicht existiert. Die Entscheidbarkeit des Existenzproblems des LCS für zyklische \mathcal{EL} -TBoxen wurde dann in [Baa04] bewiesen. Die Definition des LCS bezüglich zyklischer TBoxen in [Baa03, Baa04] erlaubt es die Terminologie zu erweitern, im Unterschied zur Definition des LCS bezüglich genereller TBoxen in [BST04, PT10b], die auch in dieser Arbeit verwendet wird. Die Nicht-Existenz des LCS wurde jedoch auch in diesem Fall in [BST04] gezeigt. Kriterien für die Existenz bzw. Nicht-Existenz des LCS wurden allerdings nicht angegeben.

Daraufhin wurden in [BST04, PT10b] Methoden für die Berechnung von Approximationen des LCS entwickelt und in [LPW10] Erweiterungen von \mathcal{EL} definiert, in denen der LCS und das MSC immer existieren. Dabei wurden in [BST04] auch erstmals \mathcal{EL} -Generalisierungen bezüglich genereller \mathcal{ALC} -Terminologien betrachtet, da in vielen Wissensbasen BLen wie \mathcal{ALC} oder Erweiterungen davon verwendet werden.

Alle Ansätze für die Berechnung von Generalisierungen basieren auf einer Charakterisierung der zu Grunde liegenden Standardinferenz. Im Fall des LCS ist das *Subsumption*, und im Fall des msc die *Instanzbeziehung*. Die ausdruckschwache Beschreibungslogik \mathcal{EL} ist in diesem Zusammenhang besonders interessant, da diese Inferenzen für generelle \mathcal{EL} -Wissensbasen in polynomieller Zeit entscheidbar sind [BBL05]. Der Algorithmus in [BBL05] basiert auf Vervollständigung, d. h. es werden alle impliziten Subsumptionsbeziehungen zwischen atomaren Konzepten mit Hilfe von Inferenzregeln abgeleitet. Auf diese Weise wird die Subsumptionshierarchie der atomaren Konzepte berechnet. Basierend auf dieser Struktur wurde in [PT10a] ein Berechnungsverfahren für Rollentiefen-beschränkte speziellere Generalisierungen bezüglich genereller \mathcal{EL} -TBoxen angegeben. Diese Approximationen sind mögliche Kandidaten für die speziellesten Generalisierungen. Das zugehörige Berechnungsverfahren und die in [LPW10] beschriebene Charakterisierung der Subsumptionsbeziehung mit kanonischen Modellen und Simulationen ist der Ausgangspunkt, für den in dieser Arbeit vorgestellten Ansatz zur Lösung

des Existenzproblems des LCS.

Für die Erweiterung der Methode zur Berechnung des Rollentiefenbeschränkten \mathcal{EL} -LCS auf generelle \mathcal{ALC} -TBoxen wird das Vervollständigungsverfahren für \mathcal{ALC} -TBoxen aus [SKH11] genutzt.

1.3 Inhaltsübersicht

Im nächsten Kapitel werden die grundlegenden Definitionen zu den Beschreibungslogiken \mathcal{EL} und \mathcal{ALC} vorgestellt. Anschließend wird eine exakte Charakterisierung der Existenz des LCS angegeben und darauf aufbauend die Entscheidbarkeit des Existenzproblems gezeigt. Außerdem wird die Komplexität des Entscheidungsverfahrens untersucht. Im darauf folgenden Kapitel wird die Methode zur Berechnung des Rollentiefenbeschränkten \mathcal{EL} -LCS auf generelle \mathcal{ALC} -TBoxen erweitert. Abschließend werden die Ergebnisse zusammengefasst und es wird ein Ausblick gegeben.

Kapitel 2

Grundlagen

In diesem Kapitel werden die formalen Definitionen zu Beschreibungslogiken und zu einigen klassischen Inferenzproblemen vorgestellt.

2.1 Beschreibungslogiken und Standardinferenzen

2.1.1 Syntax und Semantik von \mathcal{ALC}

Beschreibungslogiken sind eine Familie von Logiken zur Wissensrepräsentation, die sich hinsichtlich der verwendeten Syntax und damit der Ausdrucksstärke unterscheiden. In diesem Abschnitt wird die grundlegende BL \mathcal{ALC} mit deskriptiver Semantik vorgestellt. Es werden die Definitionen und Notation aus [BCM⁺03] verwendet.

Die grundlegenden Bausteine einer Beschreibungslogik sind die *Konzeptbeschreibungen*, die aus einer Menge von Konzeptnamen, einer Menge von Rollennamen und verschiedenen Konstruktoren aufgebaut sind.

Definition 2.1.1 (Syntax von \mathcal{ALC}). Sei N_C eine Menge von *Konzeptnamen* und N_R eine Menge von *Rollennamen*. Die Menge der \mathcal{ALC} -*Konzeptbeschreibungen* ist induktiv wie folgt definiert:

- \top, \perp sind \mathcal{ALC} -Konzeptbeschreibungen.
- Jedes $A \in N_C$ ist eine \mathcal{ALC} -Konzeptbeschreibung.
- Wenn C, D \mathcal{ALC} -Konzeptbeschreibungen sind und $r \in N_R$, dann sind auch $C \sqcap D$ (Konjunktion), $\neg C$ (Negation), $C \sqcup D$ (Disjunktion), $\exists r.C$ (Existenzrestriktion) und $\forall r.C$ (Wertesrestriktion) \mathcal{ALC} -Konzeptbeschreibungen.

Die Semantik von \mathcal{ALC} -Konzeptbeschreibungen wird mittels einer Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ definiert. Dabei ist $\Delta^{\mathcal{I}}$ eine nicht-leere Menge, die *Domäne* genannt wird. $\cdot^{\mathcal{I}}$ ist eine Abbildung, die jedes $A \in N_C$ auf eine Menge $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ abbildet. Jedem Rollennamen $r \in N_R$ wird durch

\mathcal{I} eine binäre Relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ zugeordnet. Für komplexere \mathcal{ALC} -Konzeptbeschreibungen wird die Abbildung \mathcal{I} wie folgt erweitert:

- $\top^{\mathcal{I}} := \Delta^{\mathcal{I}}; \perp^{\mathcal{I}} := \emptyset;$
- $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}; (C \sqcup D)^{\mathcal{I}} := C^{\mathcal{I}} \cup D^{\mathcal{I}};$
- $(\neg C)^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}};$
- $(\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}}\};$
- $(\forall r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \forall e \in \Delta^{\mathcal{I}} (d, e) \in r^{\mathcal{I}} \Rightarrow e \in C^{\mathcal{I}}\}$

Mit Hilfe von \mathcal{ALC} -Konzeptbeschreibungen C, D können nun Axiome der Form $C \sqsubseteq D$ formuliert werden. Diese Axiome werden als *generelle Konzeptinklusionen* (GCIs) bezeichnet. Eine *generelle TBox* ist eine Menge von GCIs. Es wird die Abkürzung $C \equiv D$ für das Paar $C \sqsubseteq D, D \sqsubseteq C$ verwendet. Für ein Individuum a aus einer Menge von Individuennamen N_I , gibt es Axiome der Form $C(a)$ (Konzeptassertion) und $r(a, b)$ (Rollenassertion). Dabei ist C eine Konzeptbeschreibung und r ein Rollenname. Die *ABox* ist eine Menge von Konzept- und Rollenassertionen. TBox und ABox bilden die Komponenten einer *Wissensbasis* $\mathcal{K} = (\mathcal{T}, \mathcal{A})$.

Die *Signatur* einer Konzeptbeschreibung C , bezeichnet mit $\text{sig}(C)$, ist die Menge der Konzept- und Rollenamen, die in C vorkommen. $N_{C, \mathcal{T}}$ bzw. $N_{R, \mathcal{T}}$ ist die Menge der Konzept- bzw. Rollenamen, die in der TBox \mathcal{T} vorkommen. $N_{C, \mathcal{T}}$ enthält zusätzlich das \top -Konzept. Die Signatur einer TBox ist dann $\text{sig}(\mathcal{T}) := N_{C, \mathcal{T}} \setminus \{\top\} \cup N_{R, \mathcal{T}}$.

Die Semantik einer Wissensbasis ist wie folgt definiert:

Definition 2.1.2. Sei $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ eine Wissensbasis mit der TBox \mathcal{T} und der ABox \mathcal{A} . Eine Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ ist ein Modell für \mathcal{T} gdw. für alle Konzeptinklusionen $C \sqsubseteq D \in \mathcal{T}$: $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ gilt. \mathcal{I} ist ein Modell für \mathcal{A} gdw. für alle Konzeptassertionen $C(a) \in \mathcal{A}$ und Rollenassertionen $r(a, b) \in \mathcal{A}$ gilt: $a^{\mathcal{I}} \in C^{\mathcal{I}}$ bzw. $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$. \mathcal{I} ist ein Modell für \mathcal{K} genau dann, wenn \mathcal{I} ein Modell für \mathcal{T} und ein Modell für \mathcal{A} ist. Eine Wissensbasis \mathcal{K} ist genau dann *konsistent*, wenn \mathcal{K} ein Modell hat.

Die Beschreibungslogik \mathcal{EL} ist eine Teilsprache von \mathcal{ALC} . Es sind nur die Konstruktoren \top , Konjunktion und Existenzrestriktion erlaubt. Ein Wissensbasis $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ bzw. eine TBox \mathcal{T} wird als \mathcal{EL} -Wissensbasis bzw. als \mathcal{EL} -TBox bezeichnet genau dann, wenn nur \mathcal{EL} -Konzeptbeschreibungen mit den genannten Konstruktoren in \mathcal{K} bzw. \mathcal{T} vorkommen.

Im folgenden Abschnitt werden einige grundlegende Schlussfolgerungsprobleme eingeführt.

2.1.2 Standardinferenzen

Auf Grundlage der formalen Semantik lassen sich Schlussfolgerungsprobleme definieren. Zunächst werden die beiden klassischen Inferenzen *Subsumption* und das *Instanzproblem* betrachtet.

Subsumption. Seien C, D Konzeptbeschreibungen und \mathcal{T} eine TBox. D subsumiert C bezüglich \mathcal{T} genau dann, wenn für alle Modelle \mathcal{I} von \mathcal{T} : $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ gilt. Dies wird abgekürzt mit $C \sqsubseteq_{\mathcal{T}} D$.

Instanzproblem. Sei $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ eine Wissensbasis, a ein Individuum und C eine Konzeptbeschreibung. a ist Instanz von C ($\mathcal{K} \models C(a)$) genau dann, wenn $a^{\mathcal{I}} \in C^{\mathcal{I}}$ für alle Modelle \mathcal{I} von \mathcal{K} .

Zwei Konzeptbeschreibungen C, D sind äquivalent bezüglich einer TBox \mathcal{T} , abgekürzt mit $C \equiv_{\mathcal{T}} D$, genau dann, wenn $C \sqsubseteq_{\mathcal{T}} D$ und $D \sqsubseteq_{\mathcal{T}} C$ gilt. An Stelle von einzelnen Subsumptions- oder Instanzanfragen sind in der Praxis vor allem die beiden darauf aufbauenden komplexen Inferenzen *Klassifikation* und *ABox-Realisierung* von Bedeutung.

Klassifikation. Es wird für eine TBox \mathcal{T} und für alle Paare von Konzeptnamen $A, B \in N_{C, \mathcal{T}}$ berechnet ob $A \sqsubseteq_{\mathcal{T}} B$ gilt.

ABox-Realisierung. Für eine Wissensbasis \mathcal{K} , alle Individuen a und alle Konzeptnamen B , die in \mathcal{K} vorkommen, wird berechnet ob $\mathcal{K} \models B(a)$ gilt.

Für \mathcal{EL} -Wissensbasen sind diese beiden Inferenzen in polynomieller Zeit berechenbar [BBL05]. Der Algorithmus aus [BBL05] für diese Inferenzen wird im Folgenden vorgestellt. Zunächst werden die TBox und ABox normalisiert.

Definition 2.1.3 (Normalform). Sei \mathcal{T} eine \mathcal{EL} -TBox und $A, B \in N_{C, \mathcal{T}}$ und $r \in N_{R, \mathcal{T}}$. \mathcal{T} ist in Normalform genau dann, wenn alle Konzeptinklusionen in \mathcal{T} eine der folgenden Formen haben:

$$\bigsqcap_i^n A_i \sqsubseteq B \quad (n \leq 2) \qquad A \sqsubseteq \exists r.B \qquad \exists r.A \sqsubseteq B$$

Sei \mathcal{A} eine ABox und $a \in N_I$ und B ein Konzeptname. \mathcal{A} ist in Normalform genau dann, wenn alle Konzeptassertionen von der Form $B(a)$ sind.

Zu jeder \mathcal{EL} -TBox \mathcal{T} kann in linearer Zeit durch Einführung von neuen Konzeptnamen eine normalisierte TBox \mathcal{T}' mit den Regeln in Abb. 2.1 berechnet werden [BBL05]. Die ABox wird normalisiert, indem eine Konzeptassertion der Form $C(a)$ mit einer komplexen Konzeptbeschreibung C durch $B(a)$ mit

$$\begin{aligned}
C \sqcap \hat{D} \sqsubseteq E &\rightarrow \{\hat{D} \sqsubseteq [\hat{D}], C \sqcap [\hat{D}] \sqsubseteq E\} \\
\exists r. \hat{C} \sqsubseteq D &\rightarrow \{\hat{C} \sqsubseteq [\hat{C}], \exists r. [\hat{C}] \sqsubseteq D\} \\
\hat{C} \sqsubseteq \hat{D} &\rightarrow \{\hat{C} \sqsubseteq [\hat{C}], [\hat{C}] \sqsubseteq D\} \\
B \sqsubseteq \exists r. \hat{C} &\rightarrow \{B \sqsubseteq \exists r. [\hat{C}], [\hat{C}] \sqsubseteq \hat{C}\} \\
B \sqsubseteq C \sqcap D &\rightarrow \{B \sqsubseteq C, B \sqsubseteq D\} \\
\hat{C}, \hat{D} \notin N_{C, \mathcal{T}}; [\hat{C}], [\hat{D}] &\text{ sind neue Konzeptnamen.}
\end{aligned}$$

Abbildung 2.1: \mathcal{EL} -Normalisierungsregeln

einem neuen Konzeptnamen B ersetzt wird und $B \equiv C$ zur TBox hinzugefügt wird.

Durch Anwendung von Vervollständigungsregeln werden neue Subsumptions- bzw. Instanzbeziehungen abgeleitet. Die Struktur, die dabei generiert wird ist ein *Vervollständigungsgraph* und wird in der folgenden Definition beschrieben.

Definition 2.1.4 (Vervollständigungsgraph). Sei $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ eine normalisierte \mathcal{EL} -Wissensbasis. N_R ist die Menge der Rollennamen und N_I die Menge der Individuen, die in \mathcal{K} vorkommen. Der *Vervollständigungsgraph* $\mathcal{G}_{\mathcal{K}} = (V, E, S)$ ist ein beschrifteter Graph mit der Knotenmenge $V = N_{C, \mathcal{T}} \cup N_I$, den Kanten $E \subseteq V \times N_R \times V$ und der Beschriftungsfunktion $S : V \rightarrow 2^{N_{C, \mathcal{T}}}$. $\mathcal{G}_{\mathcal{K}}$ erfüllt die folgenden Eigenschaften:

- $A \sqsubseteq_{\mathcal{T}} B$ genau dann, wenn $B \in S(A)$;
- $A \sqsubseteq_{\mathcal{T}} \exists r. B$ genau dann, wenn $(A, r, B') \in E$ und $B \in S(B')$;
- $\mathcal{K} \models B(a)$ genau dann, wenn $B \in S(a)$ und
- $\mathcal{K} \models (\exists r. A)(a)$ genau dann, wenn $(a, r, A') \in E$ mit $A \in S(A')$.

Aus dem Vervollständigungsgraphen können alle implizierten Subsumptions- und Instanzbeziehungen der Wissensbasis abgelesen werden. Der Vervollständigungsgraph bietet damit die Grundlage für die Berechnung von Generalisierungen. Als ein Beispiel dafür wird im nächsten Abschnitt der *Least Common Subsumer* betrachtet.

2.2 Least Common Subsumer

In diesem Abschnitt wird der Least Common Subsumer definiert und einige bekannte Resultate zur Existenz und zu Approximationen des LCS zusammengefasst.

Die Idee des Least Common Subsumer (LCS) besteht darin die Gemeinsamkeiten einer Menge von Konzeptbeschreibungen C_1, \dots, C_n zu erfassen. Es wird hier nur der binäre \mathcal{EL} -LCS bezüglich einer generellen TBox betrachtet. Der LCS ist formal wie folgt definiert.

Definition 2.2.1 (LCS). Sei \mathcal{T} eine TBox und D_1, D_2 und C \mathcal{EL} -Konzeptbeschreibungen. C ist der LCS von D_1 und D_2 bezüglich \mathcal{T} genau dann, wenn die folgenden Bedingungen erfüllt sind:

- (C1) Es gilt $D_1 \sqsubseteq_{\mathcal{T}} C$ und $D_2 \sqsubseteq_{\mathcal{T}} C$.
- (C2) Für alle \mathcal{EL} -Konzeptbeschreibungen E mit $D_1 \sqsubseteq_{\mathcal{T}} E$ und $D_2 \sqsubseteq_{\mathcal{T}} E$ gilt $C \sqsubseteq_{\mathcal{T}} E$.

Eine \mathcal{EL} -Konzeptbeschreibung C' , die $D_i \sqsubseteq_{\mathcal{T}} C'$ für $i = 1, 2$ erfüllt, wird *Common Subsumer* von D_1 und D_2 bezüglich \mathcal{T} genannt.

Eine \mathcal{EL} -Konzeptbeschreibung die (C1) erfüllt existiert immer und ist gegeben durch \top . Zudem ist der LCS modulo Äquivalenz eindeutig bestimmt. Es kann demnach keine bezüglich Subsumption nicht vergleichbaren Common Subsumer geben, die die Bedingung (C2) erfüllen. Der \mathcal{EL} -LCS existiert daher genau dann nicht, wenn die Menge der Common Subsumer unendlich ist und eine unendliche Kette von immer spezieller werdenden Common Subsumern existiert. Zum Beispiel existiert der LCS von A und B bezüglich der \mathcal{EL} -TBox $\mathcal{T} := \{A \sqsubseteq \exists r.A, B \sqsubseteq \exists r.B\}$ nicht [BST04], da es eine unendliche Folge $\exists r.\top, \exists r.\exists r.\top, \exists r.\exists r.\exists r.\top, \dots$ von spezieller werdenden Common Subsumern von A und B gibt.

Um dieses Problem zu vermeiden, wurde in [PT10b] eine Schranke für die *Rollentiefe* der Common Subsumer eingeführt. Die Rollentiefe rt ist definiert für eine \mathcal{EL} -Konzeptbeschreibung C . $rt(C)$ ist 0, wenn C ein Konzeptnamen oder \top ist. Für eine Konjunktion $C \sqcap D$ und eine \mathcal{EL} -Konzeptbeschreibung D gilt $rt(C \sqcap D) := \max(rt(C), rt(D))$ und für eine Existenzrestriktion $\exists r.D$ gilt $rt(\exists r.D) := 1 + rt(D)$. Der k -lcs von \mathcal{EL} -Konzeptbeschreibungen D_1, D_2 bezüglich einer TBox \mathcal{T} ist definiert wie in 2.2.1. Zusätzlich zu (C1) gilt $rt(C) \leq k$ und zusätzlich zu (C2) gilt $rt(E) \leq k$ für die natürliche Zahl k . Der k -lcs ist damit der speziellste Common Subsumer mit einer Rollentiefe $\leq k$. Der (binäre) k -lcs existiert immer und kann in polynomieller Zeit berechnet werden [PT10b]. Dazu wird der Produktgraph des Vervollständigungsgraphen aus Definition 2.1.4 gebildet.

Definition 2.2.2 (Produktgraph). Sei \mathcal{T} eine \mathcal{EL} -TBox und $\mathcal{G}_{\mathcal{T}'} = (V', E', S')$ der Vervollständigungsgraph für die normalisierte TBox \mathcal{T}' von \mathcal{T} . $\mathcal{G}_{\mathcal{T}'}$ ist definiert durch $\mathcal{G}_{\mathcal{T}'} := \mathcal{G}_{\mathcal{K}}$ mit $\mathcal{K} = (\mathcal{T}', \emptyset)$. Der Produktgraph ist der Graph $\mathcal{G}_{\mathcal{T}'} \times \mathcal{G}_{\mathcal{T}'} = (V, E, S)$ mit

- $V := V' \times V'$,

- $E := \{((A, B), r, (C, D)) \mid (A, r, C), (B, r, D) \in E'\}$ und
- $S((A, B)) := S'(A) \cap S'(B)$.

Der k -lcs von D_1 und D_2 bezüglich \mathcal{T} wird berechnet, indem zunächst der Produktgraph zu der TBox $\mathcal{T} \cup \{A \equiv D_1, B \equiv D_2\}$ im Knoten (A, B) bis zur Tiefe k abgewickelt wird. Das Konzept, das bei der Abwicklung entsteht wird *Traversierungskonzept* genannt und ist wie folgt definiert:

Definition 2.2.3 (Traversierungskonzept). Sei $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Produktgraph zur \mathcal{EL} -TBox \mathcal{T} , k eine natürliche Zahl und $(A, B) \in V$. Die Funktion $\text{trav}_{\mathcal{G}_{\mathcal{T}}}$ ist rekursiv definiert durch:

$$\begin{aligned} \text{trav}_{\mathcal{G}_{\mathcal{T}}}(0, (A, B)) &= \bigsqcap_{D \in S((A, B))} D; \\ \text{trav}_{\mathcal{G}_{\mathcal{T}}}(k, (A, B)) &= \bigsqcap_{D \in S((A, B))} D \sqcap \bigsqcap_{((A, B), r, (E, F)) \in E} \exists r. (\text{trav}_{\mathcal{G}_{\mathcal{T}}}(k-1, (E, F))). \end{aligned}$$

Die Berechnung des Traversierungskonzeptes beinhaltet die Einführung neuer Konzeptnamen, da bei der Berechnung des Produktgraphen die TBox normalisiert wird. Diese neuen Konzeptnamen werden *Normalisierungsnamen* genannt. Um den k -lcs bezüglich der ursprünglichen Signatur zu berechnen, werden die Normalisierungsnamen in dem Traversierungskonzept durch \top ersetzt. Das Traversierungskonzept mit den ersetzten Normalisierungsnamen wird mit $\hat{\text{trav}}$ bezeichnet.

Satz 2.2.1 ([PT10b]). Seien \mathcal{T} eine \mathcal{EL} -TBox, D_1, D_2 \mathcal{EL} -Konzeptbeschreibungen, $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Produktgraph zu $\mathcal{T} \cup \{A \equiv D_1, B \equiv D_2\}$ mit zwei neuen Konzeptnamen A, B , \mathcal{T}' die normalisierte TBox zu $\mathcal{T} \cup \{A \equiv D_1, B \equiv D_2\}$ und k eine natürliche Zahl. $\text{trav}_{\mathcal{G}_{\mathcal{T}'}}(k, (A, B))$ ist der k -lcs von D_1 und D_2 bezüglich \mathcal{T}' und $\hat{\text{trav}}_{\mathcal{G}_{\mathcal{T}'}}(k, (A, B))$ ist der k -lcs von D_1 und D_2 bezüglich \mathcal{T} .

Die Korrektheit der Ersetzung der Normalisierungsnamen folgt aus diesem Lemma:

Lemma 2.2.1 ([PT10a]). Sei \mathcal{T} eine \mathcal{EL} -TBox, \mathcal{T}' die normalisierte TBox zu \mathcal{T} nach Anwendung der Regel in Abb. 2.1, C, D \mathcal{EL} -Konzeptbeschreibungen mit $\text{sig}(C) \cup \text{sig}(D) \subseteq \text{sig}(\mathcal{T}')$ und $C'(D')$ die Konzeptbeschreibung, die aus $C(D)$ entstanden ist durch Entfernen aller Namen $A \in \text{sig}(\mathcal{T}') \setminus \text{sig}(\mathcal{T})$ aus $C(D)$. Es gilt $C \sqsubseteq_{\mathcal{T}'} D$ genau dann, wenn $C' \sqsubseteq_{\mathcal{T}} D'$.

Damit kann auch bewiesen werden, dass die Existenz des exakten LCS unabhängig von der Normalform ist.

Lemma 2.2.2. *Sei \mathcal{T} eine \mathcal{EL} -TBox und D_1, D_2 \mathcal{EL} -Konzeptbeschreibungen mit $\text{sig}(D_1) \cup \text{sig}(D_2) \subseteq \text{sig}(\mathcal{T})$. Der LCS von D_1 und D_2 bezüglich \mathcal{T} existiert genau dann, wenn der LCS für neue Konzeptnamen A, B bezüglich der normalisierten TBox \mathcal{T}' zu $\mathcal{T} \cup \{A \equiv D_1, B \equiv D_2\}$ existiert.*

Beweis. Sei K mit $\text{rt}(K) = k$ der LCS von A, B bezüglich \mathcal{T}' . Es wird gezeigt, dass der LCS von D_1, D_2 bezüglich \mathcal{T} ebenfalls existiert. Es gilt nach Satz 2.2.1:

$$K \equiv_{\mathcal{T}'} \text{trav}_{\mathcal{G}_{\mathcal{T}'}}(k, (A, B)) \quad (2.1)$$

Dann ist $\hat{K} = \hat{\text{trav}}_{\mathcal{G}_{\mathcal{T}'}}(k, (A, B))$ nach Satz 2.2.1 der k -lcs(D_1, D_2) bezüglich \mathcal{T} . Angenommen \hat{K} ist *nicht* der exakte LCS. Dann existiert ein $l > k$ und es gilt für den $\hat{L} = l$ -lcs(D_1, D_2) bezüglich \mathcal{T} :

$$\hat{L} \sqsubseteq_{\mathcal{T}} \hat{K} \text{ und } \hat{K} \not\sqsubseteq_{\mathcal{T}} \hat{L} \quad (2.2)$$

und

$$\hat{L} = \hat{\text{trav}}_{\mathcal{G}_{\mathcal{T}'}}(l, (A, B)) \quad (2.3)$$

Da K der exakte LCS bezüglich \mathcal{T}' ist, folgt mit (2.1):

$$\text{trav}_{\mathcal{G}_{\mathcal{T}'}}(k, (A, B)) \sqsubseteq_{\mathcal{T}'} \text{trav}_{\mathcal{G}_{\mathcal{T}'}}(l, (A, B)). \quad (2.4)$$

Aus Lemma 2.2.1 folgt:

$$\hat{\text{trav}}_{\mathcal{G}_{\mathcal{T}'}}(k, (A, B)) \sqsubseteq_{\mathcal{T}} \hat{\text{trav}}_{\mathcal{G}_{\mathcal{T}'}}(l, (A, B)) \quad (2.5)$$

und mit (2.3) auch $\hat{K} \sqsubseteq_{\mathcal{T}} \hat{L}$ im Widerspruch zu (2.2).

Die andere Richtung kann analog gezeigt werden, da Lemma 2.2.1 für beide Richtungen gilt. \square

Weiterhin lässt sich die folgende Charakterisierung des exakten LCS ableiten.

Lemma 2.2.3. *Sei \mathcal{T} eine \mathcal{EL} -TBox, D_1, D_2 \mathcal{EL} -Konzeptbeschreibungen und $k \in \mathbb{N}$. Es gilt $\forall l \in \mathbb{N} : k$ -lcs(D_1, D_2) $\sqsubseteq_{\mathcal{T}}$ l -lcs(D_1, D_2) genau dann, wenn k -lcs(D_1, D_2) der exakte LCS ist.*

Beweis. (Skizze) „ \Rightarrow “: Angenommen $k \in \mathbb{N}$ und es gilt $\forall l \in \mathbb{N} : k$ -lcs(D_1, D_2) $\sqsubseteq_{\mathcal{T}}$ l -lcs(D_1, D_2). Es ist zu zeigen, dass k -lcs(D_1, D_2) (C1) und (C2) aus Definition 2.2.1 erfüllt. (C1) ist erfüllt. Es wird (C2) gezeigt. Sei E eine beliebige \mathcal{EL} Konzeptbeschreibung mit $D_1 \sqsubseteq_{\mathcal{T}} E$ und $D_2 \sqsubseteq_{\mathcal{T}} E$ und $\text{rt}(E) = l'$. Nach Definition des l' -lcs folgt l' -lcs(D_1, D_2) $\sqsubseteq_{\mathcal{T}} E$ und nach Annahme gilt k -lcs(D_1, D_2) $\sqsubseteq_{\mathcal{T}}$ l' -lcs(D_1, D_2). Daraus folgt k -lcs(D_1, D_2) $\sqsubseteq_{\mathcal{T}} E$.

„ \Leftarrow “: Angenommen k -lcs(D_1, D_2) ist der exakte LCS. Alle Konzeptbeschreibungen der Form l -lcs(D_1, D_2) für eine natürliche Zahl l sind Common Subsumer von D_1 und D_2 und subsumieren den exakten LCS nach Definition 2.2.1 (C2). Die Behauptung folgt daher mit vollständiger Induktion über l . \square

Die Rollentiefen- beschränkten Approximationen des LCS sind demnach mögliche Kandidaten für den exakten LCS. Das Problem besteht darin eine geeignete Rollentiefe k zu finden. Ist die berechnete Approximation des LCS zum Beispiel zu allgemein für eine bestimmte Anwendung, dann stellt sich die Frage ob der exakte LCS existiert und wie weit der Produktgraph traversiert werden muss, um den LCS zu berechnen. Dieses Problem wird im nächsten Kapitel untersucht.

Kapitel 3

Existenz des \mathcal{EL} -LCS bezüglich genereller \mathcal{EL} -TBoxen

In diesem Kapitel wird ein Entscheidungsverfahren für das Existenzproblem des \mathcal{EL} -LCS bezüglich genereller \mathcal{EL} -TBoxen entwickelt.

Als Einstieg wird im ersten Abschnitt ein Beispiel für die Nicht-Existenz des LCS untersucht und eine hinreichende Bedingung für die Existenz skizziert. Im nachfolgenden Abschnitt wird die Charakterisierung der Subsumption bezüglich genereller \mathcal{EL} -TBoxen aus [LW10] genutzt, um daraus eine hinreichende und notwendige Bedingung für die Existenz des LCS abzuleiten. Anschließend wird die Entscheidbarkeit des Existenzproblems gezeigt und eine obere Schranke für die Rollentiefe des LCS angegeben.

3.1 Eine hinreichende Bedingung für die Existenz des LCS

Betrachtet wird die \mathcal{EL} -TBox

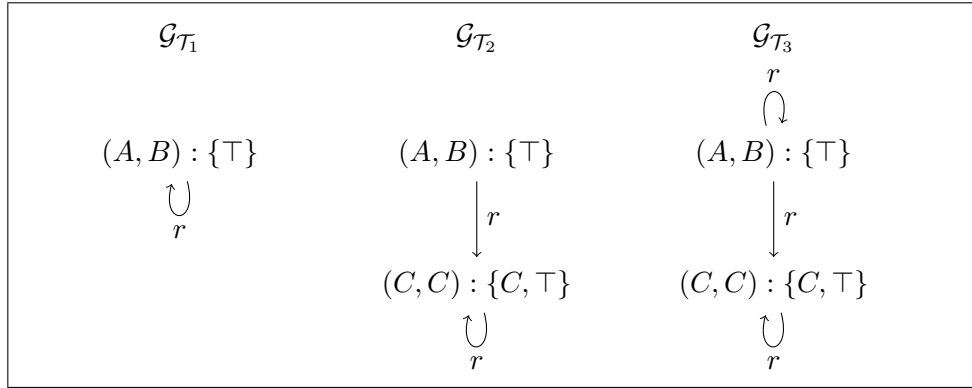
$$\mathcal{T}_1 := \{A \sqsubseteq \exists r.A, B \sqsubseteq \exists r.B\}$$

mit dem Produktgraphen $\mathcal{G}_{\mathcal{T}_1}$ in Abbildung 3.1. Wie in Abschnitt 2.2 beschrieben, existiert der LCS von A und B nicht. Der Zyklus um den Knoten (A, B) im Produktgraphen kann beliebig tief abgewickelt werden und es ergeben sich immer speziellere Common Subsumer.

Der Produktgraph $\mathcal{G}_{\mathcal{T}_2}$ zu der TBox

$$\mathcal{T}_2 := \{A \sqsubseteq \exists r.C, B \sqsubseteq \exists r.C, C \sqsubseteq \exists r.C\}$$

ist ebenfalls zyklisch mit der Schleife um den Knoten (C, C) . Doch in diesem Fall existiert der LCS von A und B bezüglich \mathcal{T}_2 und ist gegeben durch


Abbildung 3.1: zyklische Produktgraphen zu $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$

$\exists r.C$. Wird der Produktgraph weiter traversiert ergeben sich Konzeptbeschreibungen der Form $\exists r.(C \sqcap \exists r.(C \sqcap \dots))$, die nicht spezieller werden, da $\exists r.C \sqsubseteq_{\mathcal{T}_2} \exists r.(C \sqcap \exists r.(C \sqcap \dots))$ für beliebige Abwicklungen gilt. Ist der Produktgraph azyklisch, existiert der LCS immer, da der Produktgraph in diesem Fall komplett traversiert und in eine Konzept übersetzt werden kann. Jedoch ist ein azyklischer Produktgraph nur eine hinreichende Bedingung für die Existenz des LCS wie das Beispiel für \mathcal{T}_2 zeigt. Vereinfacht dargestellt gibt es demnach Zyklen im Produktgraphen wie in $\mathcal{G}_{\mathcal{T}_1}$ mit dem Knoten (A, B) , die die Nicht-Existenz des LCS verursachen, und es gibt Zyklen wie in $\mathcal{G}_{\mathcal{T}_2}$ mit dem Knoten (C, C) , die „harmlos“ sind.

In dem Produktgraphen $\mathcal{G}_{\mathcal{T}_3}$ zu der TBox

$$\begin{aligned} \mathcal{T}_3 := \{ & A \sqsubseteq \exists r.A \sqcap \exists r.C, \\ & B \sqsubseteq \exists r.B \sqcap \exists r.C, \\ & C \sqsubseteq \exists r.C \} \end{aligned}$$

sind beide Zyklen aus $\mathcal{G}_{\mathcal{T}_1}$ und $\mathcal{G}_{\mathcal{T}_2}$ enthalten. $\exists r.C \sqcap \exists r.T$ ist der 1-lcs von A und B . Die Frage ist ob $\exists r.C \sqcap \exists r.T$ der exakte LCS ist und wie dieses Problem im allgemeinen Fall entschieden werden kann.

Dazu wird im nächsten Abschnitt eine Charakterisierung der Subsumptionsbeziehung als Grundlage für eine hinreichende und notwendige Bedingung für die Existenz des LCS vorgestellt.

3.2 Charakterisierung der Subsumption mit kanonischen Modellen und Simulationen

Für eine Charakterisierung des LCS ist zunächst eine geeignete Charakterisierung der Subsumption notwendig.

In [BKM99] wurde die Subsumptionsbeziehung zwischen \mathcal{EL} -Konzeptbeschreibungen bezüglich azyklischer \mathcal{EL} -TBoxen mit Hilfe

von Homomorphismen zwischen Beschreibungsbäumen charakterisiert. Für Subsumption bezüglich genereller \mathcal{EL} -TBoxen wurde in [LPW10] ein ähnlicher Zusammenhang mit Hilfe von kanonischen Modellen und Simulationen beschrieben. Die Ergebnisse dazu werden in diesem Abschnitt zusammengefasst.

Eine *Simulation* wird als Relation zwischen Interpretationen wie folgt definiert:

Definition 3.2.1 (Simulationen). Seien \mathcal{I}_1 und \mathcal{I}_2 Interpretationen und $\Sigma := N_C \cup N_R$ eine Signatur. Eine Relation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ ist eine Simulation zwischen \mathcal{I}_1 und \mathcal{I}_2 genau dann, wenn folgende Bedingungen erfüllt sind:

- (S1) Für alle Konzeptnamen $A \in N_C$ und alle $(e_1, e_2) \in \mathcal{S}$ gilt: Wenn $e_1 \in A^{\mathcal{I}_1}$, dann $e_2 \in A^{\mathcal{I}_2}$.
- (S2) Für alle Rollennamen $r \in N_R$, alle $(e_1, e_2) \in \mathcal{S}$ und alle $e'_1 \in \Delta^{\mathcal{I}_1}$ mit $(e_1, e'_1) \in r^{\mathcal{I}_1}$ existiert $e'_2 \in \Delta^{\mathcal{I}_2}$, so dass $(e_2, e'_2) \in r^{\mathcal{I}_2}$ und $(e'_1, e'_2) \in \mathcal{S}$.

Ein Tupel (\mathcal{I}, d) mit einer Interpretation \mathcal{I} und einem ausgezeichneten Element $d \in \Delta^{\mathcal{I}}$ wird als *punktierte Interpretation* bezeichnet. Seien (\mathcal{I}_1, d_1) und (\mathcal{I}_2, d_2) punktierte Interpretationen (\mathcal{I}_2, d_2) *simuliert* (\mathcal{I}_1, d_1) bezüglich einer Signatur Σ genau dann, wenn eine Simulation \mathcal{S} bezüglich Σ existiert mit $(d_1, d_2) \in \mathcal{S}$. Dies wird abgekürzt mit $(\mathcal{I}_1, d_1) \lesssim_{\Sigma} (\mathcal{I}_2, d_2)$. $(\mathcal{I}_1, d_1) \simeq_{\Sigma} (\mathcal{I}_2, d_2)$ wird als Abkürzung für $(\mathcal{I}_1, d_1) \lesssim_{\Sigma} (\mathcal{I}_2, d_2)$ und $(\mathcal{I}_2, d_2) \lesssim_{\Sigma} (\mathcal{I}_1, d_1)$ verwendet. In der Notation wird im Folgenden Σ meist weggelassen. Dann gilt $\Sigma = \text{sig}(\mathcal{T})$ für die betrachtete TBox.

Die \lesssim -Relation ist eine Quasiordnung auf einer Menge von Interpretationen der Form (\mathcal{I}, d) . Dies folgt direkt aus der Definition.

Mit Hilfe der \lesssim -Relation können Eigenschaften des Vervollständigungsgraphen beschrieben werden, indem der Graph als *kanonisches Modell* der zugehörigen TBox aufgefasst wird. Die Modelleigenschaft dieser Struktur wurde in [BBL05] für den Beweis der Vollständigkeit des Vervollständigungsgraphen genutzt. Das folgende Lemma beschreibt eine grundlegende Eigenschaft von kanonischen Modellen.

Lemma 3.2.1 (Kanonische Modelle [LW10]). *Sei \mathcal{T} eine \mathcal{EL} -TBox und C ein \mathcal{EL} -Konzept mit $\text{sig}(C) \subseteq \text{sig}(\mathcal{T})$. Dann existiert ein kanonisches Modell $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ mit $d_C \in C^{\mathcal{I}_{\mathcal{T}, C}}$ und für alle Modelle \mathcal{J} von \mathcal{T} mit $e \in \Delta^{\mathcal{J}}$ sind folgende Aussagen äquivalent:*

1. $e \in C^{\mathcal{J}}$
2. $(\mathcal{I}_{\mathcal{T}, C}, d_C) \lesssim (\mathcal{J}, e)$

Beweis. Das Lemma wurde in [LW10] bewiesen. Das kanonische Modell $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ ist eine punktierte Interpretation und $\mathcal{I}_{\mathcal{T}, C}$ ist Modell für \mathcal{T} mit

$d_C \in C^{\mathcal{I}_{\mathcal{T}, C}}$. In [LW10] wurde eine allgemeine Definition für die kanonischen Modelle angegeben. Im Folgenden werden Modelle mit Hilfe des Vervollständigungsgraphen aus Definition 2.1.4 definiert und gezeigt, dass die Eigenschaft für diese Modelle ebenfalls gilt.

Sei \mathcal{T}' die normalisierte TBox zu $\mathcal{T} \cup \{A \equiv C\}$ und $\mathcal{G}_{\mathcal{T}'} = (V, E, S)$ der zugehörige Vervollständigungsgraph. Das kanonische Modell $\mathcal{I}_{\mathcal{T}'}$ ist für alle Konzeptnamen B und alle Rollennamen r wie folgt definiert:

$$\Delta^{\mathcal{I}_{\mathcal{T}'}} := \{d_B \mid B \in V\} \quad (3.1)$$

$$B^{\mathcal{I}_{\mathcal{T}'}} := \{d_F \mid B \in S(F)\} \quad (3.2)$$

$$r^{\mathcal{I}_{\mathcal{T}'}} := \{(d_F, d_B) \mid (F, r, B) \in E\} \quad (3.3)$$

$\mathcal{I}_{\mathcal{T}'}$ ist ein Modell für \mathcal{T}' [BBL05] und auch ein Modell für \mathcal{T} und es gilt $d_A \in C^{\mathcal{I}'}$. Es wird gezeigt, dass $(\mathcal{I}_{\mathcal{T}'}, d_A)$ die geforderten Eigenschaften erfüllt.

1. \Rightarrow 2. : Sei \mathcal{J} ein beliebiges Modell für \mathcal{T} mit $e \in C^{\mathcal{J}}$. \mathcal{J} wird wie folgt zu einem Modell \mathcal{J}' für \mathcal{T}' erweitert:

\mathcal{T}' enthält den Konzeptnamen A und Normalisierungsnamen der Form $[\hat{D}]$ mit $\text{sig}(\hat{D}) \subseteq \text{sig}(\mathcal{T})$ (nach Anwendung der Regeln aus Abb. 2.1), die nicht in \mathcal{T} vorkommen. \mathcal{J}' wird definiert mit $A^{\mathcal{J}'} := C^{\mathcal{J}}$ und $[\hat{D}]^{\mathcal{J}'} := \hat{D}^{\mathcal{J}}$. Sonst stimmt \mathcal{J}' mit \mathcal{J} überein. Daher gilt

$$(\mathcal{J}', e) \lesssim_{\text{sig}(\mathcal{T})} (\mathcal{J}, e). \quad (3.4)$$

Es wird gezeigt, dass $(\mathcal{I}_{\mathcal{T}'}, d_A) \lesssim_{\text{sig}(\mathcal{T}')} (\mathcal{J}', e)$ gilt. Dazu wird gezeigt, dass die folgende Menge \mathcal{S} eine Simulation ist:

$$\mathcal{S} := \{(d_B, e') \mid B \in N_{C, \mathcal{T}'} \wedge e' \in B^{\mathcal{J}'}\} \quad (3.5)$$

Aus $e \in C^{\mathcal{J}}$ folgt $e \in A^{\mathcal{J}}$ nach Definition von \mathcal{J}' . Daher ist $(d_A, e) \in \mathcal{S}$. Sei $(d_F, f) \in \mathcal{S}$. Dann erfüllt (d_F, f) die Bedingungen aus Definition 3.2.1:

(S1): Sei $B \in N_{C, \mathcal{T}'}$ mit $d_F \in B^{\mathcal{I}_{\mathcal{T}'}}$. Dann folgt aus (3.2) $B \in S(F)$ und mit Definition 2.1.4 folgt $F \sqsubseteq_{\mathcal{T}'} B$. Wegen (3.5) gilt $f \in F^{\mathcal{J}'}$ und da \mathcal{J}' Modell von \mathcal{T}' ist, folgt $f \in B^{\mathcal{J}'}$.

(S2): Sei $r \in N_{R, \mathcal{T}'}$ und $(d_F, d_B) \in r^{\mathcal{I}_{\mathcal{T}'}}$. Es ist zu zeigen, dass ein f' existiert mit $(f, f') \in r^{\mathcal{J}'}$ und $(d_B, f') \in \mathcal{S}$. Aus $(d_F, d_B) \in r^{\mathcal{I}_{\mathcal{T}'}}$ folgt mit (3.3) und Definition 2.1.4 $F \sqsubseteq_{\mathcal{T}'} \exists r.B$. Mit $f \in F^{\mathcal{J}'}$ folgt $(f, f') \in r^{\mathcal{J}'}$ und $f' \in B^{\mathcal{J}'}$. Mit (3.5) folgt $(d_B, f') \in \mathcal{S}$.

Da $\text{sig}(\mathcal{T}) \subseteq \text{sig}(\mathcal{T}')$ gilt, folgt aus $(\mathcal{I}_{\mathcal{T}'}, d_A) \lesssim_{\text{sig}(\mathcal{T}')} (\mathcal{J}', e)$ auch $(\mathcal{I}_{\mathcal{T}'}, d_A) \lesssim_{\text{sig}(\mathcal{T})} (\mathcal{J}', e)$. Da \lesssim transitiv ist folgt mit (3.4) $(\mathcal{I}_{\mathcal{T}'}, d_A) \lesssim_{\text{sig}(\mathcal{T})} (\mathcal{J}, e)$.

2. \Rightarrow 1. : Für eine beliebige \mathcal{EL} -Konzeptbeschreibung C mit $\text{sig}(C) \subseteq \text{sig}(\mathcal{T})$ wird $(\mathcal{I}_{\mathcal{T}'}, d_A)$ wie oben definiert und mit $(\mathcal{I}_{\mathcal{T},C}, d_C) := (\mathcal{I}_{\mathcal{T}'}, d_A)$ bezeichnet.

Angenommen es gilt $(\mathcal{I}_{\mathcal{T},C}, d_C) \lesssim (\mathcal{J}, e)$. Es wird durch strukturelle Induktion über den Aufbau von C gezeigt, dass $e \in C^{\mathcal{J}}$ gilt.

$C = A$: Es gibt eine Simulation \mathcal{S} mit $(d_C, e) \in \mathcal{S}$. Mit $d_C \in A^{\mathcal{I}_{\mathcal{T},C}}$ und (S1) folgt $e \in A^{\mathcal{J}}$.

$C = F \sqcap D$: Aus $d_C \in F^{\mathcal{I}_{\mathcal{T},C}}$ und $d_C \in D^{\mathcal{I}_{\mathcal{T},C}}$ folgt mit 1. \Rightarrow 2. und der Annahme $(\mathcal{I}_{\mathcal{T},C}, d_C) \lesssim (\mathcal{J}, e)$:

$$(\mathcal{I}_{\mathcal{T},F}, d_F) \stackrel{1.\Rightarrow 2.}{\lesssim} (\mathcal{I}_{\mathcal{T},C}, d_C) \lesssim (\mathcal{J}, e)$$

$$(\mathcal{I}_{\mathcal{T},D}, d_D) \stackrel{1.\Rightarrow 2.}{\lesssim} (\mathcal{I}_{\mathcal{T},C}, d_C) \lesssim (\mathcal{J}, e)$$

Mit der Induktionshypothese und der Transitivität von \lesssim folgt dann $e \in F^{\mathcal{J}}$ und $e \in D^{\mathcal{J}}$ und damit $e \in C^{\mathcal{J}}$.

$C = \exists r.D$: Es gibt eine Simulation \mathcal{S} mit $(d_C, e) \in \mathcal{S}$. Es existiert ein $d'_D \in D^{\mathcal{I}_{\mathcal{T},C}}$ mit $(d_C, d'_D) \in r^{\mathcal{I}_{\mathcal{T},C}}$. Nach (S2) gibt es somit ein e' mit $(e, e') \in r^{\mathcal{J}}$ und

$$(d'_D, e') \in \mathcal{S}. \quad (3.6)$$

Aus 1. \Rightarrow 2. folgt:

$$(\mathcal{I}_{\mathcal{T},D}, d_D) \stackrel{1.\Rightarrow 2.}{\lesssim} (\mathcal{I}_{\mathcal{T},C}, d'_D) \stackrel{(3.6)}{\lesssim} (\mathcal{J}, e')$$

Die Induktionshypothese wird auf $(\mathcal{I}_{\mathcal{T},D}, d_D)$ angewendet und es folgt dann $e' \in D^{\mathcal{J}}$ und damit $e \in (\exists r.D)^{\mathcal{J}}$.

□

Zu der Definition der kanonischen Modelle im Beweis dieses Lemmas wird ein Beispiel betrachtet.

Beispiel 3.2.1. Es wird nochmals das Beispiel \mathcal{T}_3 aus Abschnitt 3.1 betrachtet. Der Produktgraph $\mathcal{G}_{\mathcal{T}_3}$ in Abbildung 3.1 entspricht nach Konstruktion der kanonischen Modelle in Lemma 3.2.1 dem Produkt der kanonischen Modelle $(\mathcal{I}_{\mathcal{T}_3,A}, d_A)$ und $(\mathcal{I}_{\mathcal{T}_3,B}, d_B)$. Das kanonische Modell $(\mathcal{I}_{\mathcal{T}_3, \exists r.C}, d_{\exists r.C})$ für die Konzeptbeschreibung $\exists r.C$ ergibt sich aus dem Vervollständigungsgraphen zu der TBox $\mathcal{T}_3 \cup \{[\exists r.C] \equiv \exists r.C\}$ und kann wie folgt skizziert werden:

$$d_{\exists r.C} : \{[\exists r.C], \top\} \xrightarrow{r} d_C : \{C, [\exists r.C], \top\}$$

$$\begin{array}{c} \cup \\ r \end{array}$$

Mit Hilfe der kanonischen Modelle lässt sich nun Subsumption zwischen \mathcal{EL} -Konzepten bezüglich einer generellen \mathcal{EL} -TBox charakterisieren. Das Lemma wurde in [LW10] bewiesen und ist eine Folgerung aus Lemma 3.2.1.

Lemma 3.2.2 (Charakterisierung der Subsumption [LW10]). *Sei \mathcal{T} eine \mathcal{EL} -TBox und C, D \mathcal{EL} -Konzeptbeschreibungen mit $\text{sig}(C) \cup \text{sig}(D) \subseteq \text{sig}(\mathcal{T})$. Es gilt $C \sqsubseteq_{\mathcal{T}} D$ genau dann, wenn $(\mathcal{I}_{\mathcal{T}, D}, d_D) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$.*

In diesem Abschnitt wurde gezeigt, dass basierend auf dem Vervollständigungsgraphen einer \mathcal{EL} -TBox \mathcal{T} einem \mathcal{EL} -Konzept C ein kanonisches Modell zugeordnet werden kann. Diese Struktur ist eine punktierte Interpretation der Form $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ mit einem Modell $\mathcal{I}_{\mathcal{T}, C}$ von \mathcal{T} und einem „Zeiger“ d_C für den $d_C \in C^{\mathcal{I}_{\mathcal{T}, C}}$ gilt. Mit Hilfe der Eigenschaften dieser Modelle und der Charakterisierung der Subsumption wird im folgenden Abschnitt eine Charakterisierung der Existenz des LCS entwickelt.

3.3 Charakterisierung der Existenz des LCS

In den folgenden Abschnitten wird der Fall betrachtet, dass eine normalisierte \mathcal{EL} -TBox \mathcal{T} und zwei Konzeptnamen A, B aus \mathcal{T} vorliegen. Dies ist nach Lemma 2.2.2 keine Einschränkung des allgemeinen Falls. In diesem Abschnitt wird eine notwendige und hinreichende Bedingung für die Existenz des LCS von A, B bezüglich \mathcal{T} angegeben.

Als ein Hilfsmittel wird dazu die bekannte Operation der *Baumabwicklung* einer Interpretation genutzt. Die Elemente der Domäne der Baumabwicklung sind dabei Wörter, die einen Pfad in der ursprünglichen Interpretation repräsentieren.

Definition 3.3.1 (Baumabwicklung einer Interpretation). Sei \mathcal{I} eine Interpretation mit $d \in \Delta^{\mathcal{I}}$. Die *Baumabwicklung* \mathcal{I}_d von \mathcal{I} in d ist wie folgt definiert:

$$\begin{aligned} \Delta^{\mathcal{I}_d} &:= \{dr_1d_1r_2\dots r_nd_n \mid (d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}} \wedge i \geq 0 \wedge d_0 = d\} \\ A^{\mathcal{I}_d} &:= \{\sigma d' \mid \sigma d' \in \Delta^{\mathcal{I}_d} \wedge d' \in A^{\mathcal{I}}\} \\ r^{\mathcal{I}_d} &:= \{(\sigma, \sigma r d') \mid (\sigma, \sigma r d') \in \Delta^{\mathcal{I}_d} \times \Delta^{\mathcal{I}_d}\} \end{aligned}$$

Die Länge eines Wortes $\sigma = dr_1d_1r_2\dots r_md_m$ aus $\Delta^{\mathcal{I}_d}$ ist gegeben durch:

$$|dr_1d_1r_2\dots r_md_m| := m$$

Die Interpretation \mathcal{I}_d^l bezeichnet die Beschränkung von \mathcal{I}_d auf Wörter der Länge $\leq l$ und wird *l-Beschränkung* von \mathcal{I}_d genannt. Das bedeutet, dass der Baum mit Wurzel d in der Tiefe l abgeschnitten wird. Die Baumabwicklung wird als punktierte Interpretation aufgefasst. An Stelle von (\mathcal{I}_d, d) wird im Folgenden nur \mathcal{I}_d geschrieben.

Es gilt die bekannte Beziehung zwischen einer Interpretation und ihrer Baumabwicklung.

Lemma 3.3.1. *Sei (\mathcal{I}, d) eine Interpretation und \mathcal{I}_d die Baumabwicklung von \mathcal{I} in d . Dann gilt $(\mathcal{I}, d) \simeq \mathcal{I}_d$.*

Beweis. $\mathcal{S} := \{(d', \sigma d') \mid d' \in \Delta^{\mathcal{I}}, \sigma d' \in \Delta^{\mathcal{I}_d}\}$ bzw. die inverse Relation sind entsprechende Simulationen. \square

Wie in Satz 2.2.1 gezeigt, entsteht der k -lcs(A, B) durch Traversierung des Produktgraphen bis zur Tiefe k . Daraus ergibt sich das folgende Lemma.

Lemma 3.3.2. *Sei C der k -lcs(A, B) bezüglich der \mathcal{EL} -TBox \mathcal{T} , $\mathcal{J}_{(d_A, d_B)}$ die Baumabwicklung von $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ und $\mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Produktgraph von \mathcal{T} . Es gilt*

1. $\mathcal{J}_{(d_A, d_B)}^k \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ und
2. $(d_A, d_B) \in (\text{trav}_{\mathcal{G}_{\mathcal{T}}}(k, (A, B)))^{\mathcal{J}_{(d_A, d_B)}^k}$

für die k -Beschränkung von $\mathcal{J}_{(d_A, d_B)}$.

Beweis. Zu 1.:

Es wird durch vollständige Induktion über die Rollentiefe $rt(C)$ gezeigt, dass eine Simulation \mathcal{S} existiert, die $\mathcal{J}_{(d_A, d_B)}^{rt(C)} \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ belegt. Dazu wird genutzt, dass der Produktgraph nach Definition 2.2.2 und das Produktmodell $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ mit der zugehörigen Baumabwicklung $\mathcal{J}_{(d_A, d_B)}$ nach Lemma 3.2.1 die gleiche Struktur haben.

$rt(C) = 0$: $\mathcal{S} := \{((d_A, d_B), d_C)\}$ ist eine Simulation. Da (d_A, d_B) keine Nachfolger in $\mathcal{J}_{(d_A, d_B)}^0$ hat, genügt es die Bedingung (S1) aus Definition 3.2.1 zu überprüfen. Es gilt $(d_A, d_B) \in D^{\mathcal{J}_{(d_A, d_B)}^0}$ gdw. $(d_A, d_B) \in D^{\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}}$ gdw. $D \in S((A, B))$ (nach Lemma 3.2.1) für einen Konzeptnamen D und für die Beschriftungsfunktion S im Produktgraphen von \mathcal{T} . Nach Satz 2.2.1 gilt $C \equiv_{\mathcal{T}} \text{trav}_{\mathcal{G}_{\mathcal{T}}}(0, (A, B)) = \prod_{D' \in S((A, B))} D'$ und damit folgt $d_C \in D^{\mathcal{I}_{\mathcal{T}, C}}$, da D in der Konjunktion $\prod_{D' \in S((A, B))} D'$ enthalten ist.

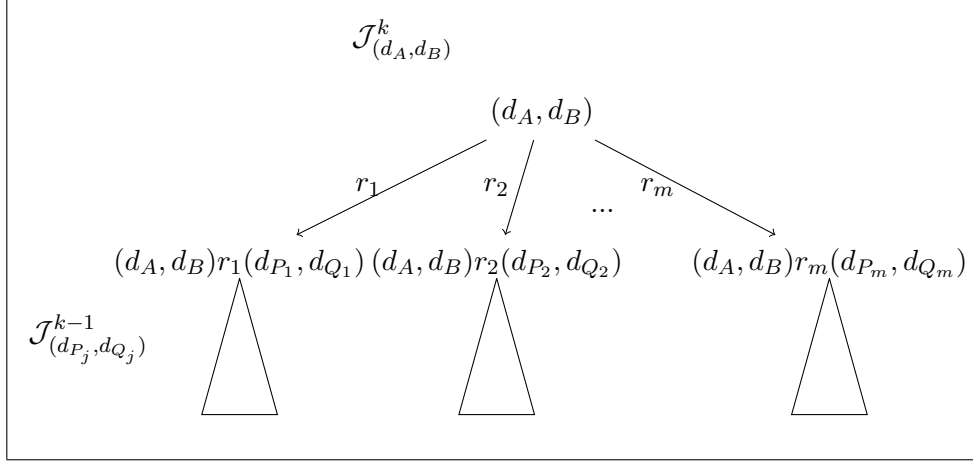
$rt(C) > 0$: Angenommen $rt(C) = k$. Es ist zu zeigen, dass $\mathcal{J}_{(d_A, d_B)}^k$ durch $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ simuliert wird.

Da C der k -lcs(A, B) bezüglich \mathcal{T} ist, gilt nach Satz 2.2.1:

$$C \equiv_{\mathcal{T}} D_1 \sqcap \dots \sqcap D_n \sqcap \exists r_1.K_1 \sqcap \dots \sqcap \exists r_m.K_m. \quad (3.7)$$

mit $D_i \in S((A, B))$ für alle $i = 1, \dots, n$ und

$$K_j = \text{trav}_{\mathcal{G}_{\mathcal{T}}}(k-1, (P_j, Q_j)) \text{ mit } ((A, B), r_j, (P_j, Q_j)) \in E \quad (3.8)$$


Abbildung 3.2: Skizze zum Beweis von Lemma 3.3.2

für alle $j = 1, \dots, m$. Mit (3.7) existiert für d_C für alle j ein $e_{K_j} \in \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ mit $(d_C, e_{K_j}) \in r^{\mathcal{I}_{\mathcal{T}, C}}$ und $e_{K_j} \in (K_j)^{\mathcal{I}_{\mathcal{T}, C}}$. Mit der Induktionshypothese, (3.8) und Lemma 3.2.1 folgt

$$\mathcal{J}_{(d_{P_j}, d_{Q_j})}^{k-1} \stackrel{\text{I.H.}}{\lesssim} (\mathcal{I}_{\mathcal{T}, K_j}, d_{K_j}) \stackrel{\text{L. 3.2.1}}{\lesssim} (\mathcal{I}_{\mathcal{T}, C}, e_{K_j}) \quad (3.9)$$

Demzufolge existiert eine Simulation $\mathcal{S}_j \subseteq \Delta^{\mathcal{J}_{(d_{P_j}, d_{Q_j})}^{k-1}} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ mit $((d_{P_j}, d_{Q_j}), e_{K_j}) \in \mathcal{S}_j$. Der Baum $\mathcal{J}_{(d_A, d_B)}^k$ hat die in Abb. 3.2 skizzierte Struktur. Aus (3.9) folgt, dass für alle als Dreiecke dargestellten Teilbäume mit den Wurzeln (d_{P_j}, d_{Q_j}) entsprechende r_j -Nachfolger e_{K_j} von d_C existieren, die diese Teilbäume simulieren. Die Simulation \mathcal{S} , die $\mathcal{J}_{(d_A, d_B)}^k \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ belegt, wird aus den zugehörigen Simulationen \mathcal{S}_j für die Teilbäume zusammengesetzt. Dabei wird berücksichtigt, dass die Knoten aus den Teilbäumen $\mathcal{J}_{(d_{P_j}, d_{Q_j})}^{k-1}$ in $\mathcal{J}_{(d_A, d_B)}^k$ das Präfix $(d_A, d_B)r_j$ haben. Die Simulation \mathcal{S} ergibt sich wie folgt:

$$\mathcal{S} := \{((d_A, d_B), d_C)\} \cup \bigcup_{j=1}^m \{((d_A, d_B)r_j\sigma, e) \mid (\sigma, e) \in \mathcal{S}_j\}.$$

Aus dem Gezeigten folgt, dass \mathcal{S} eine Simulation ist und damit $\mathcal{J}_{(d_A, d_B)}^k \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ folgt.

Zu 2.:

Ähnlich wie in 1. wird die Behauptung durch Induktion über k gezeigt.

$k = 0$: Nach Definition 2.2.3 gilt $\text{trav}_{G_{\mathcal{T}}}(0, (A, B)) = \prod_{D' \in S((A, B))} D'$. Aus Lemma 3.2.1 und Definition 3.3.1 folgt

$$(d_A, d_B) \in \left(\prod_{D \in S((A, B))} D \right)^{\mathcal{J}_{(d_A, d_B)}^0}.$$

$k > 0$: Nach Definition 2.2.3 gilt

$$\text{trav}_{\mathcal{G}_{\mathcal{T}}}(k, (A, B)) = \prod_{D \in S((A, B))} D \cap \prod_{((A, B), r, (P, Q)) \in E} \exists r. (\text{trav}_{\mathcal{G}_{\mathcal{T}}}(k-1, (P, Q))).$$

Aus Lemma 3.2.1 und Definition 3.3.1 folgt

$$(d_A, d_B) \in \left(\prod_{D \in S((A, B))} D \right)^{\mathcal{J}_{(d_A, d_B)}^k}$$

und mit der Induktionshypothese folgt

$$(d_P, d_Q) \in (\text{trav}_{\mathcal{G}_{\mathcal{T}}}(k-1, (P, Q)))^{\mathcal{J}_{(d_P, d_Q)}^{k-1}}.$$

Mit $((A, B), r, (P, Q)) \in E$ und Lemma 3.2.1 folgt $(d_A, d_B)r(d_P, d_Q) \in \Delta^{\mathcal{J}_{(d_A, d_B)}^k}$ und

$$(d_A, d_B)r(d_P, d_Q) \in (\exists r. \text{trav}_{\mathcal{G}_{\mathcal{T}}}(k-1, (P, Q)))^{\mathcal{J}_{(d_A, d_B)}^k}.$$

Damit folgt $(d_A, d_B) \in (\text{trav}_{\mathcal{G}_{\mathcal{T}}}(k, (A, B)))^{\mathcal{J}_{(d_A, d_B)}^k}$.

□

Das obige Lemma zeigt, dass der k -lcs(A, B) die k -Beschränkung des Baumes $\mathcal{J}_{(d_A, d_B)}$ simuliert. $\mathcal{J}_{(d_A, d_B)}$ ist ein möglicherweise unendlicher Baum, der äquivalent bezüglich \simeq zu dem Produktmodell $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ ist. Um den LCS zu charakterisieren, wird gezeigt, dass ein Common Subsumer C genau dann der exakte LCS von A und B bezüglich \mathcal{T} ist, wenn das kanonische Modell von C den gesamten Baum $\mathcal{J}_{(d_A, d_B)}$ simuliert. Dazu wird ein weiteres Lemma zu Produkten von punktierten Interpretationen benötigt, das in [LPW10] bewiesen wurde.

Lemma 3.3.3 ([LPW10]). *Seien (\mathcal{J}, e) , (\mathcal{I}_1, d_1) und (\mathcal{I}_2, d_2) beliebige punktierte Interpretationen. Wenn $(\mathcal{J}, e) \lesssim (\mathcal{I}_1, d_1)$ und $(\mathcal{J}, e) \lesssim (\mathcal{I}_2, d_2)$, dann $(\mathcal{J}, e) \lesssim (\mathcal{I}_1 \times \mathcal{I}_2, (d_1, d_2))$.*

Nun wird die folgende Charakterisierung des LCS bewiesen.

Satz 3.3.1 (Charakterisierung des LCS). *Sei \mathcal{T} eine \mathcal{EL} -TBox, $A, B \in N_{C, \mathcal{T}}$ und C ein \mathcal{EL} -Konzept mit $A \sqsubseteq_{\mathcal{T}} C$ und $B \sqsubseteq_{\mathcal{T}} C$. C ist genau dann der LCS von A und B bezüglich \mathcal{T} , wenn*

$$(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B)) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C).$$

Beweis. Für die „ \Rightarrow “-Richtung des Beweises wird eine Simulation konstruiert, die $\mathcal{J}_{(d_A, d_B)} \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ zeigt. Diese Simulation wird aus den Simulationen der k -Beschränkungen von $\mathcal{J}_{(d_A, d_B)}$ konstruiert, die nach Lemma 3.3.2 existieren. Im Beweis der „ \Leftarrow “-Richtung wird die Charakterisierung der Subsumption aus Lemma 3.2.2, und Lemma 3.3.3 angewendet.

\Rightarrow : Angenommen C ist der LCS von A und B bezüglich \mathcal{T} . Es wird gezeigt, dass eine Simulation zwischen der Baumabwicklung $\mathcal{J}_{(d_A, d_B)}$ von $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ und dem kanonischen Modell $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ existiert. Nach Lemma 3.3.1 existiert dann auch eine Simulation zwischen $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ und $(\mathcal{I}_{\mathcal{T}, C}, d_C)$. Zunächst kann $\mathcal{J}_{(d_A, d_B)}$ als Vereinigung der Teilbäume $\mathcal{J}_{(d_A, d_B)}^k$ geschrieben werden.

$$\Delta^{\mathcal{J}_{(d_A, d_B)}} = \bigcup_{k=0} \Delta^{\mathcal{J}_{(d_A, d_B)}^k}, \quad (3.10)$$

$$A^{\mathcal{J}_{(d_A, d_B)}} = \bigcup_{k=0} A^{\mathcal{J}_{(d_A, d_B)}^k}, \forall A \in N_{C, \mathcal{T}} \quad (3.11)$$

$$r^{\mathcal{J}_{(d_A, d_B)}} = \bigcup_{k=0} r^{\mathcal{J}_{(d_A, d_B)}^k}, \forall r \in N_{R, \mathcal{T}} \quad (3.12)$$

Sei $C_k = \text{k-lcs}(A, B)$. Aus Lemma 3.3.2 folgt

$$\mathcal{J}_{(d_A, d_B)}^k \lesssim (\mathcal{I}_{\mathcal{T}, C_k}, d_{C_k}) \quad (3.13)$$

Es gilt $C \sqsubseteq_{\mathcal{T}} C_k$ nach Definition des LCS und damit gilt nach Lemma 3.2.2: $(\mathcal{I}_{\mathcal{T}, C_k}, d_{C_k}) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$. Aus (3.13) und der Transitivität von \lesssim folgt damit $\mathcal{J}_{(d_A, d_B)}^k \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$.

Sei $\mathcal{S}_k \subseteq \Delta^{\mathcal{J}_{(d_A, d_B)}^k} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ die maximale Simulation. Wegen $\mathcal{J}_{(d_A, d_B)}^k \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ ist $((d_A, d_B), d_C) \in \mathcal{S}_k$. Für die Folge aller Teilbäume $\mathcal{J}_{(d_A, d_B)}^0, \mathcal{J}_{(d_A, d_B)}^1, \dots$ existiert eine Folge von maximalen Simulationen:

$$\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \dots \quad (3.14)$$

Mit dieser Folge wird gezeigt, dass eine Simulation \mathcal{S} existiert, die $\mathcal{J}_{(d_A, d_B)} \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ belegt.

Dazu wird ein beliebiges σ aus $\Delta^{\mathcal{J}_{(d_A, d_B)}}$ mit $|\sigma| \leq k'$ betrachtet. σ kommt in $\Delta^{\mathcal{J}_{(d_A, d_B)}^{k'}}$ vor. Daher gibt es nach (3.13) ein Element $d' \in \Delta^{\mathcal{I}_{\mathcal{T}, C}}$, das σ simuliert, d. h. $(\sigma, d') \in \mathcal{S}_{k'}$. Eine Menge \mathcal{S}_k^σ wird so definiert, dass genau diese Paare (σ, d') für ein festes σ und ein festes k in der Menge enthalten sind:

$$\mathcal{S}_k^\sigma := \{(\sigma', d') \in \Delta^{\mathcal{J}_{(d_A, d_B)}} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}} \mid \sigma' = \sigma \wedge (\sigma, d') \in \mathcal{S}_k\}. \quad (3.15)$$

Für alle $k \geq |\sigma|$ sind die \mathcal{S}_k^σ nicht-leer. Es gilt die folgende Behauptung:

Behauptung 1. Sei $l = |\sigma|$. Es gelten die folgenden Inklusionen:

$$\mathcal{S}_l^\sigma \supseteq \mathcal{S}_{l+1}^\sigma \supseteq \mathcal{S}_{l+2}^\sigma \dots$$

Beweis. Durch vollständige Induktion über alle $n \geq l$ wird gezeigt:

$$\mathcal{S}_n^\sigma \subseteq \mathcal{S}_{n-1}^\sigma \subseteq \dots \subseteq \mathcal{S}_{l+1}^\sigma \subseteq \mathcal{S}_l^\sigma \quad (3.16)$$

Der Induktionsanfang mit $n = l$ ist trivial.

Sei $n > l$ und $(\sigma, d) \in \mathcal{S}_n^\sigma$ und $\mathcal{S}_n \subseteq \Delta^{\mathcal{J}_{(d_A, d_B)}^n} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ die maximale Simulation (aus der Folge (3.14)). $\mathcal{S}_{n|_{n-1}}$ ist definiert durch

$$\mathcal{S}_{n|_{n-1}} := \mathcal{S}_n \cap (\Delta^{\mathcal{J}_{(d_A, d_B)}^{n-1}} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}})$$

als die Beschränkung von \mathcal{S}_n auf Wörter der Länge $\leq n - 1$. Sei \mathcal{S}_{n-1} die zugehörige maximale Simulation aus der Folge (3.14). Da $\mathcal{S}_{n|_{n-1}}$ auch eine Simulation ist, die $\mathcal{J}_{(d_A, d_B)}^{n-1} \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$ belegt und \mathcal{S}_{n-1} maximal ist, gilt $\mathcal{S}_{n|_{n-1}} \subseteq \mathcal{S}_{n-1}$.

Es gilt $(\sigma, d) \in \mathcal{S}_{n|_{n-1}}$, da $|\sigma| < n$. Daraus folgt $(\sigma, d) \in \mathcal{S}_{n-1}$ und mit (3.15) folgt $(\sigma, d) \in \mathcal{S}_{n-1}^\sigma$. Daraus folgt mit der Induktionshypothese:

$$\mathcal{S}_n^\sigma \subseteq \mathcal{S}_{n-1}^\sigma \stackrel{\text{I.H.}}{\subseteq} \dots \stackrel{\text{I.H.}}{\subseteq} \mathcal{S}_{l+1}^\sigma \stackrel{\text{I.H.}}{\subseteq} \mathcal{S}_l^\sigma$$

□

Aus der Behauptung folgt, dass für ein festes $\sigma \in \Delta^{\mathcal{J}_{(d_A, d_B)}}$ die Menge der Elemente aus $\Delta^{\mathcal{I}_{\mathcal{T}, C}}$, die σ in der Folge (3.14) simulieren, mit wachsendem k kleiner wird. Für alle σ existiert daher ein Index l' , so dass gilt:

$$\mathcal{S}_{l'}^\sigma = \bigcap_{l \geq |\sigma|} \mathcal{S}_l^\sigma \quad (3.17)$$

l' wird als Grenzwert von σ bezeichnet. $\mathcal{S} \subseteq \Delta^{\mathcal{J}_{(d_A, d_B)}} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ wird nun wie folgt definiert:

$$\mathcal{S} := \bigcup_{\sigma \in \Delta^{\mathcal{J}_{(d_A, d_B)}}} \bigcap_{l \geq |\sigma|} \mathcal{S}_l^\sigma$$

Es ist zu zeigen, dass \mathcal{S} eine Simulation mit $((d_A, d_B), d_C) \in \mathcal{S}$ ist. Für alle k gilt $((d_A, d_B), d_C) \in \mathcal{S}_k^{(d_A, d_B)}$ und damit auch $((d_A, d_B), d_C) \in \mathcal{S}$.

(S1) Sei $(\sigma, d) \in \mathcal{S}$ und $\sigma \in D^{\mathcal{J}_{(d_A, d_B)}}$ für einen Konzeptnamen D . Es existiert ein k mit $(\sigma, d) \in \mathcal{S}_k$. Aus $\sigma \in D^{\mathcal{J}_{(d_A, d_B)}}$ folgt mit (3.11) auch $\sigma \in D^{\mathcal{J}_{(d_A, d_B)}^k}$ und $d \in D^{\mathcal{I}_{\mathcal{T}, C}}$, weil \mathcal{S}_k eine Simulation ist.

(S2) Sei $(\sigma, d) \in \mathcal{S}$ und $(\sigma, \sigma re) \in r^{\mathcal{J}(d_A, d_B)}$. Sei l das Maximum der Grenzwerte von σ und σre . Mit (3.17) folgt $\mathcal{S}_l^\sigma = \bigcap_{i \geq |\sigma|} \mathcal{S}_i^\sigma$ und $\mathcal{S}_l^{\sigma re} = \bigcap_{j \geq |\sigma re|} \mathcal{S}_j^{\sigma re}$. Daraus folgt $\mathcal{S}_l^\sigma \subseteq \mathcal{S}$ und $\mathcal{S}_l^{\sigma re} \subseteq \mathcal{S}$ nach Definition von \mathcal{S} . Da $(\sigma, d) \in \mathcal{S}_l$ und $(\sigma, \sigma re) \in r^{\mathcal{J}^l(d_A, d_B)}$ (wegen (3.12)), existiert ein d' mit $(d, d') \in r^{\mathcal{I}_{\mathcal{T}, C}}$ und $(\sigma re, d') \in \mathcal{S}_l$, weil \mathcal{S}_l eine Simulation ist. Mit $\mathcal{S}_l^{\sigma re} \subseteq \mathcal{S}$ folgt dann $(\sigma re, d') \in \mathcal{S}$.

\Leftarrow : Angenommen es gilt $A \sqsubseteq_{\mathcal{T}} C$, $B \sqsubseteq_{\mathcal{T}} C$ und

$$(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B)) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C). \quad (3.18)$$

Es ist zu zeigen, dass C der LCS von A und B ist. Sei E ein Konzept mit $A \sqsubseteq_{\mathcal{T}} E$ und $B \sqsubseteq_{\mathcal{T}} E$ mit Lemma 3.2.2 folgt $(\mathcal{I}_{\mathcal{T}, E}, d_E) \lesssim (\mathcal{I}_{\mathcal{T}, A}, d_A)$ und $(\mathcal{I}_{\mathcal{T}, E}, d_E) \lesssim (\mathcal{I}_{\mathcal{T}, B}, d_B)$. Daraus folgt mit Lemma 3.3.3 $(\mathcal{I}_{\mathcal{T}, E}, d_E) \lesssim (\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$. Aus (3.18) und der Transitivität von \lesssim folgt $(\mathcal{I}_{\mathcal{T}, E}, d_E) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$. Mit Lemma 3.2.2 folgt $C \sqsubseteq_{\mathcal{T}} E$. □

In diesem Abschnitt wurde gezeigt, dass ein Common Subsumer C von A und B genau dann der LCS von A und B bezüglich \mathcal{T} ist, wenn das zugehörige kanonische Modell $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ das Produktmodell $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ simuliert. Für den Rollentiefen- beschränkten LCS kann jetzt getestet werden ob diese Approximation der exakte LCS ist. Im nächsten Abschnitt wird der in Satz 3.3.1 gezeigte Zusammenhang genutzt, um zu entscheiden, ob der LCS existiert oder nicht.

3.4 Entscheidbarkeit des Existenzproblems

Die Frage nach der Existenz des LCS kann, wie im letzten Abschnitt bewiesen, auf die Frage reduziert werden ob eine natürliche Zahl k existiert, so dass das kanonische Modell des $k\text{-lcs}(A, B)$ das Produktmodell $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ simuliert.

Zunächst wird gezeigt, dass das Produktmodell vereinfacht werden kann. Das Modell $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ entspricht wie in Lemma 3.2.1 gezeigt dem Produktgraphen nach Definition 2.2.2, da \mathcal{T} normalisiert ist und A, B Konzeptnamen aus \mathcal{T} sind. Es wird eine Unterscheidung der Knoten des Produktgraphen bzw. der Elemente von $\Delta^{\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}}$ in *synchrone* Elemente und *asynchrone* Elemente bzw. Knoten eingeführt. Ein Knoten (P, Q) des Produktgraphen bzw. das Element (d_P, d_Q) wird als *synchron* bezeichnet, wenn $P = Q$ gilt. Alle anderen Knoten bzw. Elemente werden als *asynchrone* Knoten bzw. Elemente bezeichnet.

Im Beispiel $\mathcal{G}_{\mathcal{T}_3}$ in Abbildung 3.1 ist demzufolge (C, C) ein synchroner Knoten und (A, B) ist asynchron. Es wird gezeigt, dass das Produktmodell so reduziert werden kann, dass ein synchrones Element nur Nachfolger hat die synchron sind.

Definition 3.4.1 (Reduziertes Produktmodell). Sei $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ das Produktmodell. Sei $\Delta^{cut} \subseteq \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$. Es gilt $(d_P, d_Q) \in \Delta^{cut}$ genau dann, wenn

1. (d_P, d_Q) asynchron ist und
2. ein Element $(d_F, d_F) \in \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$ existiert, so dass (d_F, d_F) synchron ist und (d_P, d_Q) von (d_F, d_F) durch einen Pfad in $\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}$ erreichbar ist.

Das reduzierte Produktmodell $(\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}, (d_A, d_B))$ wird wie folgt definiert:

$$\begin{aligned} \Delta^{\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}} &:= \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \setminus \Delta^{cut} \\ A^{\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}} &:= A^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \cap \Delta^{\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}} \\ r^{\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}} &:= r^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \cap (\Delta^{\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}} \times \Delta^{\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}}) \end{aligned}$$

Die zugehörige Baumabwicklung zu $(\mathcal{I}'_{\mathcal{T},A} \times \mathcal{I}'_{\mathcal{T},B}, (d_A, d_B))$ wird mit $\mathcal{J}_{(d_A, d_B)}^{cut}$ bezeichnet.

Im nächsten Lemma wird gezeigt, dass das Produktmodell durch das reduzierte Produktmodell simuliert werden kann. Dies folgt aus den Eigenschaften des Vervollständigungsgraphen und des Produktgraphen.

Lemma 3.4.1. *Sei $\mathcal{J}_{(d_A, d_B)}$ die Baumabwicklung des Produktmodells und $\mathcal{J}_{(d_A, d_B)}^{cut}$ das zugehörige reduzierte Baummodell nach Definition 3.4.1. Es gilt $\mathcal{J}_{(d_A, d_B)} \simeq \mathcal{J}_{(d_A, d_B)}^{cut}$.*

Beweis. Es wird gezeigt, dass eine Simulation \mathcal{S} für $\mathcal{J}_{(d_A, d_B)} \lesssim \mathcal{J}_{(d_A, d_B)}^{cut}$ existiert. Sei $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph zu \mathcal{T} und sei $\sigma \in \Delta^{\mathcal{J}_{(d_A, d_B)}} \setminus \Delta^{\mathcal{J}_{(d_A, d_B)}^{cut}}$. Dann ist σ von der Form:

$$\sigma = \sigma'(d_F, d_F)r_1(d_{P_1}, d_{Q_1})r_2 \dots r_n(d_{P_n}, d_{Q_n})$$

und σ' enthält nur asynchrone Elemente. Dabei ist (d_F, d_F) synchron und (d_{P_n}, d_{Q_n}) asynchron. Es gilt $(F, r_1, P_1) \in E$, $(F, r_1, Q_1) \in E$, $(P_{i-1}r_i, P_i) \in E$ und $(Q_{i-1}r_i, Q_i) \in E$ für alle $i = 2, \dots, n$ nach Definition des Produktmodells. Dann existieren die Wörter $\sigma_P, \sigma_Q \in \Delta^{\mathcal{J}_{(d_A, d_B)}^{cut}}$ mit

$$\begin{aligned} \sigma_P &= \sigma'(d_F, d_F)r_1(d_{P_1}, d_{P_1})r_2 \dots r_n(d_{P_n}, d_{P_n}) \\ \sigma_Q &= \sigma'(d_F, d_F)r_1(d_{Q_1}, d_{Q_1})r_2 \dots r_n(d_{Q_n}, d_{Q_n}) \end{aligned}$$

und σ wird durch σ_P und σ_Q simuliert, da für die Beschriftungsfunktion $S_{\mathcal{G} \times \mathcal{G}}$ des Produktgraphen gilt:

$$S_{\mathcal{G} \times \mathcal{G}}(P_i, Q_i) = S(P_i) \cap S(Q_i) = S_{\mathcal{G} \times \mathcal{G}}(P_i, P_i) \cap S_{\mathcal{G} \times \mathcal{G}}(Q_i, Q_i).$$

Jedem Element aus $\Delta^{\mathcal{J}(d_A, d_B)} \setminus \Delta^{\mathcal{J}^{cut}(d_A, d_B)}$ kann daher ein simulierendes Element aus $\Delta^{\mathcal{J}^{cut}(d_A, d_B)}$ zugeordnet werden. Daraus folgt $\mathcal{J}(d_A, d_B) \lesssim \mathcal{J}^{cut}(d_A, d_B)$. Die andere Richtung folgt direkt, weil $\mathcal{J}^{cut}(d_A, d_B)$ ein Teilbaum von $\mathcal{J}(d_A, d_B)$ ist. \square

Im Folgenden wird davon ausgegangen, dass das Produktmodell $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$ nach Definition 3.4.1 reduziert ist.

Im nächsten Schritt wird die Struktur des kanonischen Modells des k -lcs(A, B) genauer untersucht. Dazu wird zunächst die k -Beschränkung $\mathcal{J}^k_{(d_A, d_B)}$ aus Lemma 3.3.2 zu einem Modell von \mathcal{T} erweitert, indem synchrone Elemente hinzugefügt werden.

Definition 3.4.2 (Kanonisches Modell des k -lcs). Sei $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph zu \mathcal{T} und $\mathcal{J}^k_{(d_A, d_B)}$ die k -Beschränkung der Baumabwicklung von $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$. Die Erweiterung $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ von $\mathcal{J}^k_{(d_A, d_B)}$ wird wie folgt definiert:

$\Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$ ist die kleinste Menge, die die folgenden Bedingungen erfüllt:

1. $\Delta^{\mathcal{J}^k_{(d_A, d_B)}} \subseteq \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$
2. $\widehat{\sigma} = \sigma d_0 r_1 d_1 r_2 \dots r_n d_n \in \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$ genau dann, wenn folgende Bedingungen gelten:
 - (a) $\widehat{\sigma} \in \Delta^{\mathcal{J}(d_A, d_B)}$ und $\sigma d_0 \in \Delta^{\mathcal{J}^k_{(d_A, d_B)}}$;
 - (b) Für alle $i = 1, \dots, n$ ist d_i ein synchrones Element;
 - (c) Es existiert ein Konzeptname D mit $\sigma d_0 \in D^{\mathcal{J}^k_{(d_A, d_B)}}$, so dass $(D, r_1, F) \in E$ und $d_1 = (d_F, d_F)$ gilt.

Für alle Konzeptnamen A und Rollennamen r gilt

$$\begin{aligned} A^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} &:= A^{\mathcal{J}(d_A, d_B)} \cap \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} \text{ und} \\ r^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} &:= r^{\mathcal{J}(d_A, d_B)} \cap (\Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} \times \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}). \end{aligned}$$

Es wird gezeigt, dass $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ ein Modell von \mathcal{T} ist.

Lemma 3.4.2. $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ ist ein Modell von \mathcal{T} .

Beweis. Sei $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph zu \mathcal{T} und $\mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}} = (V_{\mathcal{G} \times \mathcal{G}}, E_{\mathcal{G} \times \mathcal{G}}, S_{\mathcal{G} \times \mathcal{G}})$ der Produktgraph. Es ist zu zeigen, dass $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ alle Inklusionen aus \mathcal{T} (in Normalform) erfüllt. Es gilt folgender Zusammenhang nach Definition 3.4.2 für alle $\sigma \in \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$:

$$\sigma \in D^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} \Leftrightarrow \sigma \in D^{\mathcal{J}_{(d_A, d_B)}} \quad (3.19)$$

für alle Konzeptnamen D aus \mathcal{T} . Da $\mathcal{J}_{(d_A, d_B)}$ nach Definition 2.1.4, Lemma 2.1.4 und Definition 2.2.2 ein Modell für \mathcal{T} ist, erfüllt $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ alle Inklusionen der Form $D \sqsubseteq F \in \mathcal{T}$. Für alle Inklusionen $\exists r. D \sqsubseteq F \in \mathcal{T}$ gilt

$$\sigma \in (\exists r. D)^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} \Rightarrow \sigma \in (\exists r. D)^{\mathcal{J}_{(d_A, d_B)}} \Rightarrow \sigma \in F^{\mathcal{J}_{(d_A, d_B)}}$$

und mit (3.19) folgt $\sigma \in F^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$. Es bleibt zu zeigen, dass alle Inklusionen der Form $D \sqsubseteq \exists r. F$ ebenfalls erfüllt sind. Sei $\rho d_0 \in D^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$ und $d_0 = (d_Q, d_{Q'})$. Dann ist $D \in S_{\mathcal{G} \times \mathcal{G}}((Q, Q'))$ und $(Q, r, F), (Q', r, F) \in E$. Daraus folgt $((Q, Q'), r, (F, F)) \in E_{\mathcal{G} \times \mathcal{G}}$. Für $\widehat{\rho} = \rho d_0 r d_1$ mit $d_1 = (d_F, d_F)$ gilt dann $\widehat{\rho} \in \Delta^{\mathcal{J}_{(d_A, d_B)}}$ und entsprechend $(\rho d_0, \widehat{\rho}) \in r^{\mathcal{J}_{(d_A, d_B)}}$ und $\widehat{\rho} \in F^{\mathcal{J}_{(d_A, d_B)}}$, da $F \in S_{\mathcal{G} \times \mathcal{G}}((F, F))$. Da $\widehat{\rho}$ die Bedingungen aus der Definition von $\Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$ erfüllt, gilt auch $(\rho d_0, \widehat{\rho}) \in r^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$ und $\widehat{\rho} \in F^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$. Daraus folgt dann $\rho d_0 \in (\exists r. F)^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$. \square

Als nächstes wird gezeigt, dass $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ äquivalent bezüglich \simeq zum kanonischen Modell des k -lcs(A, B) ist.

Lemma 3.4.3. *Sei $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ das kanonische Modell des k -lcs(A, B) bezüglich \mathcal{T} und $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ die Erweiterung der k -Beschränkung von $\mathcal{J}_{(d_A, d_B)}$. Es gilt*

$$\widehat{\mathcal{J}}^k_{(d_A, d_B)} \simeq (\mathcal{I}_{\mathcal{T}, C}, d_C)$$

Beweis. Sei $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph zu \mathcal{T} und $\mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}} = (V_{\mathcal{G} \times \mathcal{G}}, E_{\mathcal{G} \times \mathcal{G}}, S_{\mathcal{G} \times \mathcal{G}})$ der Produktgraph.

Es wird zunächst gezeigt, dass eine Simulation von $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ durch $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ existiert. Nach Lemma 3.3.1 existiert eine Simulation $\mathcal{S} \subseteq \Delta^{\mathcal{J}^k_{(d_A, d_B)}} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ mit $((d_A, d_B), d_C) \in \mathcal{S}$. Diese wird zu einer Simulation $\widehat{\mathcal{S}} \subseteq \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ erweitert.

Dazu wird ein Element $\sigma_{pre} \in \Delta^{\mathcal{J}^k_{(d_A, d_B)}}$ der Länge k betrachtet mit dem simulierenden Element $d_{pre} \in \Delta^{\mathcal{I}_{\mathcal{T}, C}}$, d. h. es gilt $(\sigma_{pre}, d_{pre}) \in \mathcal{S}$. Angenommen es existiert in $\Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$ ein Element $\sigma_{succ} = \sigma_{pre} r d'$. Nach Konstruktion von $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ ist d' ein synchrones Element. In der folgenden Behauptung wird gezeigt, dass dann auch ein $d_{succ} \in \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ als r -Nachfolger von d_{pre} existiert, so dass σ_{succ} durch d_{succ} simuliert wird.

Behauptung 1. Sei $\sigma_{succ} \in \Delta^{\widehat{\mathcal{J}}^k(d_A, d_B)} \setminus \Delta^{\mathcal{J}^k(d_A, d_B)}$ mit $\sigma_{succ} = \sigma_{pre} r d'$ und $\sigma_{pre} \in \Delta^{\mathcal{J}^k(d_A, d_B)}$ und $d_{pre} \in \Delta^{\mathcal{I}_{\mathcal{T}, C}}$ mit $(\sigma_{pre}, d_{pre}) \in \mathcal{S}$. Es existiert ein d_{succ} mit $(d_{pre}, d_{succ}) \in r^{\mathcal{I}_{\mathcal{T}, C}}$, so dass $(\widehat{\mathcal{J}}^k(d_A, d_B), \sigma_{succ}) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_{succ})$ gilt.

Beweis. Nach Definition 3.4.2 2. (c) gibt es einen Konzeptnamen D mit $\sigma_{pre} \in D^{\mathcal{J}^k(d_A, d_B)}$ mit $(D, r, K) \in E$ und $d' = (d_K, d_K)$. Da $(\sigma_{pre}, d_{pre}) \in \mathcal{S}$, gilt $d_{pre} \in D^{\mathcal{I}_{\mathcal{T}, C}}$ und $d_{pre} \in (\exists r.K)^{\mathcal{I}_{\mathcal{T}, C}}$, da $\mathcal{I}_{\mathcal{T}, C}$ ein Modell von \mathcal{T} ist. Daher existiert ein d_{succ} mit $(d_{pre}, d_{succ}) \in r^{\mathcal{I}_{\mathcal{T}, C}}$ und $d_{succ} \in K^{\mathcal{I}_{\mathcal{T}, C}}$. Nach Definition 3.4.2 gilt, dass σ_{succ} und alle Nachfolger von σ_{succ} in $\widehat{\mathcal{J}}^k(d_A, d_B)$ als letztes Element ein synchrones Element haben. Eine Menge $\mathcal{S}_{\sigma_{succ}}$ wird daher wie folgt definiert:

$$\mathcal{S}_{\sigma_{succ}} := \{(\sigma_{succ} r_0 d_0 \dots r_n d_n, t) \in \Delta^{\widehat{\mathcal{J}}^k(d_A, d_B)} \times \Delta^{\mathcal{I}_{\mathcal{T}, C}} \mid n \geq 0 \\ d_n = (d_P, d_P) \wedge t \in P^{\mathcal{I}_{\mathcal{T}, C}}\} \cup \{(\sigma_{succ}, d_{succ})\}$$

Es wird gezeigt, dass $\mathcal{S}_{\sigma_{succ}}$ die Bedingungen (S1) und (S2) erfüllt.

(S1): Sei $(\rho g, h) \in \mathcal{S}_{\sigma_{succ}}$. Es gilt $g = (d_Q, d_Q)$ für einen Konzeptnamen Q . Es gelten die folgenden Äquivalenzen:

$$\rho g \in Q^{\widehat{\mathcal{J}}^k(d_A, d_B)} \Leftrightarrow Q' \in S_{\mathcal{G} \times \mathcal{G}}((Q, Q)) \Leftrightarrow Q' \in S(Q)$$

Da $h \in Q^{\mathcal{I}_{\mathcal{T}, C}}$ nach Definition von $\mathcal{S}_{\sigma_{succ}}$ und $\mathcal{I}_{\mathcal{T}, C}$ Modell ist, folgt $h \in Q^{\mathcal{I}_{\mathcal{T}, C}}$.

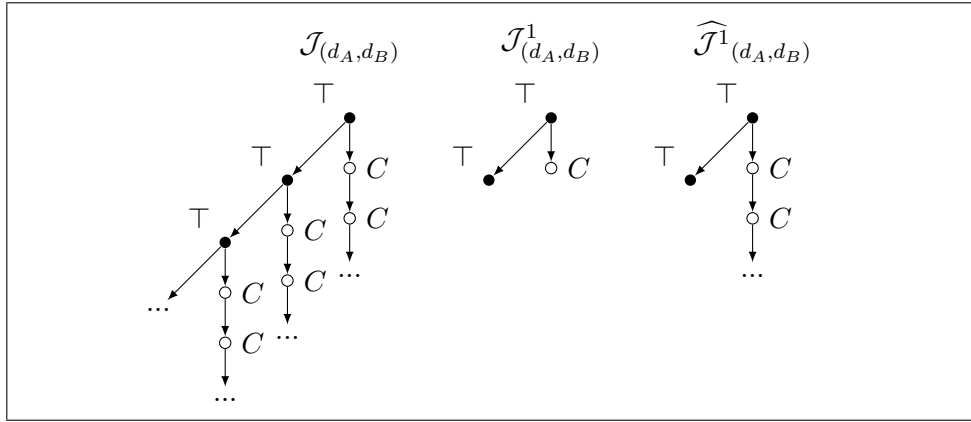
(S2): Sei $(\rho g, h) \in \mathcal{S}_{\sigma_{succ}}$ und $(\rho g, \rho g r g') \in r^{\widehat{\mathcal{J}}^k(d_A, d_B)}$ mit $g = (d_Q, d_Q)$ und $g' = (d_{Q'}, d_{Q'})$. Dann gibt es die Kanten $((Q, Q), r, (Q', Q')) \in E_{\mathcal{G} \times \mathcal{G}}$ und $(Q, r, Q') \in E$. Mit $h \in Q^{\mathcal{I}_{\mathcal{T}, C}}$ folgt $h \in (\exists r.Q')^{\mathcal{I}_{\mathcal{T}, C}}$. Es existiert also ein h' mit $(h, h') \in r^{\mathcal{I}_{\mathcal{T}, C}}$ und $h' \in Q'^{\mathcal{I}_{\mathcal{T}, C}}$. Daraus folgt $(\rho g r g', h') \in \mathcal{S}_{\sigma_{succ}}$.

□

Sei $\Delta_{succ} := \{\sigma d r d' \mid \sigma d r d' \in \Delta^{\widehat{\mathcal{J}}^k(d_A, d_B)} \setminus \Delta^{\mathcal{J}^k(d_A, d_B)} \wedge \sigma d \in \Delta^{\mathcal{J}^k(d_A, d_B)}\}$. Die Menge der Elemente aus Δ_{succ} erfüllen die Bedingungen aus Behauptung 1. Alle Elemente die neu zu $\widehat{\mathcal{J}}^k(d_A, d_B)$ hinzugefügt liegen in einem Teilbaum von $\widehat{\mathcal{J}}^k(d_A, d_B)$ mit einer Wurzel aus Δ_{succ} . Für diesen Teilbaum wurde in Behauptung 1 eine entsprechende Simulation konstruiert. Die Simulation $\widehat{\mathcal{S}}$ kann also wie folgt zusammengesetzt werden:

$$\widehat{\mathcal{S}} := \mathcal{S} \cup \bigcup_{\sigma \in \Delta_{succ}} \mathcal{S}_{\sigma}$$

Nach Konstruktion folgt damit $\widehat{\mathcal{J}}^k(d_A, d_B) \lesssim (\mathcal{I}_{\mathcal{T}, C}, d_C)$.


 Abbildung 3.3: Baumabwicklungen von $\mathcal{G}_{\mathcal{T}_3}$

Nach Satz 2.2.1 folgt $C \equiv_{\mathcal{T}} \text{trav}_{\mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}}}(k, (A, B))$. Mit Lemma 3.3.2 folgt $(d_A, d_B) \in (\text{trav}_{\mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}}}(k, (A, B)))^{\mathcal{J}^k_{(d_A, d_B)}}$ und damit auch $(d_A, d_B) \in (\text{trav}_{\mathcal{G}_{\mathcal{T}} \times \mathcal{G}_{\mathcal{T}}}(k, (A, B)))^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}}$. Nach Lemma 3.4.2 ist $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ Modell von \mathcal{T} . Aus Lemma 3.2.1 folgt damit $(\mathcal{I}_{\mathcal{T}, C}, d_C) \lesssim \widehat{\mathcal{J}}^k_{(d_A, d_B)}$. \square

Aus diesem Lemma folgt, dass das kanonische Modell des k -lcs als spezielle Abwicklung des Produktmodells aufgefasst werden kann. Dazu wird ein Beispiel betrachtet.

Beispiel 3.4.1. Das Beispiel \mathcal{T}_3 aus Abschnitt 3.1 wird fortgesetzt. In Abb. 3.3 ist mit $\mathcal{J}_{(d_A, d_B)}$ die Baumabwicklung von $\mathcal{G}_{\mathcal{T}_3}$ aus Abb. 3.1 skizziert. Die schwarzen Knoten sind die Elemente mit (d_A, d_B) als letztem Element. Die weißen Knoten haben das synchrone Element (d_C, d_C) an letzter Position. Daneben ist die 1-Beschränkung des Baummodells $\mathcal{J}^1_{(d_A, d_B)}$ dargestellt. Da $C \sqsubseteq \exists r. C \in \mathcal{T}_3$ und die weißen Knoten mit C beschriftet sind, wird $\mathcal{J}^1_{(d_A, d_B)}$ um unendliche viele (d_C, d_C) -Knoten erweitert. Der Pfad mit dem schwarzen Knoten wird hingegen nicht erweitert. Daraus ergibt sich die Skizze des Modells $\widehat{\mathcal{J}}^1_{(d_A, d_B)}$. Da die schwarzen Knoten durch die weißen Knoten simuliert werden können, wird $\mathcal{J}_{(d_A, d_B)}$ durch $\widehat{\mathcal{J}}^1_{(d_A, d_B)}$ simuliert. In Lemma 3.4.3 wurde gezeigt, dass $\widehat{\mathcal{J}}^1_{(d_A, d_B)}$ äquivalent zum kanonischen Modell des 1-lcs (A, B) ist. Aus dem bisher Gezeigten folgt damit, dass der 1-lcs $(A, B) = \exists r. \top \sqcap \exists r. C$ der exakte LCS ist.

Wenn der LCS von Konzeptnamen A, B bezüglich \mathcal{T} existiert wie in Beispiel 3.4.1, dann gibt es ein k , so dass das Produktmodell $\mathcal{J}_{(d_A, d_B)}$ durch $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ simuliert wird. Beide Baummodelle können jedoch unendliche Pfade enthalten. Diese unendlichen Pfade entstehen durch Zyklen in $(\mathcal{I}_{\mathcal{T}, A} \times \mathcal{I}_{\mathcal{T}, B}, (d_A, d_B))$. Enthält $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ einen unendlichen Pfad,

dann gibt es in $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ eine Zyklus mit synchronen Elementen. In $\mathcal{J}_{(d_A, d_B)}$ kann es zusätzlich auch unendliche Pfade mit nur asynchronen Elementen geben, wenn es einen entsprechenden Zyklus in $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ gibt. Vereinfacht dargestellt existiert ein k , so dass $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ das Modell $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ bzw. $\mathcal{J}_{(d_A, d_B)}$ simuliert, wenn die Zyklen in $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ mit asynchronen Elementen durch Zyklen mit synchronen Elementen simuliert werden. Um diese spezielle Eigenschaft einer Simulation zu beschreiben, wird der Begriff eines *Simulationsgraphen* eingeführt. In einem Simulationsgraphen werden von einem Startelement ausgehend die Paare der Simulation zu einem Graphen verbunden. Ein Simulationsgraph ist wie folgt definiert:

Definition 3.4.3 (Simulationsgraph). Seien \mathcal{I}_1 und \mathcal{I}_2 Interpretationen und $\mathcal{S} \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ eine Simulation bezüglich der Signatur $\Sigma = N_C \cup N_R$. Ein *Simulationsgraph* $\mathcal{G}_{sim} = (\mathcal{S}, E_{sim}, I)$ besteht aus der Knotenmenge \mathcal{S} , den Kanten $E_{sim} \subseteq \mathcal{S} \times N_R \times \mathcal{S}$ und dem Startknoten $I \in \mathcal{S}$. Sei $I = (d_0, e_0)$. E_{sim} erfüllt die folgenden Bedingungen:

1. Für alle $d \in \Delta^{\mathcal{I}_1}$, die von d_0 in \mathcal{I}_1 erreichbar sind, existiert für alle d' mit $(d, d') \in r^{\mathcal{I}_1}$ eine Kante $((d, e), r, (d', e')) \in E_{sim}$ mit $(e, e') \in r^{\mathcal{I}_2}$.
2. Wenn $((d, e), r, (d', e')) \in E_{sim}$, dann $(d, d') \in r^{\mathcal{I}_1}$ und $(e, e') \in r^{\mathcal{I}_2}$.
3. Aus $((d, e), r, (d', e')) \in E_{sim}$ und $((d, e), r, (d', f')) \in E_{sim}$ folgt $e' = f'$.
4. Wenn $((d, e), r, (d', e')) \in E_{sim}$, dann existiert ein Pfad $(d_0, e_0)r_1(d_1, e_1)r_2\dots r_n(d_n, e_n)$ mit $((d_i, e_i), r_{i+1}, (d_{i+1}, e_{i+1})) \in E_{sim}$ für alle $i = 0, \dots, n-1$ und $(d_n, e_n) = (d, e)$.

Zusätzlich werden Projektionen für eine Simulation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ mit $\text{pr}_i : \mathcal{S} \rightarrow \Delta^{\mathcal{I}_i}$ und $\text{pr}_i(d_1, d_2) := d_i$ für $i = 1, 2$ definiert. Die Projektionen werden auf Pfade in \mathcal{G}_{sim} erweitert. Für einen Pfad $\pi = (d_0, e_0)r_0(d_1, e_1)r_1\dots$ in \mathcal{G}_{sim} ist $\text{pr}_1(\pi) := d_0r_0d_1r_1\dots$ und $\text{pr}_2(\pi) := e_0r_0e_1r_1\dots$

Nach der Definition können für eine Simulation und einen gewählten Startknoten mehrere verschiedene Simulationsgraphen existieren. Dazu wird ein Beispiel betrachtet.

Beispiel 3.4.2. Betrachtet wird der Produktgraph $\mathcal{G}_{\mathcal{T}_3}$ bzw. das Produktmodell $(\mathcal{I}_{\mathcal{T}_3,A} \times \mathcal{I}_{\mathcal{T}_3,B}, (d_A, d_B))$ in Abb. 3.1. und die folgende Simulation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_{\mathcal{T}_3,A} \times \mathcal{I}_{\mathcal{T}_3,B}} \times \Delta^{\mathcal{I}_{\mathcal{T}_3,A} \times \mathcal{I}_{\mathcal{T}_3,B}}$ mit

$$\mathcal{S} := \{((d_A, d_B), (d_A, d_B)), ((d_A, d_B), (d_C, d_C)), ((d_C, d_C), (d_C, d_C))\}.$$

Die beiden möglichen Simulationsgraphen zu \mathcal{S} mit dem Startknoten $((d_A, d_B), (d_A, d_B))$ sind in Abb. 3.4 dargestellt. \mathcal{G}_{sim}^1 enthält ebenfalls den Knoten $((d_A, d_B), (d_C, d_C))$, der aber nicht in E_{sim}^1 verbunden ist, da E_{sim}^1 die Bedingungen 3. und 4. aus Definition 3.4.3 erfüllt.

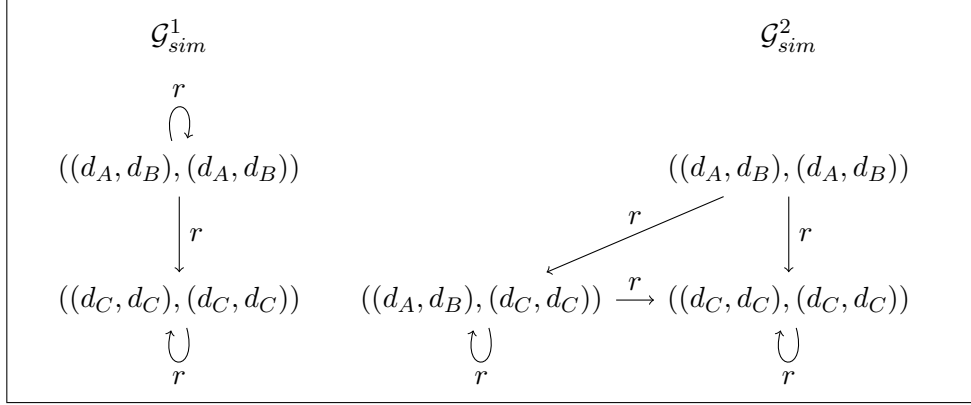


Abbildung 3.4: Simulationsgraphen zu Beispiel 3.4.2

Im Folgenden werden Eigenschaften von Simulationen auf dem reduzierten Produktmodell $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ beschrieben. Sei $\mathcal{S} \subseteq \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \times \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$ und $(d, e) \in \mathcal{S}$. Das Element (d, e) heißt *synchronisiert* genau dann, wenn e ein synchrones Element ist. Betrachtet werden außerdem *starke Zusammenhangskomponenten* (SZKs) in den Simulationsgraphen. Es wird die übliche Definition von starken Zusammenhangskomponenten in gerichteten Graphen verwendet. Eine SZK Z in einem Simulationsgraphen \mathcal{G}_{sim} wird genau dann synchronisiert genannt, wenn Z ein synchronisiertes Element enthält. Außerdem heißt eine SZK Z genau dann *nicht-trivial*, wenn Z genau ein Element mit einer Schleife oder mindestens zwei Elemente enthält.

Satz 3.4.1. *Sei \mathcal{T} eine normalisierte \mathcal{EL} -TBox, A, B Konzeptnamen aus \mathcal{T} und $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ das reduzierte Produktmodell. Der LCS C von A und B bezüglich \mathcal{T} existiert genau dann, wenn eine Simulation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \times \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$ und ein Simulationsgraph $\mathcal{G}_{sim} = (\mathcal{S}, E_{sim}, ((d_A, d_B), (d_A, d_B)))$ existieren, so dass alle nicht-trivialen starken Zusammenhangskomponenten in \mathcal{G}_{sim} synchronisiert sind.*

Beweis. \Rightarrow : Sei C der LCS von A, B bezüglich \mathcal{T} mit $rt(C) = k$, so dass $k\text{-lcs}(A, B) \equiv_{\mathcal{T}} C$ gilt. Aus Lemma 3.4.3 folgt dann $\mathcal{J}_{(d_A, d_B)} \lesssim \widehat{\mathcal{J}}^k_{(d_A, d_B)}$. Daher existiert eine Simulation

$$\mathcal{S}_{\mathcal{J}} \subseteq \Delta^{\mathcal{J}_{(d_A, d_B)}} \times \Delta^{\widehat{\mathcal{J}}^k_{(d_A, d_B)}} \quad (3.20)$$

mit $((d_A, d_B), (d_A, d_B)) \in \mathcal{S}_{\mathcal{J}}$. Daraus ergibt sich eine Simulation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \times \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$ mit

$$\mathcal{S} := \{(d, d') \mid (\sigma d, \sigma' d') \in \mathcal{S}_{\mathcal{J}}\}. \quad (3.21)$$

Angenommen es gibt einen Simulationsgraphen $\mathcal{G}_{sim} = (\mathcal{S}, E_{sim}, ((d_A, d_B), (d_A, d_B)))$ und eine Menge

$\mathcal{Z} := \{Z_1, \dots, Z_n\}$ von nicht synchronisierten, nicht-trivialen SZKs Z_i in \mathcal{G}_{sim} . Da das reduzierte Produktmodell betrachtet wird, sind nach Definition 3.4.1 alle Knoten in den Z_i nicht synchronisiert. Es wird gezeigt, dass aus \mathcal{G}_{sim} ein Simulationsgraph konstruiert werden kann, der nur synchronisierte nicht-triviale SZKs enthält.

Zu jeder $Z_i \in \mathcal{Z}$ existiert ein zyklischer unendlicher Pfad π_i in \mathcal{G}_{sim} der Form

$$\pi_i = (d_0, d'_0)r_1(d_1, d'_1)r_2 \dots r_n(d_n, d'_n)r_{n+1} \dots r_m(d_n, d'_n)r_{n+1} \dots$$

mit $(d_0, d'_0) = ((d_A, d_B), (d_A, d_B))$, alle (d_j, d'_j) in π_i sind nicht synchronisiert und in dem Zyklus $(d_n, d'_n)r_{n+1} \dots r_m(d_n, d'_n)$ ist ein Knoten aus Z_i enthalten. Sei (d_k, d'_k) das Element an Position k in π_i mit dem r_{k+1} Nachfolger (d_{k+1}, d'_{k+1}) . Mit (3.21) und (3.20) folgt, dass ein Wort $\sigma d_k \in \Delta^{\mathcal{J}(d_A, d_B)}$ und ein Wort $\sigma' d'_k \in \Delta^{\widehat{\mathcal{J}}^k(d_A, d_B)}$ existiert mit $(\sigma d_k, \sigma' d'_k) \in \mathcal{S}_{\mathcal{J}}$. Für den r_{k+1} -Nachfolger $\sigma d_k r_{k+1} d_{k+1} \in \Delta^{\mathcal{J}(d_A, d_B)}$ gibt es einen r_{k+1} -Nachfolger $\sigma' d'_k r_{k+1} e'_{k+1} \in \Delta^{\widehat{\mathcal{J}}^k(d_A, d_B)}$ mit $(\sigma d_k r_{k+1} d_{k+1}, \sigma' d'_k r_{k+1} e'_{k+1}) \in \mathcal{S}_{\mathcal{J}}$, da $\mathcal{S}_{\mathcal{J}}$ eine Simulation ist. Nach Definition von $\widehat{\mathcal{J}}^k(d_A, d_B)$ ist e'_{k+1} ein synchrones Element. Mit (3.21) folgt $(d_{k+1}, e'_{k+1}) \in \mathcal{S}$. Sei $\widehat{\mathcal{G}}_{sim} = (\mathcal{S}, \widehat{E}_{sim}, (d_{k+1}, e'_{k+1}))$ ein Simulationsgraph mit dem Startknoten (d_{k+1}, e'_{k+1}) . e'_{k+1} ist ein synchrones Element und nach Definition des reduzierten Produktmodells hat e'_{k+1} nur synchrone Nachfolger. Demzufolge enthält $\widehat{\mathcal{G}}_{sim}$ nur synchronisierte SZKs. Die Idee besteht darin, das von dem Knoten (d_{k+1}, d'_{k+1}) in \mathcal{G}_{sim} erreichbare Fragment durch den Simulationsgraphen $\widehat{\mathcal{G}}_{sim} = (\mathcal{S}, \widehat{E}_{sim}, (d_{k+1}, e'_{k+1}))$ zu ersetzen.

Der Simulationsgraph $\mathcal{G}'_{sim} = (\mathcal{S}, E'_{sim}, ((d_A, d_B), (d_A, d_B)))$ entsteht wie folgt aus $\mathcal{G}_{sim} = (\mathcal{S}, E_{sim}, ((d_A, d_B), (d_A, d_B)))$:

$$\begin{aligned} E' &:= (E_{sim} \setminus \{(d_k, d'_k)r_{k+1}(d_{k+1}, d'_{k+1})\}) \\ &\cup \{(d_k, d'_k)r_{k+1}(d_{k+1}, e'_{k+1})\} \cup \widehat{E}_{sim} \end{aligned}$$

E' erfüllt nicht notwendigerweise die Bedingungen 3. und 4. aus Definition 3.4.3. Durch das Entfernen der Kante $(d_k, d'_k)r_{k+1}(d_{k+1}, d'_{k+1})$ kann Bedingung 4. verletzt sein und durch Hinzufügen der Kanten aus \widehat{E}_{sim} kann Bedingung 3. verletzt sein.

Die folgenden Schritte werden so lange angewendet, bis E' alle Bedingungen aus Definition 3.4.3 erfüllt.

1. Sei $((f, g), r, (h, k)) \in E'$ und (f, g) ist nicht von $((d_A, d_B), (d_A, d_B))$ in E' erreichbar. Dann wird $((f, g), r, (h, k))$ aus E' entfernt.

2. Seien $((f, g), r, (h, k)), ((f, g), r, (h, k')) \in E'$ mit $k \neq k'$. Dann gilt entweder $((f, g), r, (h, k)) \in \widehat{E}_{sim}$ oder $((f, g), r, (h, k')) \in \widehat{E}_{sim}$. Angenommen es gilt $((f, g), r, (h, k')) \in \widehat{E}_{sim}$. Dann wird $((f, g), r, (h, k))$ aus E' entfernt.

Die resultierende Menge ist E'_{sim} . Der Graph \mathcal{G}'_{sim} enthält nach Konstruktion mindestens eine nicht synchronisierte, nicht-triviale SZK weniger als \mathcal{G}_{sim} , da mindestens ein Element aus Z_i durch ein synchronisiertes Element ersetzt wurde. Die beschriebene Ersetzung kann mit \mathcal{G}'_{sim} fortgesetzt werden bis nur noch synchronisierte nicht-triviale starke SZKs vorkommen.

\Leftarrow : Sei $\mathcal{S} \subseteq \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \times \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$ eine Simulation und $\mathcal{G}_{sim} = (\mathcal{S}, E_{sim}, ((d_A, d_B), (d_A, d_B)))$ ein Simulationsgraph in dem nur synchronisierte SZKs vorkommen. Da das reduzierte Produktmodell betrachtet wird, haben synchronisierte Knoten in \mathcal{G}_{sim} nur synchronisierte Nachfolger. Daher existiert eine natürliche Zahl k , so dass für alle Pfade π in \mathcal{G}_{sim} mit einer Länge $\geq k$ der Form:

$$\pi = ((d_A, d_B), (d_A, d_B))r_1(d_1, d'_1)r_2 \dots r_k(d_k, d'_k)r_{k+1}(d_{k+1}, d'_{k+1}) \dots$$

alle Knoten (d_j, d'_j) mit $j \geq k$ synchronisiert sind, d. h. d'_j ist ein synchrones Element. Für alle Pfade π in \mathcal{G}_{sim} und die entsprechenden Projektionen

$$\text{pr}_2(\pi) = (d_A, d_B)r_1d'_1r_2d'_2r_3 \dots$$

gibt es einen entsprechenden Pfad \hat{p} in der Baumabwicklung von $\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}$ mit

$$\hat{p} = (d_A, d_B)r_1\sigma_1d'_1r_2\sigma_2d'_2r_3$$

und \hat{p} ist in dem Modell $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ enthalten, da \hat{p} ab Position k nur synchrone Elemente enthält. Daraus folgt, dass jeder Pfad in $\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}$ beginnend mit (d_A, d_B) durch einen Pfad in $\widehat{\mathcal{J}}^k_{(d_A, d_B)}$ beginnend mit der Wurzel simuliert werden kann. Damit folgt

$$(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B)) \lesssim \widehat{\mathcal{J}}^k_{(d_A, d_B)}.$$

Mit Lemma 3.4.3 und Satz 3.3.1 folgt, dass $\text{k-lcs}(A, B)$ der LCS von A, B bezüglich \mathcal{T} ist. □

Für den Spezialfall eines azyklischen Produktgraphen folgt damit die Existenz des LCS direkt. Der Simulationsgraph zu der Simulation, in der jeder Knoten sich selbst simuliert, erfüllt die geforderten Bedingungen. Gleiches gilt für Produktgraphen, die nur Zyklen mit synchronen Knoten enthalten wie Beispiel $\mathcal{G}_{\mathcal{T}_2}$ in Abschnitt 3.1.

Zusammenfassend ergibt sich aus diesem Satz die folgende Schlussfolgerung:

Korollar 3.4.1. *Seien A und B Konzeptnamen aus einer normalisierten \mathcal{EL} -TBox. Das Problem ob der LCS von A und B bezüglich \mathcal{T} existiert ist entscheidbar. Existiert der LCS C , dann gilt für die Rollentiefe $rt(C) \leq n^4$ mit $n = |N_{C,\mathcal{T}}|$.*

Beweis. Das Entscheidungsverfahren für das Existenzproblem des LCS besteht aus den folgenden Schritten.

1. Zunächst wird das Modell $(\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}, (d_A, d_B))$ berechnet. In Lemma 3.2.1 wurde gezeigt, dass sich dieses Modell aus dem Produktgraphen nach Definition 2.2.2 ergibt.
2. Die maximale Simulation $\mathcal{S} \subseteq \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}} \times \Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}$ kann ebenfalls in polynomieller Zeit berechnet werden.
3. Es gibt endlich viele Simulationsgraphen der Form $\mathcal{G}_{sim} = (\mathcal{S}, E_{sim}, ((d_A, d_B), (d_A, d_B)))$ für \mathcal{S} . Für diese Graphen werden die nicht-trivialen starken Zusammenhangskomponenten berechnet. Gibt es einen Graphen, dessen nicht-triviale starke Zusammenhangskomponenten synchronisiert sind, existiert der LCS von A, B bezüglich \mathcal{T} nach Satz 3.4.1.
4. Ist einer solcher Graph gegeben, der die Bedingung aus Satz 3.4.1 erfüllt, kann die natürliche Zahl k aus der Rückrichtung des Beweises von Satz 3.4.1 bestimmt werden. Nach der Konstruktion im Beweis ergibt sich k aus der maximalen Länge eines Pfades π in dem Simulationsgraph, so dass in π nur nicht synchronisierte Element durchlaufen werden. Für k gilt dann $k = |\pi| + 1$. Da der Simulationsgraph nach Voraussetzung nur synchronisierte nicht-triviale SZKs enthält, existiert ein solches maximales k . Dieses k ist nach oben durch die Anzahl der Knoten in dem Simulationsgraph beschränkt. Es gilt $|\Delta^{\mathcal{I}_{\mathcal{T},A} \times \mathcal{I}_{\mathcal{T},B}}| \leq |N_{C,\mathcal{T}}|^2$ und entsprechend gilt $k \leq |\mathcal{S}| \leq |N_{C,\mathcal{T}}|^4$. k entspricht der Rollentiefe des LCS, wie in Satz 3.4.1 gezeigt wurde.

□

Kapitel 4

Der \mathcal{EL} -LCS bezüglich einer generellen \mathcal{ALC} -TBox

In diesem Kapitel wird der \mathcal{EL} -LCS von \mathcal{EL} -Konzeptbeschreibungen bezüglich einer generellen \mathcal{ALC} -TBox betrachtet.

In [BST04] wurde dieses Problem bereits untersucht. Auf Basis der Subsumptionshierarchie der Konjunktionen von Konzeptnamen aus der generellen \mathcal{ALC} -TBox wurden Approximationen des \mathcal{EL} -LCS berechnet. Das Ziel dieses Kapitels ist es, einen Vervollständigungsgraphen zu berechnen, der nicht nur alle impliziten Subsumptionsbeziehungen der Form

$$A_1 \sqcap \dots \sqcap A_n \sqsubseteq_{\mathcal{T}} B_1 \sqcap \dots \sqcap B_m \quad (4.1)$$

berücksichtigt sondern auch alle Subsumptionsbeziehungen der Form

$$A_1 \sqcap \dots \sqcap A_n \sqsubseteq_{\mathcal{T}} \exists r.(B_1 \sqcap \dots \sqcap B_m). \quad (4.2)$$

Auf einen Graphen mit diesen Eigenschaften können die Methoden zur Berechnung des Rolltiefen- beschränkten LCS aus [PT10b] und das Entscheidungsverfahren für die Existenz des exakten LCS aus Kapitel 3 angewendet werden.

Als Ausgangspunkt dient dabei das Vervollständigungsverfahren aus [SKH11] für generelle \mathcal{ALC} -TBoxen. Im ersten Abschnitt werden die Normalform und die Inferenzregeln aus [SKH11] vorgestellt. Da das Verfahren in [SKH11] nur vollständig bezüglich Subsumptionsbeziehungen der Form (4.1) ist, wird im nächsten Abschnitt eine Erweiterung der Methode vorgestellt. Anschließend wird die Vollständigkeit und Korrektheit des Verfahrens bewiesen und die Komplexität untersucht.

In diesem Kapitel werden die folgenden Bezeichnungen verwendet:

- H, K bezeichnen Konjunktionen und M, N Disjunktionen von Konzeptnamen;

- H, K und M, N werden auch als Mengen aufgefasst und es werden entsprechende Mengenoperationen verwendet. Die leere Konjunktion H oder K bezeichnet \top und die leere Disjunktion bezeichnet \perp .

In den folgenden Abschnitten wird davon ausgegangen, dass eine konsistente \mathcal{ALC} -TBox vorliegt.

4.1 Normalisierung und Inferenzregeln für generelle \mathcal{ALC} -TBoxen

Das Ziel der Normalisierung ist es, eine möglichst einfache und flache Struktur der TBox zu erhalten. Dazu werden für komplexe Teilkonzepte aus der TBox \mathcal{T} neue Konzeptnamen eingeführt.

In [SKH11] wurde gezeigt, dass zu einer generelle \mathcal{ALC} -TBox \mathcal{T} eine \mathcal{ALC} -TBox $\hat{\mathcal{T}}$ berechnet werden kann, die nur Konzeptinklusionen der Form

$$(n1) \prod_i A_i \sqsubseteq \bigsqcup_j B_j \quad (n2) A \sqsubseteq \exists r.B \quad (n3) A \sqsubseteq \forall r.B \quad (n4) \exists r.A \sqsubseteq B$$

enthält.

Zusätzlich wird die folgende Normalisierungsregel betrachtet:

$$\exists r.A \sqsubseteq B \rightarrow \{ \top \sqsubseteq [\forall r. \neg A] \sqcup B, [\forall r. \neg A] \sqsubseteq \forall r. [\neg A], \quad [\neg A] \sqcap A \sqsubseteq \perp \} \quad (4.3)$$

Dabei sind $[\forall r. \neg A]$ und $[\neg A]$ neue Konzeptnamen. Daraus folgt, dass Inklusionen der Form (n4) auch mit Hilfe von Inklusionen der Form (n1) und (n3) ersetzt werden können, so dass Subsumptionsbeziehungen zwischen Konzeptbeschreibungen mit Symbolen aus der ursprünglichen Signatur bewahrt bleiben.

Lemma 4.1.1. *Sei \mathcal{T} eine \mathcal{ALC} -TBox. Es kann zu \mathcal{T} eine TBox \mathcal{T}' berechnet werden, die nur Inklusionen der Form (n1), (n2) oder (n3) enthält und jedes Modell von \mathcal{T} kann zu einem Modell von \mathcal{T}' erweitert werden und umgekehrt ist jedes Modell von \mathcal{T}' auch ein Modell für \mathcal{T} .*

Beweis. Das Lemma gilt nach [SKH11] für die Normalform mit den Inklusionen (n1-4). Sei $\hat{\mathcal{T}}$ eine TBox in dieser Form und bezeichne \mathcal{T}' die TBox, nachdem die Regel (4.3) erschöpfend auf $\hat{\mathcal{T}}$ angewendet wurde. Für eine beliebige \mathcal{ALC} -TBox \mathcal{T} gilt $\exists r.A \sqsubseteq_{\mathcal{T}} B$ genau dann, wenn $\top \sqsubseteq_{\mathcal{T}} \forall r. \neg A \sqcup B$. Daher erfüllt \mathcal{T}' die geforderten Bedingungen. \square

Der zusätzliche Normalisierungsschritt wird nur eingeführt um die Darstellung und die Beweise im Folgenden zu vereinfachen.

Der nächste Schritt besteht darin die normalisierte TBox mit Hilfe von Inferenzregeln zu vervollständigen. In Abbildung 4.1 sind die Inferenzregeln für

$$\begin{array}{l}
\text{CR}_0 \frac{A \in H}{H \sqsubseteq A} \\
\\
\text{CR}_1 \frac{\{H \sqsubseteq N_i \sqcup A_i\}_{i=1}^n \quad \prod_{i=1}^n A_i \sqsubseteq M}{H \sqsubseteq \bigsqcup_{i=1}^n N_i \sqcup M} \\
\\
\text{CR}_2 \frac{H \sqsubseteq N \sqcup A \quad A \sqsubseteq \exists r.B}{H \sqsubseteq N \sqcup \exists r.B} \\
\\
\text{CR}_3 \frac{H \sqsubseteq M \sqcup \exists r.K \quad K \sqsubseteq \perp}{H \sqsubseteq M} \\
\\
\text{CR}_4 \frac{H \sqsubseteq M \sqcup \exists r.K \quad H \sqsubseteq N \sqcup A \quad A \sqsubseteq \forall r.B}{H \sqsubseteq M \sqcup N \sqcup \exists r.(K \sqcap B)}
\end{array}$$

Abbildung 4.1: Inferenzregeln für \mathcal{ALC} [SKH11]

\mathcal{ALC} dargestellt. Da im Vergleich zu dem in [SKH11] beschriebenen Verfahren eine vereinfachte Normalform verwendet wird, sind nicht alle Inferenzregeln aus [SKH11] notwendig. Es werden Konzeptinklusionen der Form

$$H \sqsubseteq M \sqcup \exists r.K \qquad H \sqsubseteq M$$

abgeleitet. Die Korrektheit der Inferenzregeln wurde in [SKH11] gezeigt. Die vervollständigte TBox \mathcal{T}_{sat} entsteht aus der normalisierten TBox \mathcal{T} , indem die Regeln erschöpfend auf \mathcal{T} angewendet werden. Dabei wird angenommen, dass die Inklusionen $H \sqsubseteq M \sqcup \exists r.K$ und $H \sqsubseteq M$ nur genau dann abgeleitet werden, wenn diese nicht bereits in der TBox enthalten sind. Die Anwendung der Regeln terminiert. H , M und K sind Mengen von Konzeptnamen. Es gibt demnach jeweils $2^{|N_{C,\mathcal{T}}|}$ viele solcher Mengen und damit maximal n

$$n = 2^{|N_{C,\mathcal{T}}|} \cdot 2^{|N_{C,\mathcal{T}}|} \cdot |N_{R,\mathcal{T}}| \cdot 2^{|N_{C,\mathcal{T}}|} \quad (4.4)$$

Inklusionen der Form $H \sqsubseteq M \sqcup \exists r.K$.

In [SKH11] wurde bewiesen, dass mit den Inferenzregeln alle inkonsistenten Konjunktionen H abgeleitet werden können. Das bedeutet $H \sqsubseteq_{\mathcal{T}} \perp$ genau dann, wenn $H \sqsubseteq \perp \in \mathcal{T}_{sat}$. \mathcal{T}_{sat} ist jedoch in der Hinsicht unvollständig, dass nicht alle impliziten Subsumptionsbeziehungen der Form $H \sqsubseteq \exists r.K$ in \mathcal{T}_{sat} enthalten sind. Dazu wird ein Beispiel betrachtet.

Beispiel 4.1.1. Betrachtet wird die folgende normalisierte \mathcal{ALC} -TBox \mathcal{T} .

$$\begin{aligned} \mathcal{T} := \{ & A \sqsubseteq B \sqcup C, \\ & B \sqsubseteq \exists r.D, C \sqsubseteq \exists r.E, \\ & D \sqsubseteq F, E \sqsubseteq F \} \end{aligned}$$

Für die linke Seite A enthält \mathcal{T}_{sat} die folgenden Inklusionen:

$$\begin{aligned} \mathcal{T}_{sat} := \{ & \dots, \\ & A \sqsubseteq A, \\ & A \sqsubseteq B \sqcup C, \\ & A \sqsubseteq B \sqcup \exists r.E, \\ & A \sqsubseteq C \sqcup \exists r.D, \\ & \dots \} \end{aligned}$$

Es folgt $A \sqsubseteq_{\mathcal{T}} \exists r.D \sqcup \exists r.E$ und damit folgt auch $A \sqsubseteq_{\mathcal{T}} \exists r.F$. Es gibt jedoch keine Inklusionen $A \sqsubseteq \exists r.F' \in \mathcal{T}_{sat}$ und $F' \sqsubseteq F \in \mathcal{T}_{sat}$.

Das Beispiel zeigt, dass \mathcal{T}_{sat} unvollständig bezüglich Inklusionen der Form (4.2) ist. Ein möglicher Lösungsansatz ist, die Inferenzregeln zu erweitern und zum Beispiel auch Disjunktionen von Existenzrestriktionen abzuleiten, so dass in Beispiel 4.1.1 die Inklusion $A \sqsubseteq \exists r.D \sqcup \exists r.E$ abgeleitet werden kann. Durch die Weitererestriktionen entstehen dann jedoch Disjunktionen von Konjunktionen von Konzeptnamen und es kann keine exponentielle obere Schranke wie (4.4) für die Zahl der möglichen Inklusionen angegeben werden.

Im Folgenden Abschnitt wird daher eine alternative Methode vorgestellt, um alle impliziten Subsumptionsbeziehungen zu berechnen.

4.2 Vervollständigung von \mathcal{ALC} -TBoxen

In diesem Abschnitt wird auf Grundlage von \mathcal{T}_{sat} ein Vervollständigungsgraph konstruiert. Die Idee ist für alle H mit $H \not\sqsubseteq_{\mathcal{T}} \perp$ jeweils einen Knoten einzuführen, der mit allen Konzeptnamen beschriftet ist, die H subsumieren. Die Knoten H, K werden durch eine r -Kante verbunden genau dann, wenn $H \sqsubseteq_{\mathcal{T}} \exists r.K$ gilt. Da die Inklusionen in \mathcal{T}_{sat} Disjunktionen enthalten, muss eine Menge von möglichen Beschriftungen und eine Menge von möglichen Kanten berücksichtigt werden. Dazu wird für ein H aus \mathcal{T}_{sat} eine Menge von *Typen* definiert.

Definition 4.2.1 (Typen). Sei \mathcal{T}_{sat} die vervollständigte TBox. Eine Menge von H -Typen bezüglich \mathcal{T}_{sat} wird für alle H mit $H \sqsubseteq \perp \notin \mathcal{T}_{sat}$ definiert. Mit \mathcal{M}_H^{min} wird die Menge der minimalen Disjunktionen M , für die

$H \sqsubseteq M \in \mathcal{T}_{sat}$ gilt, bezeichnet.

$$\mathcal{M}_H^{min} := \{M \mid H \sqsubseteq M \in \mathcal{T}_{sat} \wedge \nexists M' \text{ mit } H \sqsubseteq M' \in \mathcal{T}_{sat} \text{ und } M' \subset M\} \quad (4.5)$$

Ein Typ τ_H wird als minimale Menge von Konzeptnamen so definiert, dass alle minimale Disjunktionen aus \mathcal{M}_H^{min} überdeckt werden. τ_H erfüllt die folgenden Bedingungen:

1. Für alle $M \in \mathcal{M}_H^{min}$ gilt $\tau_H \cap M \neq \emptyset$.
2. τ_H ist eine minimale Menge, die 1. erfüllt.

Zusätzlich wird für jede Rolle r in Abhängigkeit von τ_H eine Menge von Konjunktionen $\tilde{\tau}_{H,r}$ definiert, um alle Inklusionen $H \sqsubseteq M \sqcup \exists r.K \in \mathcal{T}_{sat}$, für die τ_H noch keinen Konzeptnamen aus M enthält, zu erfüllen.

$$\tilde{\tau}_{H,r} := \{K \mid H \sqsubseteq M \sqcup \exists r.K \in \mathcal{T}_{sat} \wedge M \cap \tau_H = \emptyset\} \quad (4.6)$$

Aus dieser Menge werden die maximalen K ausgewählt.

$$\tau_{H,r} := \{K \mid K \in \tilde{\tau}_{H,r} \wedge \nexists K' \in \tilde{\tau}_{H,r} \text{ mit } K \subset K'\} \quad (4.7)$$

Die Menge aller H -Typen τ_H bezüglich \mathcal{T}_{sat} wird mit T_H bezeichnet. Abschließend werden in $\mathbf{types}(\mathcal{T}_{sat})$ alle H -Typen für \mathcal{T}_{sat} zusammengefasst.

$$\mathbf{types}(\mathcal{T}_{sat}) := \{\tau_H \mid \tau_H \in T_H, H \sqsubseteq \perp \notin \mathcal{T}_{sat}\} \quad (4.8)$$

Bevor einige Eigenschaften dieser H -Typen bewiesen werden, wird das Beispiel 4.1.1 fortgesetzt.

Beispiel 4.2.1. Betrachtet wird die TBox aus Beispiel 4.1.1 mit dem Konzeptnamen A . Für A gibt es die folgenden Typen $T_A = \{\tau_A^1, \tau_A^2\}$ mit

$$\begin{aligned} \tau_A^1 &= \{A, B, \top\}, \tau_{A,r}^1 = \{D\} \\ \tau_A^2 &= \{A, C, \top\}, \tau_{A,r}^2 = \{E\} \end{aligned}$$

Lemma 4.2.1. Sei $\tau_H \in \mathbf{types}(\mathcal{T}_{sat})$.

1. Für alle $H \sqsubseteq M \in \mathcal{T}_{sat}$ gilt $M \cap \tau_H \neq \emptyset$.
2. Für alle $A \in \tau_H$ existiert $H \sqsubseteq N \sqcup A \in \mathcal{T}_{sat}$ mit $\tau_H \cap N = \emptyset$.
3. Für alle $K \in \tau_{H,r}$ ist $K \sqsubseteq \perp \notin \mathcal{T}_{sat}$.
4. Es gilt $H \subseteq \tau_H$.

Beweis. 1. Nach Bedingung (1) hat τ_H einen nicht-leeren Schnitt mit allen Mengen aus \mathcal{M}_H^{min} (4.5). Da für alle M mit $H \sqsubseteq M \in \mathcal{T}_{sat}$ eine minimale Menge $M^{min} \in \mathcal{M}_H^{min}$ existiert mit $M^{min} \subseteq M$, gilt auch $M \cap \tau_H \neq \emptyset$.

2. Sei $A \in \tau_H$. Angenommen für alle Disjunktionen $M = N \sqcup A$ mit $H \sqsubseteq N \sqcup A$, die A enthalten, gilt $N \cap \tau_H \neq \emptyset$. Daraus folgt, dass $\tau_H \setminus \{A\}$ ebenfalls Bedingung 1 aus der Definition von τ_H erfüllt. Damit ist τ_H nicht minimal und es ergibt sich ein Widerspruch zu Bedingung 2.
3. Angenommen $K \in \tau_{H,r}$ und $K \sqsubseteq \perp \in \mathcal{T}_{sat}$. Für $K \in \tau_{H,r}$ existiert nach (4.6) $H \sqsubseteq M \sqcup \exists r.K \in \mathcal{T}_{sat}$ mit $M \cap \tau_H = \emptyset$. Mit Anwendung der Inferenzregel CR_3 auf $H \sqsubseteq M \sqcup \exists r.K$ und $K \sqsubseteq \perp$ folgt $H \sqsubseteq M \in \mathcal{T}_{sat}$. Daher ist nach 1. $M \cap \tau_H \neq \emptyset$. Das ist ein Widerspruch und es folgt $K \sqsubseteq \perp \notin \mathcal{T}_{sat}$.
4. Sei $H = A_1 \sqcap \dots \sqcap A_n$. Es ist zu zeigen, dass $A_i \in \tau_H$ für alle $i = 1, \dots, n$. Es gilt $A_i \in H$ und mit Inferenzregel CR_0 folgt $H \sqsubseteq A_i \in \mathcal{T}_{sat}$. Damit ist $\{A_i\}$ in \mathcal{M}_H^{min} (4.5) enthalten und es folgt $A_i \in \tau_H$.

□

Der Vervollständigungsgraph zu \mathcal{T}_{sat} wird als Produkt der einzelnen Typen gebildet und ist wie folgt definiert:

Definition 4.2.2 (Vervollständigungsgraph). Sei \mathcal{T} eine \mathcal{ALC} -TBox, \mathcal{T}_{sat} die Vervollständigung und $T_H := \{\tau_H^1, \dots, \tau_H^n\}$ die Menge der H -Typen. Der Vervollständigungsgraph $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ ergibt sich wie folgt:

- $V := \{H \mid H \sqsubseteq \perp \notin \mathcal{T}_{sat}\}$;
- $S(H) := \bigcap_{i=1}^n \tau_H^i$ und
- $E := \{(H, r, K) \mid \exists (K_1, \dots, K_n) \in \tau_{H,r}^1 \times \dots \times \tau_{H,r}^n, K = \bigcap_{i=1}^n S(K_i)\}$.

Beispiel 4.2.2. Es wird wieder der Konzeptnamen A aus Beispiel 4.2.1 betrachtet mit den zugehörigen Typen:

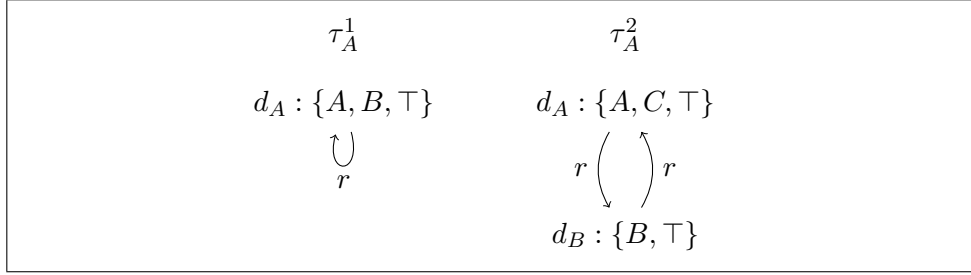
$$\begin{aligned} \tau_A^1 &= \{A, B, \top\}, \tau_{A,r}^1 = \{D\} \\ \tau_A^2 &= \{A, C, \top\}, \tau_{A,r}^2 = \{E\} \\ \tau_D &= \{D, F, \top\}, \tau_{D,r} = \emptyset \\ \tau_E &= \{E, F, \top\}, \tau_{E,r} = \emptyset \end{aligned}$$

Für den Knoten A ergeben sich damit im Vervollständigungsgraphen zu \mathcal{T} die Beschriftung $S(A) = \tau_A^1 \cap \tau_A^2 = \{A, \top\}$ und die Kante (A, r, F) .

Um die Vollständigkeit und Korrektheit zu zeigen, werden wie in [BBL05] und in [SKH11] kanonische Modelle konstruiert. Dazu wird zunächst ein weiteres Beispiel betrachtet.

Beispiel 4.2.3. Betrachtet wird die folgende normalisierte \mathcal{ALC} -TBox \mathcal{T} .

$$\begin{aligned} \mathcal{T} &:= \{A \sqsubseteq B \sqcup C, \\ &\quad B \sqsubseteq \exists r.A, \\ &\quad C \sqsubseteq \exists r.B\} \end{aligned}$$

**Abbildung 4.2:** Kanonische Modelle zu Beispiel 4.2.3

Mit den Typen:

$$\begin{array}{ll}
 \tau_A^1 = \{A, B, \top\}, & \tau_{A,r}^1 = \{A\} \\
 \tau_A^2 = \{A, C, \top\}, & \tau_{A,r}^2 = \{B\} \\
 \tau_B = \{B, \top\}, & \tau_{B,r} = \{A\} \\
 \tau_C = \{C, \top\}, & \tau_{C,r} = \{B\}
 \end{array}$$

Werden die kanonischen Modelle so konstruiert, dass für jedes H genau ein Element d_H eingeführt wird, dann ergeben sich im Beispiel für das Element d_A genau zwei mögliche kanonische Modelle mit den beiden Typen τ_A^1 und τ_A^2 aus T_A . Beide Möglichkeiten sind in Abb. 4.2 dargestellt. In beiden Modellen hat d_A eine r -Nachfolger in B . Es gilt aber $A \not\sqsubseteq_{\mathcal{T}} \exists r.B$, da Elemente aus A in einem Modell von \mathcal{T} auch zum Beispiel einen r -Nachfolger haben können, der in A und C und nicht in B liegt. Ein Element d_A vom Typ τ_A^1 kann also auch einen r -Nachfolger d'_A vom Typ τ_A^2 haben. Für die Definition der kanonischen Modelle müssen diese Fälle berücksichtigt werden.

Basierend auf der Menge der Typen $\text{types}(\mathcal{T}_{sat})$ wird eine Menge von kanonischen Interpretationen wie folgt definiert:

Definition 4.2.3 (Kanonische Interpretationen für \mathcal{T}_{sat}). Eine *kanonische Interpretation* besteht aus einer Interpretation \mathcal{I} und einer totalen Funktion θ , die jedem Element $d \in \Delta^{\mathcal{I}}$ einen H -Typen $\theta(d) \in \text{types}(\mathcal{T}_{sat})$ zuordnet. Wenn $\theta(d') = \tau_H$ für ein $d' \in \Delta^{\mathcal{I}}$ gilt, dann bezeichnet $\theta(d)_r$ die Menge $\tau_{H,r}$. Seien $d, e \in \Delta^{\mathcal{I}}$. Für $A^{\mathcal{I}}$ und $r^{\mathcal{I}}$ gelten folgende Bedingungen:

- $d \in A^{\mathcal{I}}$ genau dann, wenn $A \in \theta(d)$.
- Wenn $K \in \theta(d)_r$, dann existiert ein $e \in \Delta^{\mathcal{I}}$ mit $\theta(e) \in T_K$ und $(d, e) \in r^{\mathcal{I}}$.
- Wenn $(d, e) \in r^{\mathcal{I}}$, dann gilt $\theta(e) \in T_K$ für ein $K \in \theta(d)_r$.

$\text{kan}(\mathcal{T}_{sat})$ bezeichnet die Menge aller kanonischen Interpretationen. $\mathcal{I}, \theta \in \text{kan}(\mathcal{T}_{sat})$ bezeichnet eine kanonische Interpretation \mathcal{I} mit der zugehörigen Funktion θ .

Zunächst wird gezeigt, dass die Menge $\text{kan}(\mathcal{T}_{\text{sat}})$ für eine konsistente \mathcal{ALC} -TBox nicht-leer ist.

Lemma 4.2.2. *Sei \mathcal{T}_{sat} eine vervollständigte \mathcal{ALC} -TBox. Es existiert eine kanonische Interpretation $\mathcal{I}, \theta \in \text{kan}(\mathcal{T}_{\text{sat}})$.*

Beweis. Es ist zu zeigen, dass eine Interpretation \mathcal{I} und eine Funktion $\theta : \Delta^{\mathcal{I}} \rightarrow \text{types}(\mathcal{T}_{\text{sat}})$ existiert, so dass die Bedingungen aus Definition 4.2.3 erfüllt sind.

Dazu wird für jedes H mit $H \sqsubseteq \perp \notin \mathcal{T}_{\text{sat}}$ genau ein Element d_H eingeführt.

$$\Delta^{\mathcal{I}} := \{d_H \mid H \sqsubseteq \perp \notin \mathcal{T}_{\text{sat}}\}$$

Für jedes H existiert nach Definition 4.2.1 eine nicht-leere Mengen von Typen T_H . Daher gibt es eine Funktion θ mit $\theta(d_H) := \tau_H(\in T_H)$. Mit dieser Funktion werden $A^{\mathcal{I}}$ und $r^{\mathcal{I}}$ für alle Konzept- und Rollennamen wie folgt definiert:

$$\begin{aligned} A^{\mathcal{I}} &:= \{d_H \mid A \in \tau_H\} \\ r^{\mathcal{I}} &:= \{(d_H, d_K) \mid K \in \tau_{H,r}\} \end{aligned}$$

Aus Lemma 4.2.1 3. folgt, dass $d_K \in \Delta^{\mathcal{I}}$ für alle $K \in \tau_{H,r}$. Damit erfüllen $A^{\mathcal{I}}$ und $r^{\mathcal{I}}$ die Bedingungen aus Definition 4.2.3 und es folgt $\mathcal{I}, \theta \in \text{kan}(\mathcal{T}_{\text{sat}})$. \square

Nachdem gezeigt wurde, dass die Menge $\text{kan}(\mathcal{T}_{\text{sat}})$ nicht-leer ist, wenn mindestens ein H mit $H \sqsubseteq \perp \in \mathcal{T}_{\text{sat}}$ existiert, wird nun bewiesen, dass alle $\mathcal{I}, \theta \in \text{kan}(\mathcal{T}_{\text{sat}})$ Modelle für \mathcal{T} sind. Dazu werden die Inferenzregeln aus Abbildung 4.1 im Beweis angewendet.

Lemma 4.2.3. *Sei $\mathcal{I}, \theta \in \text{kan}(\mathcal{T}_{\text{sat}})$. \mathcal{I} ist ein Modell für \mathcal{T} .*

Beweis. Sei $\mathcal{I} \in \text{kan}(\mathcal{T}_{\text{sat}})$ und θ die zugehörige Funktion nach Definition 4.2.3. Es wird gezeigt, dass alle Inklusionen aus \mathcal{T} durch \mathcal{I} erfüllt werden.

$\prod_{i=1}^n A_i \sqsubseteq M$: Sei $d \in A_i^{\mathcal{I}}$ für alle $i = 1, \dots, n$, mit $\theta(d) \in T_H$. Nach Lemma 4.2.1 2. existieren für alle i Inklusionen $H \sqsubseteq N_i \sqcup A_i \in \mathcal{T}_{\text{sat}}$ mit $A_i \in \theta(d)$ und $N_i \cap \theta(d) = \emptyset$. Aus $\{H \sqsubseteq N_i \sqcup A_i\}_{i=1}^n \in \mathcal{T}_{\text{sat}}$ und $\prod_{i=1}^n A_i \sqsubseteq M \in \mathcal{T}_{\text{sat}}$ folgt mit der Inferenzregel CR₁, dass $H \sqsubseteq \bigsqcup_{i=1}^n N_i \sqcup M \in \mathcal{T}_{\text{sat}}$. Nach Lemma 4.2.1 1. gilt $(\bigsqcup_{i=1}^n N_i \sqcup M) \cap \theta(d) \neq \emptyset$. Für alle i gilt $N_i \cap \theta(d) = \emptyset$. Daher ist $d \in M^{\mathcal{I}}$.

$A \sqsubseteq \exists r.B$: Sei $d \in A^{\mathcal{I}}$ mit $\theta(d) \in T_H$. Es gibt $H \sqsubseteq N \sqcup A \in \mathcal{T}_{\text{sat}}$ mit $A \in \theta(d)$ und $N \cap \theta(d) = \emptyset$ nach Lemma 4.2.1 2. Da CR₂ erschöpfend angewendet wurde ist auch $H \sqsubseteq N \sqcup \exists r.B \in \mathcal{T}_{\text{sat}}$. Da $N \cap \theta(d) = \emptyset$, ist $B \in \theta(d)_r$ nach (4.6). In $\theta(d)_r$ gibt es ein maximales K' , das B enthält. Nach (4.7) ist $K' \in \theta(d)_r$ und nach Definition 4.2.3 gibt es

ein $e \in \Delta^{\mathcal{I}}$ mit $\theta(e) \in T_{K'}$ und $(d, e) \in r^{\mathcal{I}}$. Weil B in der Konjunktion K' enthalten ist, folgt mit Lemma 4.2.1 4. $B \in \theta(e)$ und $e \in B^{\mathcal{I}}$ und damit auch $d \in (\exists r.B)^{\mathcal{I}}$.

$A \sqsubseteq \forall r.B$: Sei $d \in A^{\mathcal{I}}$ mit $\theta(d) \in T_H$ und $(d, e) \in r^{\mathcal{I}}$ mit $\theta(e) \in T_K$. Es ist zu zeigen, dass $e \in B^{\mathcal{I}}$ gilt. Es gibt $H \sqsubseteq N \sqcup A, H \sqsubseteq M \sqcup \exists r.K \in \mathcal{T}_{sat}$ mit $N \cap \theta(d) = \emptyset$ und $M \cap \theta(d) = \emptyset$. Wegen CR_4 ist $H \sqsubseteq N \sqcup M \sqcup \exists r.(K \sqcap B) \in \mathcal{T}_{sat}$. Nach (4.6) ist $K \sqcap B \in \theta(d)_r$. Nach Annahme ist $K \in \theta(d)_r$, da e ein r -Nachfolger von d ist. Daher ist K maximal in $\theta(d)_r$ nach (4.7). Da $K \sqcap B \in \theta(d)_r$ muss $B \in K$ gelten. Nach Lemma 4.2.1 4. folgt $e \in B^{\mathcal{I}}$.

□

Mit Hilfe der kanonischen Modelle kann gezeigt werden, dass \mathcal{T}_{sat} vollständig bezüglich $H \sqsubseteq_{\mathcal{T}} A$ ist.

Lemma 4.2.4. *Sei \mathcal{T} eine \mathcal{ALC} -TBox und A ein Konzeptname aus \mathcal{T} .*

1. *Wenn $H \sqsubseteq_{\mathcal{T}} \perp$, dann gilt auch $H \sqsubseteq \perp \in \mathcal{T}_{sat}$.*
2. *Wenn $H \sqsubseteq_{\mathcal{T}} A$, dann gilt auch $H \sqsubseteq A \in \mathcal{T}_{sat}$.*

Beweis. Für den Beweis wird das kanonische Modell $\mathcal{I} \in \text{kan}(\mathcal{T}_{sat})$ mit der Funktion θ aus dem Beweis von Lemma 4.2.2 benutzt. Mit diesem Modell werden die Behauptungen durch Widerspruch bewiesen.

Zu 1.:

Dazu wird angenommen, dass $H \sqsubseteq_{\mathcal{T}} \perp$ und $H \sqsubseteq \perp \notin \mathcal{T}_{sat}$ gilt. Wenn $H \sqsubseteq \perp \notin \mathcal{T}_{sat}$, dann gilt nach Lemma 4.2.1 4. $H \sqsubseteq \theta(d_H)$ und damit $d_H \in H^{\mathcal{I}}$. Da \mathcal{I}, θ nach Lemma 4.2.3 Modell für \mathcal{T} ist, ergibt sich ein Widerspruch zur Annahme $H \sqsubseteq_{\mathcal{T}} \perp$.

Zu 2.:

Sei A ein beliebiger Konzeptname und $H \sqsubseteq_{\mathcal{T}} A$ und $H \not\sqsubseteq_{\mathcal{T}} \perp$. Da \mathcal{I} Modell von \mathcal{T} ist, gilt $d_H \in A^{\mathcal{I}}$. Nach Definition (4.2.3) ist $A \in \theta(d_H)$. Nach Definition (4.5) gibt es eine minimale Disjunktion $N \sqcup A \in \mathcal{M}_H^{min}$ mit $H \sqsubseteq N \sqcup A \in \mathcal{T}_{sat}$. Durch Widerspruch soll gezeigt werden, dass es eine solche minimale Disjunktion gibt mit $N = \emptyset$ d. h. $H \sqsubseteq A \in \mathcal{T}_{sat}$. Angenommen $N \neq \emptyset$ für alle minimalen Disjunktionen $N \sqcup A$, die A enthalten. Es wird ein H -Typ τ'_H konstruiert, der Bedingung 1. und 2. aus Definition 4.2.1 erfüllt und A nicht enthält.

1. $\tau'_H := \emptyset; \mathcal{D}'_H := \mathcal{M}_H^{min}$
2. Es wird eine beliebige Menge $M = \{B_1, \dots, B_n\}$ aus \mathcal{D}'_H ausgewählt. Es wird genau ein Element B_i mit $i \in \{1, \dots, n\}$ und $B_i \neq A$ ausgewählt. Als nächstes werden die Mengen τ'_H und \mathcal{D}'_H aktualisiert.

3. B_i ist der in Schritt 2 ausgewählte Konzeptname. B_i wird zu τ'_H hinzugefügt und alle Mengen aus \mathcal{D}'_H , die B_i enthalten, werden aus \mathcal{D}'_H entfernt.

$$\begin{aligned}\tau'_H &:= \tau'_H \cup \{B_i\} \\ \mathcal{D}'_H &:= \mathcal{D}'_H \setminus \{M^{min} \in \mathcal{D}'_H \mid B_i \in M^{min}\}\end{aligned}$$

4. Wenn $\mathcal{D}'_H = \emptyset$ ist die Konstruktion beendet. Sonst wird mit Schritt 2 fortgefahren.

Auf Grund der Annahme, dass A nicht einzeln in einer minimalen Disjunktion vorkommen kann, kann in jedem Schritt ein entsprechendes Element $B_i \neq A$ ausgewählt werden. Daher terminiert die Konstruktion, weil \mathcal{D}'_H in jedem Durchlauf verkleinert wird. τ'_H hat nach Konstruktion mit jeder Menge aus \mathcal{M}_H^{min} einen nicht-leeren Schnitt. Somit ist 1. aus Definition 4.2.1 erfüllt. Als Nächstes ist zu zeigen, dass τ'_H nicht verkleinert werden kann. Sei dazu $B \in \tau'_H$. Dann gibt es eine Menge $M^{min} \in \mathcal{M}_H^{min}$, die im 2. Schritt ausgewählt wurde und B enthält. Für die zugehörige Disjunktion $N \sqcup B$ zu M^{min} gilt dann $\tau'_H \cap N = \emptyset$. Das bedeutet, dass τ'_H nicht verkleinert werden kann. Daher erfüllt τ'_H auch Bedingung 2. aus Definition 4.2.1.

Das Modell \mathcal{I} wird geändert, indem $\theta(d_H) := \tau'_H$ gesetzt wird. Weil $A \notin \tau'_H$ gilt, ist $d_H \notin A^\perp$. Das ist ein Widerspruch zu $H \sqsubseteq_{\mathcal{T}} A$ und es folgt $H \sqsubseteq A \in \mathcal{T}_{sat}$. \square

Aus diesem Lemma und der Definition des Vervollständigungsgraphen folgt die Korrektheit und Vollständigkeit der Beschriftungsfunktion.

Lemma 4.2.5. *Sei \mathcal{T}_{sat} eine vervollständigte \mathcal{ALC} -TBox und $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph. Es gilt $A \in S(H)$ genau dann, wenn $H \sqsubseteq_{\mathcal{T}} A$ und $H \not\sqsubseteq_{\mathcal{T}} \perp$.*

Beweis. Zuerst wird die Richtung „ \Leftarrow “ gezeigt. Angenommen $A \in S(H)$, dann ist $A \in \tau_H$ für alle $\tau_H \in T_H$. Dann gibt es Inklusionen der Form $H \sqsubseteq N \sqcup A \in \mathcal{T}_{sat}$, so dass $N \sqcup A \in \mathcal{M}_H^{min}$ (siehe (4.5)). Wird angenommen, dass A nicht einzeln in einer solchen Disjunktion vorkommt, dann kann wie im Beweis von Lemma 4.2.4 ein τ'_H konstruiert werden, so dass $A \notin \tau'_H$ gilt. Das ist ein Widerspruch zu der Annahme, dass A nicht einzeln in einer Disjunktion vorkommt, und damit folgt $H \sqsubseteq A \in \mathcal{T}_{sat}$.

Zum Beweis der Richtung „ \Rightarrow “ wird angenommen, dass $H \sqsubseteq_{\mathcal{T}} A$ gilt. Aus Lemma 4.2.4 folgt $H \sqsubseteq A \in \mathcal{T}_{sat}$ und damit $A \in \tau_H$ für alle $\tau_H \in T_H$. Es folgt $A \in S(H)$. \square

Abschließend wird die Korrektheit und Vollständigkeit der Kanten im Vervollständigungsgraph bewiesen.

Lemma 4.2.6. *Sei $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph zu einer \mathcal{ALC} -TBox \mathcal{T} . Wenn $(H, r, K) \in E$, dann $H \sqsubseteq_{\mathcal{T}} \exists r.K$.*

Beweis. Es wird ein beliebiges Modell \mathcal{J} betrachtet. Die Idee des Beweises besteht darin einem $e \in \Delta^{\mathcal{J}}$ mit $e \in H^{\mathcal{J}}$ einen passenden H -Typ zuzuordnen.

Behauptung 1. Sei \mathcal{J} ein Modell von \mathcal{T} und $e \in H^{\mathcal{J}}$. Es existiert ein $\tau_H \in \text{types}(\mathcal{T}_{sat})$, so dass für alle $A \in \tau_H$, $e \in A^{\mathcal{J}}$ gilt.

Beweis. Sei $e \in H^{\mathcal{J}}$ für ein Modell \mathcal{J} von \mathcal{T} und M_e die folgende Menge von Konzeptnamen.

$$M_e := \{A \mid e \in A^{\mathcal{J}}, A \in M, M \in \mathcal{M}_H^{min}\}$$

Sei $M \in \mathcal{M}_H^{min}$. Es gilt $H \sqsubseteq M \in \mathcal{T}_{sat}$. Da \mathcal{J} ein Modell für \mathcal{T}_{sat} ist, folgt $e \in M^{\mathcal{J}}$. Daher ist der Schnitt von M_e mit allen minimalen Disjunktionen für H nicht-leer und nach Definition 4.2.1 existiert ein τ_H mit $\tau_H \subseteq M_e$. Daraus folgt die Behauptung. \square

Eine ähnliche Behauptung kann für die zugehörige Menge $\tau_{H,r}$ gezeigt werden.

Behauptung 2. Sei \mathcal{J} ein Modell von \mathcal{T} und $e \in H^{\mathcal{J}}$. Es existiert ein $\tau_{H,r}$, so dass für alle $K' \in \tau_{H,r}$, $e \in (\exists r.K')^{\mathcal{J}}$ erfüllt ist.

Beweis. Sei τ_H der H -Typ aus Behauptung 1 und $K' \in \tau_{H,r}$. Dann gibt es eine Inklusion der Form

$$H \sqsubseteq N_0 \sqcup N_1 \sqcup \dots \sqcup N_n \sqcup \exists r.(B_0 \sqcap B_1 \sqcap \dots \sqcap B_n) \quad (4.9)$$

in \mathcal{T}_{sat} mit $K' = \prod_{i=0}^n B_i$ und $(\bigcup_{i=0}^n N_i) \cap \tau_H = \emptyset$. Wenn die Inklusion (4.9) in \mathcal{T}_{sat} enthalten ist, dann gibt es auch die folgenden Inklusionen in \mathcal{T}_{sat} :

$$\begin{aligned} H &\sqsubseteq N_0 \sqcup A_0, A_0 \sqsubseteq \exists r.B_0 \\ H &\sqsubseteq N_1 \sqcup A_1, A_1 \sqsubseteq \forall r.B_1 \\ &\dots \\ H &\sqsubseteq N_n \sqcup A_n, A_n \sqsubseteq \forall r.B_n \end{aligned}$$

aus denen (4.9) mit Hilfe der Regeln CR_2 und CR_4 abgeleitet wurde. Aus $(\bigcup_{i=0}^n N_i) \cap \tau_H = \emptyset$ und Lemma 4.2.1 1. folgt $A_i \in \tau_H$ für alle $i = 0, \dots, n$. Mit Behauptung 1 folgt $e \in (A_i)^{\mathcal{J}}$ für alle i . Da \mathcal{J} ein Modell ist, gilt entsprechend $e \in (\exists r.B_0)^{\mathcal{J}}$ und $e \in (\forall r.B_j)^{\mathcal{J}}$ für alle $j = 1, \dots, n$. Daraus folgt $e \in (\exists r.K')^{\mathcal{J}}$. \square

Sei $(H, r, K) \in E$, \mathcal{J} ein Modell von \mathcal{T} mit $e \in H^{\mathcal{J}}$ und τ_H der zugehörige Typ aus Behauptung 1. Nach Definition 4.2.2 gibt es ein $K' \in \tau_{H,r}$ mit $K \subseteq S(K')$. Mit Lemma 4.2.5 und 4.2.4 gilt für alle Konzeptnamen $A \in K : K' \sqsubseteq A \in \mathcal{T}_{sat}$. Daraus folgt $K' \sqsubseteq_{\mathcal{T}} K$. Aus Behauptung 2 folgt $e \in (\exists r.K')^{\mathcal{J}}$. Daraus folgt $e \in (\exists r.K)^{\mathcal{J}}$ und $H \sqsubseteq_{\mathcal{T}} \exists r.K$, da e und \mathcal{J} beliebig gewählt waren. \square

Im nächsten Lemma wird die Vollständigkeit des Graphen bezüglich der Kanten bewiesen.

Lemma 4.2.7. *Sei \mathcal{T} eine \mathcal{ALC} -TBox und $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der zugehörige Vervollständigungsgraph. Wenn $H \sqsubseteq_{\mathcal{T}} \exists r.K$ und $H \not\sqsubseteq_{\mathcal{T}} \perp$, dann existiert $(H, r, K) \in E$.*

Beweis. Die Behauptung wird indirekt bewiesen. Dazu wird angenommen, dass $H \sqsubseteq_{\mathcal{T}} \exists r.K$ gilt, aber die geforderte Kante im Vervollständigungsgraphen *nicht* existiert. Mit dieser Annahme kann dann gezeigt werden, dass ein kanonisches Modell nach Definition 4.2.3 existiert, das $H \sqsubseteq_{\mathcal{T}} \exists r.K$ widerlegt.

Die geforderte Kante (H, r, K) existiert nach Definition 4.2.2 genau dann, wenn für alle Typen $\tau_H \in T_H$ ein $K' \in \tau_{H,r}$ mit $K \subseteq S(K')$ existiert. Es wird das Gegenteil angenommen. Dann existiert ein $\tau'_H \in T_H$ mit $\tau'_{H,r} = \{K'_1, \dots, K'_n\}$, so dass für jedes K'_i mit $i = 1, \dots, n$ ein Konzeptnamen $A_i \in K$ existiert mit $K'_i \sqsubseteq A_i \notin \mathcal{T}_{sat}$ (d. h. es gilt $K \not\subseteq S(K'_i)$ nach Lemma 4.2.4). Daraus folgt, dass für alle i ein Typ $\tau'_{K'_i}$ existiert mit $A_i \notin \tau'_{K'_i}$. Mit Hilfe dieser beschriebenen Typen wird ein kanonisches Modell $\mathcal{I} \in \text{kan}(\mathcal{T}_{sat})$ mit einer Funktion $\theta : \Delta^{\mathcal{I}} \rightarrow \text{types}(\mathcal{T}_{sat})$ definiert. Die Domäne von \mathcal{I} wird wie folgt definiert:

$$\Delta^{\mathcal{I}} := \{d_H, d_{K'_1}, \dots, d_{K'_n}\} \cup \{d_{H'} \mid H' \sqsubseteq \perp \notin \mathcal{T}_{sat}\}$$

Den Elementen aus $\{d_H, d_{K'_1}, \dots, d_{K'_n}\}$ sollen die oben beschriebenen Typen τ'_H bzw. $\tau'_{K'_i}$ zugewiesen werden, um den Widerspruch zu erzeugen. Die Elemente der Menge $\{d_{H'} \mid H' \sqsubseteq \perp \notin \mathcal{T}_{sat}\}$ werden benötigt um ein Modell zu erhalten. Diesen Elementen wird ein beliebiger Typ aus $T_{H'}$ zugewiesen. Für θ gilt entsprechend $\theta(d_H) := \tau'_H$ und $\theta(d_{K'_i}) := \tau'_{K'_i}$ für alle i . Für alle übrigen Elemente $d_{H'} \in \Delta^{\mathcal{I}}$ gelte $\theta(d_{H'}) \in T_{H'}$. Eine solche Funktion θ existiert, weil nach Definition der Domäne $H' \sqsubseteq \perp \notin \mathcal{T}_{sat}$ gilt und damit die Mengen $T_{H'}$ nicht leer sind.

Die Interpretation der Konzeptnamen A und Rollen r ergibt sich wie folgt: Sei $\theta(d_{\hat{H}}) = \tau_{\hat{H}}$. Dann wird $\tau_{\hat{H},r}$ mit $\theta(d_{\hat{H}})_r$ bezeichnet.

$$\begin{aligned} A^{\mathcal{I}} &:= \{d \in \Delta^{\mathcal{I}} \mid A \in \theta(d)\} \\ r^{\mathcal{I}} &:= \{(d_{\hat{H}}, d_{\hat{K}}) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid \hat{K} \in \theta(d_{\hat{H}})_r\} \end{aligned}$$

\mathcal{I} erfüllt die Bedingungen aus 4.2.3 und ist damit nach Lemma 4.2.3 ein Modell für \mathcal{T} . Nach Konstruktion gilt $d_H \in H^{\mathcal{I}}$. Sei $(d_H, d_{K'}) \in r^{\mathcal{I}}$. Dann gibt es nach Konstruktion ein $i \in \{1, \dots, n\}$ mit $K' \in \tau'_{H,r}$ und $\theta(d_{K'}) = \tau'_{K'_i}$. $\tau'_{K'_i}$ war so gewählt, dass ein $A_i \in K$ existiert mit $A_i \notin \tau'_{K'_i}$. Daraus folgt $d_{K'} \notin A_i$ und damit auch $d_{K'} \notin K^{\mathcal{I}}$ und $d_H \notin (\exists r.K)^{\mathcal{I}}$. \square

Die Ergebnisse werden in dem folgenden Satz zusammengefasst:

Satz 4.2.1. *Sei \mathcal{T} eine \mathcal{ALC} -TBox, \mathcal{T}_{sat} die Vervollständigung, $\mathcal{G}_{\mathcal{T}} = (V, E, S)$ der Vervollständigungsgraph, H, K Konjunktionen von Konzeptnamen aus \mathcal{T} und A ein Konzeptname aus \mathcal{T} . Es gelten die folgenden Eigenschaften:*

- $H \sqsubseteq_{\mathcal{T}} A$ und $H \not\sqsubseteq_{\mathcal{T}} \perp$ genau dann, wenn $A \in S(H)$ und
- $H \sqsubseteq_{\mathcal{T}} \exists r.K$ und $H \not\sqsubseteq_{\mathcal{T}} \perp$ genau dann, wenn $(H, r, K) \in E$.

Der berechnete Vervollständigungsgraph erfüllt alle Eigenschaften bezüglich der TBox, wie der in Definition 2.1.4 definierte \mathcal{EL} -Vervollständigungsgraph. Die Methode zur Berechnung des Rollentiefen- beschränkten \mathcal{EL} -LCS und die Ergebnisse aus Kapitel 3 können damit auch bezüglich einer generellen \mathcal{ALC} -TBox angewendet werden.

4.3 Komplexität der Vervollständigung

In diesem Abschnitt wird die Komplexität der Berechnung des Vervollständigungsgraphen zu einer \mathcal{ALC} -TBox \mathcal{T} untersucht.

Die Berechnung des Vervollständigungsgraphen besteht aus drei Schritten. Der erste Schritt ist die Berechnung, der in Abschnitt 4.1 beschriebenen Normalform. Diese kann in polynomieller Zeit berechnet werden [SKH11]. Im zweiten Schritt wird die normalisierte \mathcal{ALC} -TBox mit Hilfe von Inferenzregeln vervollständigt. Die Größe der vervollständigten TBox ist exponentiell in der Größe der ursprünglichen TBox [SKH11]. Im dritten Schritt wird aus der vervollständigten TBox der Vervollständigungsgraph berechnet. Der Graph enthält für jede Konjunktion H mit $H \not\sqsubseteq_{\mathcal{T}} \perp$ genau einen Knoten. Die Anzahl der Knoten ist somit exponentiell in der Anzahl der Konzeptnamen aus \mathcal{T} . Für die Berechnung der Beschriftungsfunktion und der Kanten des Graphen wird für jeden Knoten H die Menge der Typen T_H bestimmt. Für eine Konjunktion H kann es exponentiell viele Typen in der Anzahl der Konzeptnamen in \mathcal{T} geben. Allgemein formuliert ist ein Typ eine minimale Menge, die einen nicht-leeren Schnitt mit einer Reihe von Mengen hat. In der Literatur ist für eine solche Menge der Begriff *minimal hitting set* bekannt. Die Berechnung aller Typen ist außerdem äquivalent zu dem Problem *transversal hypergraph generation* [Hag09]. Das Problem für eine gegebene Menge T' zu entscheiden ob $T' = T_H$ gilt, ist demnach in coNP . Die exakte Komplexität ist jedoch ein offenes Problem [Hag09].

Kapitel 5

Zusammenfassung und Ausblick

In dieser Arbeit wurde das Existenzproblem des \mathcal{EL} -LCS bezüglich einer generellen TBox mit deskriptiver Semantik untersucht.

Basierend auf der Charakterisierung der Subsumption aus [LW10, LPW10] wurde eine notwendige und hinreichende Bedingung für die Existenz des LCS angegeben. Weiterhin wurde gezeigt, dass es entscheidbar ist, ob eine \mathcal{EL} -Konzeptbeschreibung existiert, die diese Bedingung erfüllt. Das vorgestellte Entscheidungsverfahren basiert auf kanonischen Modellen, die mit Hilfe des Algorithmus zur Vervollständigung von generellen \mathcal{EL} -TBoxen berechnet werden können [BBL05], und Simulationsrelationen zwischen diesen Modellen.

Außerdem wurde eine polynomielle obere Schranke für die Rollentiefe des LCS angegeben.

Des Weiteren wurde eine Methode zur Berechnung eines \mathcal{EL} -Vervollständigungsgraphen für eine generelle \mathcal{ALC} -TBox vorgestellt und die Komplexität des Berechnungsverfahrens untersucht. Damit ist es möglich das Berechnungsverfahren für den Rollentiefen- beschränkten \mathcal{EL} -LCS aus [PT10b] auch bezüglich genereller \mathcal{ALC} -TBoxen anzuwenden. Auf Grund der sehr hohen Komplexität ist es jedoch unklar, ob das Berechnungsverfahren in der Praxis anwendbar ist.

In praktischen Anwendungen von Generalisierungsinferenzen sind der LCS und das MSC sehr eng miteinander verbunden. Eine Erweiterung der vorgestellten Methoden auf das Existenzproblem des MSC sind daher für zukünftige Forschungsarbeiten interessant. Beide Inferenzen sind strukturell sehr ähnlich und wie für den LCS, existiert auch für das MSC bereits eine analoge Charakterisierung der zu Grunde liegenden Standardinferenz [LPW10]. Eine Erweiterung der gezeigten Resultate auf das Existenzproblem des MSC ist daher prinzipiell möglich.

Der LCS und das MSC sind auch für ausdrucksstärkere Beschreibungslogiken wie $\mathcal{AL}\mathcal{E}$ und Erweiterungen von \mathcal{EL} definiert. Wie für \mathcal{EL} existieren auch für diese BLen Verfahren zur Approximation des LCS [MET11, BST07]. Erweiterungen der vorgestellten Ansätze in diese Richtung sind daher ebenfalls interessant.

Literaturverzeichnis

- [Baa03] BAADER, Franz: Computing the least common subsumer in the description logic \mathcal{EL} w.r.t. terminological cycles with descriptive semantics. In: *Proceedings of the 11th International Conference on Conceptual Structures, ICCS 2003* Bd. 2746, Springer-Verlag, 2003 (Lecture Notes in Artificial Intelligence), S. 117–130
- [Baa04] BAADER, Franz: A Graph-Theoretic Generalization of the Least Common Subsumer and the Most Specific Concept in the Description Logic \mathcal{EL} . In: HROMKOVIC, J. (Hrsg.) ; NAGL, M. (Hrsg.): *Proceedings of the 30th International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2004)* Bd. 3353. Bad Honnef, Germany : Springer-Verlag, 2004 (Lecture Notes in Computer Science), S. 177–188
- [BBL05] BAADER, Franz ; BRANDT, Sebastian ; LUTZ, Carsten: Pushing the \mathcal{EL} Envelope. In: *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05*. Edinburgh, UK : Morgan-Kaufmann Publishers, 2005
- [BCM⁺03] BAADER, Franz (Hrsg.) ; CALVANESE, Diego (Hrsg.) ; MCGUINNESS, Deborah L. (Hrsg.) ; NARDI, Daniele (Hrsg.) ; PATELSCHNEIDER, Peter F. (Hrsg.): *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003 . – ISBN 0-521-78176-0
- [BKM99] BAADER, Franz ; KÜSTERS, Ralf ; MOLITOR, Ralf: Computing Least Common Subsumers in Description Logics with Existential Restrictions. In: DEAN, T. (Hrsg.): *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI'99)*, Morgan Kaufmann, 1999, S. 96–101
- [BST04] BAADER, Franz ; SERTKAYA, Baris ; TURHAN, Anni-Yasmin: Computing the Least Common Subsumer w.r.t. a Background Terminology. In: *Proceedings of the 2004 International Workshop on Description Logics (DL2004)*, 2004 (CEUR-WS)

- [BST07] BAADER, Franz ; SERTKAYA, Barış ; TURHAN, Anni-Yasmin: Computing the Least Common Subsumer w.r.t. a Background Terminology. In: *Journal of Applied Logic* 5 (2007), Nr. 3, S. 392–420
- [CBH92] COHEN, William W. ; BORGIDA, Alexander ; HIRSH, Haym: Computing Least Common Subsumers in Description Logics. In: *AAAI*, 1992, S. 754–760
- [Hag09] HAGEN, Matthias: Lower bounds for three algorithms for transversal hypergraph generation. In: *Discrete Appl. Math.* 157 (2009), April, Nr. 7, 1460–1469. <http://dx.doi.org/10.1016/j.dam.2008.10.004>. – DOI 10.1016/j.dam.2008.10.004. – ISSN 0166–218X
- [HKP⁺09] HITZLER, Pascal (Hrsg.) ; KRÖTZSCH, Markus (Hrsg.) ; PARSIA, Bijan (Hrsg.) ; PATEL-SCHNEIDER, Peter F. (Hrsg.) ; RUDOLPH, Sebastian (Hrsg.): *OWL 2 Web Ontology Language: Primer*. W3C Recommendation, 27 October 2009. – Available at <http://www.w3.org/TR/owl2-primer/>
- [LPW10] LUTZ, Carsten ; PIRO, Robert ; WOLTER, Frank: Enriching \mathcal{EL} -Concepts with Greatest Fixpoints. In: *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI10)*, IOS Press, 2010
- [LW10] LUTZ, Carsten ; WOLTER, Frank: Deciding inseparability and conservative extensions in the description logic \mathcal{EL} . In: *Journal of Symbolic Computation* 45 (2010), Nr. 2, S. 194–228
- [MET11] MENDEZ, Julian ; ECKE, Andreas ; TURHAN, Anni-Yasmin: Implementing completion-based inferences for the \mathcal{EL} - family. In: ROSATI, Riccardo (Hrsg.) ; RUDOLPH, Sebastian (Hrsg.) ; ZAKHARYASCHEV, Michael (Hrsg.): *Proceedings of the international Description Logics workshop* Bd. 745, CEUR, 2011
- [PT10a] PEÑALOZA, Rafael ; TURHAN, Anni-Yasmin: Completion-based computation of least common subsumers with limited role-depth for \mathcal{EL} and Prob- \mathcal{EL}^{01} / Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology. Germany, 2010 (LTCS-10-02). – LTCS-Report. – See <http://lat.inf.tu-dresden.de/research/reports.html>.
- [PT10b] PEÑALOZA, Rafael ; TURHAN, Anni-Yasmin: Role-depth Bounded Least Common Subsumers by Completion for \mathcal{EL} - and Prob- \mathcal{EL} -TBoxes. In: HAARSLEV, V. (Hrsg.) ; TOMAN, D. (Hrsg.)

- ; WEDDELL, G. (Hrsg.): *Proc. of the 2010 Description Logic Workshop (DL'10)* Bd. 573, 2010 (CEUR-WS)
- [SKH11] SIMANČÍK, František ; KAZAKOV, Yevgeny ; HORROCKS, Ian: Consequence-Based Reasoning beyond Horn Ontologies. In: WALSH, Toby (Hrsg.): *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11)*, AAAI Press/IJCAI, 2011, S. 1093–1098