

Und nun ...

Beschreibungslogik: Einführung

Sommersemester 2018 Thomas Schneider

AG Theorie der künstlichen Intelligenz (TdKI)

<http://tinyurl.com/ss18-bl>

① Organisatorisches

② Einleitung

③ Vorlesungsüberblick

Organisatorisches

Zeit und Ort

Mo. 10–12 MZH 6210

Mi. 16–18 MZH 4140

Vortragender

Thomas Schneider

Cartesium, Raum 1.56

Tel. (218) 64432

ts[ÄT]cs.uni-bremen.de

Fragen sind jederzeit willkommen!**Position im Curriculum**

Master-Basis, Modul-Bereich Theorie, Master-Profile KIKR, DMI

Organisatorisches

Form K4 (in der Regel 3V, 1Ü)Fragen und Diskussion **jederzeit erwünscht.****Voraussetzungen**Grundkenntnisse aus Theoret. Informatik 1+2 hilfreich,
aber nicht zwingend erforderlich**Vorlesungsmaterial:**

- Folien und Aufgabenblätter: tinyurl.com/ss18-bl
- Folien werden online gestellt, enthalten aber nicht alle Details. (Beweise, Beispiele etc. von der Tafel bitte mitschreiben.)
- Inoffizielle Vorlesungsmitschrift von Teilnehmern (SoSe 16/17): <https://github.com/sasjonge/bl-lecture-notes>
- Literatur: (allgemein) siehe Ende dieses Foliensatzes (speziell) wird bei jedem Kapitel bekannt gegeben

Prüfungsmodalitäten

Übungsaufgaben & Fachgespräch:

- Übungsaufgaben ca. jede zweite Woche; voraussichtlich 6 Blätter, mit Zusatzaufgaben
- Werden in Gruppen (2–3 Personen) bearbeitet, abgegeben und korrigiert – jede_r muss mindestens einmal vorrechnen
- Aus der erreichten Gesamtpunktzahl aller Blätter ergibt sich die vorläufige Note für diesen Kurs
- Fachgespräche am Ende des Semesters (Prüfungsleistung, Änderung der Note möglich)
Voraussetzung: insgesamt 50 % der Punkte in Übungsaufgaben

oder

Mündliche Prüfung

Wiederholungsregelungen auf der nächsten Folie ...

Übungsbetrieb

Terminübersicht (geplant)

Blatt	Erscheinen (geplant)	Abgabe	Besprechung, Übungstermin
1	Fr. 6. 4.	Mo. 16. 4.	Mi. 18. 4.
2	Mo. 16. 4.	Mo. 30. 4.	Mi. 2. 5.
3	Mo. 7. 5.	Mo. 21. 5.*	Mi. 23. 5.
4	Mo. 21. 5.	Mo. 4. 6.	Mi. 6. 6.
5	Mo. 4. 6.	Mo. 18. 6.	Mi. 20. 6.
6	Mo. 18. 6.	Mo. 2. 7.	Mi. 4. 7.

*Pfingstmontag

Prüfungsmodalitäten

Wiederholungsregelungen

- Fachgespräch nicht bestanden?
→ 1 Wiederholungsversuch im selben Semester möglich
- Weitere Wiederholungsversuche (wenn nötig):
 - mündliche Prüfung in den folgenden 4 Semestern (je 1 Versuch pro Semester)
 - oder ggf. im SoSe 2019 erneute Teilnahme am Übungsbetrieb (wenn ich die Veranstaltung wieder halte)

Dies ist bereits Deine Wiederholung?

- Erneute Teilnahme am Übungsbetrieb ist möglich. Punkte aus vergangenen Semestern werden *nicht* anerkannt.
- Bei Nutzung früherer eigener Lösungen: Quellenangabe nötig

Und nun ...

- 1 Organisatorisches
- 2 Einleitung
- 3 Vorlesungsüberblick

Wissensrepräsentation

Ziel: allgemeines Hintergrundwissen dem Rechner verfügbar machen

Grundlage für:

- „Intelligente Systeme“, die Daten nicht nur verarbeiten, sondern auch interpretieren können
- Zahlreiche Teildisziplinen der künstlichen Intelligenz, wie z. B. Robotik, automatisches Planen, usw.
- ein „semantisches Web“, in dem Suchmaschinen und andere Agenten Webinhalte interpretieren können (im Gegensatz zu schlüsselwortbasierter Suche)

Wissensrepräsentation

Das Ziel der Wissensrepräsentation etwas konkreter:

Entwicklung von Formalismen, mittels derer Wissen über die Welt in abstrakter Weise beschrieben werden kann und die effektiv verwendet werden können, um **intelligente** Anwendungen zu realisieren.

[Ronald J. Brachman & Daniele Nardi 2003]

Dies ist **anwendungsorientierte** Sicht auf WR, keine philosophische oder kognitionswissenschaftliche.

Wissensrepräsentation – ein Beispiel

Betrachte die medizinische Datenbank

Patient(p_1)	Patient(p_2)
diagnose(p_1, d_1)	diagnose(p_2, d_2)
Tachykardie(d_1)	Hypertonie(d_2)

und die Anfrage:

$\text{Patient}(x) \wedge \exists y (\text{diagnose}(x, y) \wedge \text{Herz-Kreislauf-Erkrankung}(y))$

also: „gib mir alle Patient_innen mit Herz-Kreislauf-Erkrankungen“

Ein klassisches Datenbanksystem liefert **keine Antworten**, denn es **fehlt das Wissen**, dass Tachykardie und Hypertonie HKE sind.

Beispiel-Repräsentation

The screenshot shows the website of the University of Bremen. At the top, there is a navigation bar with the university logo and name, and links for 'Universität', 'Studium', 'Forschung', and 'Kooperationen'. Below the navigation bar is a large banner image of a group of people, with a text overlay that reads 'Bremer Studienpreis 2017 vergeben' and a quote: 'Wir brauchen Ihre Kompetenz und Bereitschaft zur Reflexibilität. Wir brauchen Sie als kritische Geister.' Below the banner is a section titled 'Aktuelles' with two smaller images: one showing hands holding a 'CROWD INVEST' sign, and another showing a man speaking. The date '29.03.2018' is visible below the images.

Typische Webpage besteht aus:

- Rendering-Information (z. B. Schriftgröße, -farbe)
- Hyperlinks auf verwandte Inhalte

Inhalt ist gut zugänglich für Menschen, aber kaum für Computer

Beispiel-Repräsentation

Uns zugängliche Informationen ...

- Universität Bremen
- Deutsch, Englisch
- Bremer Studienpreis 2017 vergeben
- Konrektor Andreas Breiter und weitere Personen
- Fotograf Harald Rehling
- Aktuelles ...
- Kontakt:
Universität Bremen
Bibliothekstraße 1
28359 Bremen
Tel.: +49 421 218-1

Beispiel-Repräsentation

Lösung: XML-Markup mit „sinnvollen“ Tags?

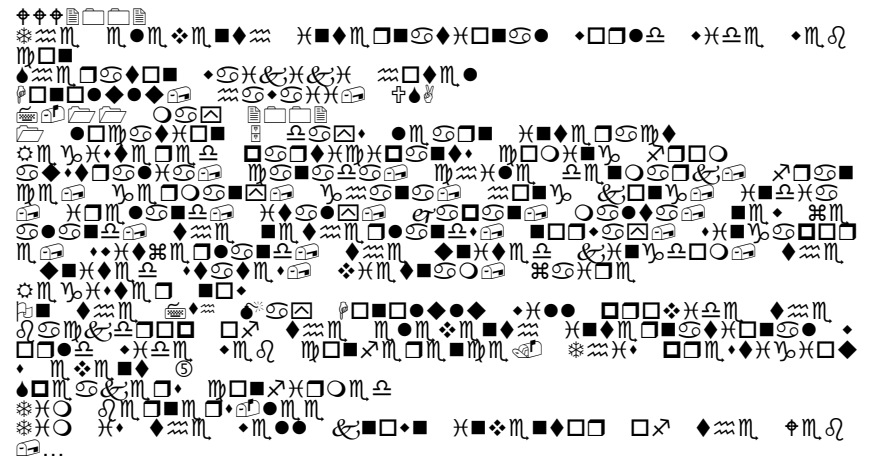
```

<university>
  <logo>
  <address>
  <topic>
  <topic>
  <topic>
  <contact>

```

Beispiel-Repräsentation

Dem Computer zugängliche Informationen ...



Beispiel-Repräsentation

Aber was, wenn ...

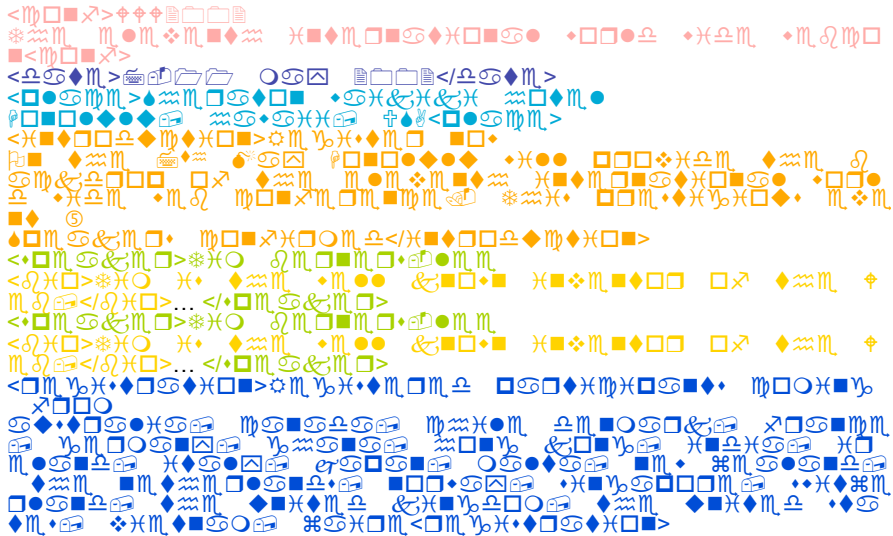
```

<university>
  <image>
  <activity>
  <activity>
  <activity>
  <contact>

```

Beispiel-Repräsentation

Die Maschine sieht immer nur ...



Ein Fall für Semantik!

- shoulder_catches_during_movement
- shoulder_feels_like_it_will_slip_out_of_place
- shoulder_joint_feels_like_it_may_slip_out_of_place
- shoulder_joint_pain_better_after_rest
- shoulder_joint_pain_causes_difficulty_lying_on_affected_side
- shoulder_joint_pain_causes_inability_to_sleep
- shoulder_joint Maschinenlesbar
- shoulder_joint aber nicht
- shoulder_joint für Maschinen „verständlich“
- shoulder_joint_pain_improved_by_medication
- shoulder_joint_pain_improves_during_exercise_returns_later
- shoulder_joint_pain_incr_by_raising_arm_above_shoulder_level
- shoulder_joint_pain_increased_by
- shoulder_joint_pain_increased_by_lifting
- shoulder_joint_pain_increased_by_moving_arm_across_chest
- ⋮

Ein Fall für Semantik!

Externe Vereinbarung über Bedeutung von Annotationen?

- z. B. **Dublin Core** zum Annotieren bibliographischer Info
 ↳ Einigung auf Bedeutung einer Menge von Annotation-Tags
- (!) ist unflexibel und
- (!) hat nur begrenztes Ausdrucksvermögen (nur simple Terme)

Verwendung von Vokabularen oder Ontologien, um Bedeutung von Annotationen zu spezifizieren

- Ontologien stellen Vokabular von Termen zur Verfügung
- Neue Terme können durch Kombination existierender Terme gebildet werden
- Bedeutung der Terme ist per Semantik **eindeutig festgelegt**

... nicht nur im Web, sondern allgemein in Informationssystemen

Wissensrepräsentation

Hier nochmal das Ziel der Wissensrepräsentation:

Entwicklung von **Formalisten**, mittels derer Wissen über die Welt in **abstrakter** Weise beschrieben werden kann und die **effektiv** verwendet werden können, um **intelligente** Anwendungen zu realisieren.

[Ronald J. Brachman & Daniele Nardi 2003]

Dies ist **anwendungsorientierte** Sicht auf WR, keine philosophische oder kognitionswissenschaftliche.

Mögliche Formalismen

Prädikatenlogik erster Stufe (FO)

- Formale Syntax, deklarative Semantik
- Hohe, dennoch beschränkte Ausdrucksstärke
- Sehr hohe Berechnungskomplexität: unentscheidbar

Aussagenlogik (AL)

- Formale Syntax, deklarative Semantik
- sehr beschränkte Ausdrucksstärke
- relativ geringe Berechnungskomplexität
- effektive Schlussfolgerungssysteme („SAT-Solver“) verfügbar

Beschreibungslogik (BL)

ist Kompromiss bzgl. Ausdrucksstärke und Komplexität

Ursprünge von BL

Historisch hervorgegangen sind BLen aus ...

- *structured inheritance networks* [Ronald J. Brachman 1978]
- einem bekannten historischen WR-System namens **KL-ONE** [Ronald J. Brachman & James G. Schmolze 1985]

Diese frühen Formalismen

- waren nicht logikbasiert
- hatten keine formale Syntax und Semantik
- haben dennoch fast alle grundlegenden Ideen heutiger BLen eingeführt

Beschreibungslogik

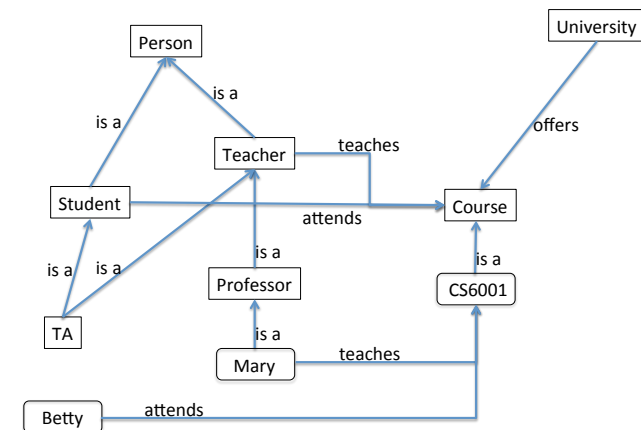
... ist wichtige Familie von Formalismen zur Wissensrepräsentation

Andere Formalismen: siehe KI-Vorlesung

BLen sind eine **Familie** von Logiken

- zur Repräsentation konzeptuellen Wissens
- mit entscheidbaren Schlussfolgerungsproblemen
- für die viele effiziente Reasoner zur Verfügung stehen

Ein solches Netzwerk



- Was genau **repräsentiert**/besagt es?
- Was kann man **schlussfolgern**? \rightsquigarrow Ist Betty ein_e Student_in?

Problem: fehlende Semantik

Terminologisches Wissen

DLs wurden geschaffen, um **terminologisches** oder **konzeptuelles Wissen** zu repräsentieren

- Einem Anwendungsfeld zugrunde liegende Terminologie (Begriffswelt) wird formalisiert und in einer **TBox** gespeichert
 \rightsquigarrow erlaubt, Schlussfolgerungen über **Konzepte** zu ziehen, z. B.:
 - Kann es Säugetiere geben?
 - Ist jedes Säugetier ein Tier?
 - Sind Frösche Reptilien?
- Fakten über Individuen werden in einer **ABox** gespeichert
 \rightsquigarrow Schlussfolgerungen über **Individuen & Konzepte**, z. B.:
 - Sind meine Fakten konsistent mit meiner Terminologie?
 - Ist Kermit ein Frosch?

Beschreibungslogik

Zentrale Elemente der Wissensrepräsentation mit BL:

- **TBoxen** (terminologische Boxen)
 Definieren Konzepte und setzen diese **zueinander in Beziehung**
 Konzeptdefinition z. B.

$$\text{Student} \equiv \text{Mensch} \sqcap \exists \text{hört.Vorlesung}$$

Allgemeines Hintergrundwissen / Constraint z. B.

$$\text{Student} \sqcap \text{Vorlesungssaal} \sqsubseteq \perp$$

Beschreibungslogik

Zentrale Elemente der Wissensrepräsentation mit BL:

- **Konzepte**
 Konzept beschreibt **Klasse von Objekten**, z. B. Bereich Universität:
 Vorlesung, Student, Universität, Fachbereich, etc.
 Können durch logische Ausdrücke (Formeln) beschrieben werden:

$$\text{Mensch} \sqcap \exists \text{hört.Vorlesung}$$
- **Rollen**
 ... sind **binäre Relationen** zwischen Objekten
 hört, lehrt, istTeilVon, etc.
 können (meist) nicht durch komplexe Ausdrücke beschrieben werden

Beschreibungslogik

Zentrale Elemente der Wissensrepräsentation mit BL:

- **ABoxen** (assertionale Boxen)
 Beschreiben **Individuen** (= Objekte) und deren **Eigenschaften**, z. B.

$$\text{Student}(\text{hans})$$

$$\text{Vorlesung} \sqcap \exists \text{hatThema.InformatikThema}(\text{blv})$$

$$\text{hört}(\text{hans}, \text{blv})$$

Ontologien

Ein sehr populärer Ansatz zur Wissensrepräsentation ist die Verwendung von **Ontologien**.

Ontologien dienen der Repräsentation von konzeptuellem Wissen, sehr ähnlich den TBoxen der Beschreibungslogik.

Im Prinzip kann eine Ontologie in beliebiger (meist logischer) Sprache verfasst werden.

Besonders populär ist aber **OWL: the Web Ontology Language**

- Standardisiert vom W3C (World Wide Web Consortium)
- Zugeschnitten aufs Web: XML-Syntax etc.
- Im Wesentlichen eine Beschreibungslogik

Beispiele für Ontologien

Biologie

- **GO** (Gene Ontology)

kontrolliertes Vokabular von Begriffen relevant für Genetik



- **Bioportal** <http://bioportal.bioontology.org/>

Website für Zugriff auf
688 biomedizinische Ontologien
(Stand 4. 4. 2018)



Semantisches Web

- Daten werden mittels Termen annotiert, die in TBox definiert werden
- Beim Stellen von Anfragen wird TBox verwendet



Beispiele für Ontologien

Medizin

- **SNOMED CT**

(Systematized Nomenclature of Medicine – Clinical Terms)

- klinische Terminologie, international verwendet
- 327.000 Konzepte (Stand 4. 4. 2018)



- **NCI Thesaurus** (NCI = National Cancer Institute of the USA)

- Vokabular für Krankenhausversorgung, Grundlagen- und praxisbezogene Forschung, Information der Öffentlichkeit, Administration



- 138.000 Konzepte (Stand 4. 4. 2018)

- **ICD-11** (International Classification of Diseases)

wird weltweit (auch in DE)
für Gesundheitsstatistiken verwendet



Beispielontologie SNOMED



Im Gesundheitswesen ist standardisierte Terminologie wichtig.

Beispiel: Austausch medizinischer Daten zwischen Ärzten, Krankenhäusern, Krankenkassen, etc.

Es sollten

- dieselben Begriffe (Konzepte) verwendet werden
- verwendete Begriffe in derselben Weise verstanden werden

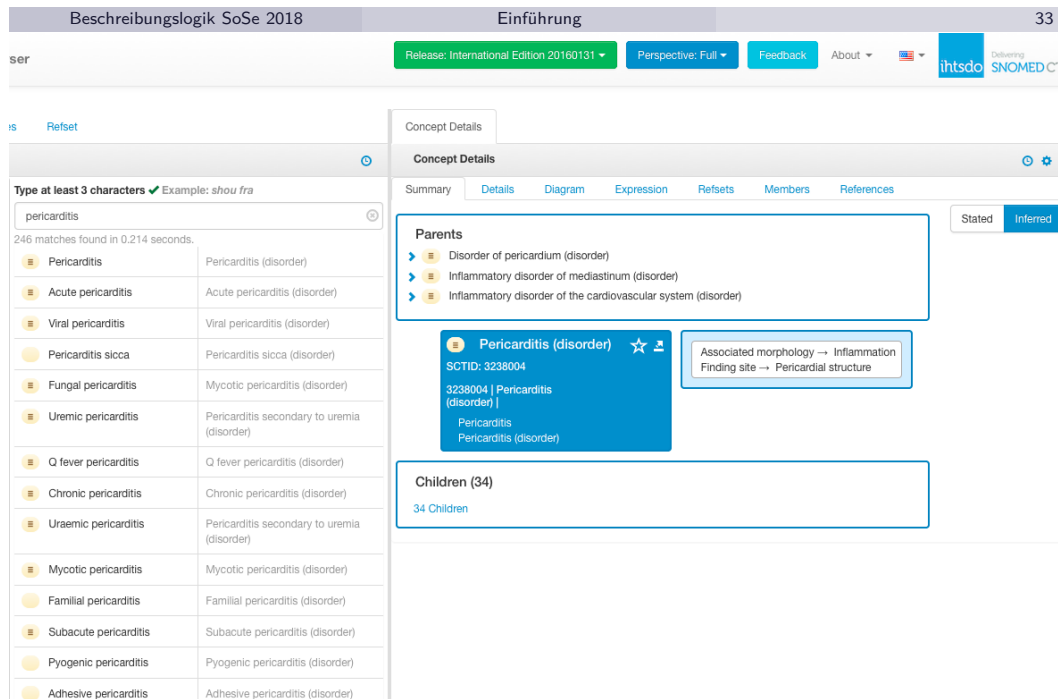
SNOMED CT ist einer von verschiedenen Standards

(ICD-11 ist ein anderer)

SNOMED CT ...

- wird von einem internationalen Nonprofit-Konsortium entwickelt (IHTSDO)
- standardisiert über 300.000 medizinische Begriffe aus allen Bereichen (Krankheiten, Behandlungen, Medikamente, etc.)
- wird regelmäßig aktualisiert, hat > 40 Autoren gleichzeitig
- basiert auf einer **Ontologie**, in der diese Begriffe formal mittels einer **Beschreibungslogik** definiert werden

Die Mediziner_innen erhalten natürlich keine logischen Ausdrücke, sondern einen Katalog von Begriffen.



The screenshot shows the IHTSDO SNOMED CT Browser interface. The search bar contains 'pericarditis' and shows 246 matches. The results list various types of pericarditis, such as 'Pericarditis (disorder)', 'Acute pericarditis (disorder)', and 'Viral pericarditis (disorder)'. The 'Pericarditis (disorder)' entry is highlighted, showing its SCTID (3238004) and associated morphology (Inflammation) and finding site (Pericardial structure). The interface also shows a 'Parents' section with 'Disorder of pericardium (disorder)', 'Inflammatory disorder of mediastinum (disorder)', and 'Inflammatory disorder of the cardiovascular system (disorder)'. A 'Children (34)' section is also visible.

Ausschnitt aus dem IHTSDO SNOMED CT Browser:
<http://browser.ihtsdotools.org>

Fragment aus SNOMED CT (vereinfacht)

Perikardium \sqsubseteq Gewebe \sqcap \exists teilVon.Herz
 Perikarditis \equiv Entzündung \sqcap \exists ort.Perikardium
 Entzündung \sqsubseteq Krankheit \sqcap \exists wirktAuf.Gewebe

Katalog



Die Web Ontology Language OWL



E. Shephard, aus Winnie-the-Pooh von A. A. Milne

OWL quick facts



OWL ...

- basiert auf einer ausdrucksstarken Beschreibungslogik (später mehr ...)
- mit präziser Semantik
- hat sich entwickelt aus jahrzehntelanger Forschung an
 - Techniken zum automatischen Schlussfolgern, auf denen Tools basieren, die Domänenexpert_innen bei Design und Pflege einer Ontologie unterstützen
 - Berechnungskomplexität (hängt eng mit Ausdrucksstärke zusammen!)
 - Modelltheorie
 - ...

Beispiel für OWL-Axiome

In Manchester-Syntax:

Inflammation	<i>SubClassOf</i>	Disease
HeartDisease	<i>EquivalentClass</i>	Disease and hasLoc some Heart
Endocarditis	<i>EquivalentClass</i>	Inflammation and hasLoc some Endocardium

aus dem **NCI Thesaurus**

- ca. 138.000 Begriffe (Classes)
- seit 2003 in OWL, monatliche Version, + 800 Begriffe/Monat
- veröffentlicht als Thesaurus (inferierte Hierarchie der Classes) sowie als OWL-Ontologie (mit logischen Axiomen), siehe BioPortal <http://bioportal.bioontology.org/>

OWL quick facts



Der Standard OWL ist mehr als nur eine Logik, stellt Infrastruktur zur Verfügung:

- Entitynamen sind IRIs
z. B. <http://purl.bioontology.org/ontology/SNOMEDCT/3238004> für Pericarditis
- mehrere Web- und nutzerfreundliche Syntaxen (RDF/XML, OWL/XML, ..., Manchester-Syntax)
- Importmechanismus
- Versionsmechanismus
- Annotationen von Entitys, Axiomen, ...
- ...

(Aussprache: [αʊl] – „aul“)

OWL versus DL

In Manchester-Syntax und DL-Syntax:

Inflammation	<i>SubClassOf</i>	Disease
Inflammation	\sqsubseteq	Disease
HeartDisease	<i>EquivalentClass</i>	Disease and hasLoc some Heart
HeartDisease	\equiv	Disease \sqcap \exists hasLoc.Heart
Endocarditis	<i>EquivalentClass</i>	Inflammation and hasLoc some Endocardium
Endocarditis	\equiv	Inflammation \sqcap \exists hasLoc.Endocardium

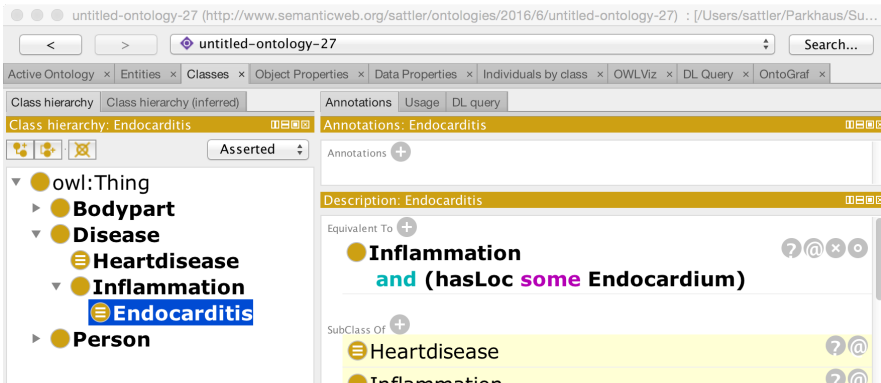
Der Ontologie-Editor Protégé



Unsere Beispiel-Axiome in Protégé:

```

Inflammation ⊆ Disease
HeartDisease ≡ Disease ⊓ ∃hasLoc.Heart
Endocarditis ≡ Inflammation ⊓ ∃hasLoc.Endocardium
  
```



Und nun ...

- 1 Organisatorisches
- 2 Einleitung
- 3 Vorlesungsüberblick

Der Ontologie-Editor Protégé



... ist ein OWL-Editor

- in der 5. Version
- basierend auf der OWL API
<http://owlcs.github.io/owlapi/>
- mit direktem Zugang zu OWL-Reasonern (= DL-Reasonern!)
- siehe <https://protege.stanford.edu/products.php>

Übersicht Vorlesung

- Kapitel 1: Einleitung ✓
- Kapitel 2: Grundlagen
- Kapitel 3: Ausdrucksstärke und Modellkonstruktionen
- Kapitel 4: Tableau-Algorithmen
- Kapitel 5: Komplexität
- Kapitel 6: Effiziente Beschreibungslogiken
- Kapitel 7: ABoxen und Anfragebeantwortung

Kapitelüberblick

Kapitel 2: Grundlagen

- Basis-Beschreibungslogik ALC
- TBoxen
- Schlussfolgerungsprobleme
- Erweiterungen von ALC

Kapitelüberblick

Kapitel 4: Tableau-Algorithmen

Wir zeigen, dass relevante Schlussfolgerungsprobleme **entscheidbar** sind und **vertretbare Komplexität** haben.

Außerdem: **praxistaugliche Algorithmen!**

- Tableau-Verfahren und dessen Korrektheit
- Komplexitätsanalyse
- Optimierungen
- In manchen Fällen: Sicherstellen von Terminierung

Praxisaufgabe geplant

Kapitelüberblick

Kapitel 3: Ausdrucksstärke und Modellkonstruktionen

- Nicht-Ausdrückbarkeit von Eigenschaften in ALC und Erweiterungen
- Baummodelleigenschaft
- Endliche Modelleigenschaft

Kapitelüberblick

Kapitel 5: Komplexität

Ziel: exakte Komplexitätsanalyse,
Kennenlernen von Standardtechniken

- Obere Schranken mittels Verfahren aus Modallogik
- Untere Schranken mittels Reduktion von spieltheoretischen Problemen
- Unentscheidbare Erweiterungen

Kapitelüberblick

Kapitel 6: Effiziente Beschreibungslogiken

Betrachten \mathcal{EL} :

geringere Ausdrucksstärke, aber effizientes Schlussfolgern

(SNOMED CT basiert im Wesentlichen auf \mathcal{EL} !)

- Grundbegriffe
- Polynomieller Algorithmus für Schlussfolgerungsprobleme
- Erweiterungen von \mathcal{EL} mit höherer Komplexität

Kleines Programmierprojekt geplant

Beim nächsten Mal ...

Kapitel 1: Einleitung

Kapitel 2: Grundlagen

Kapitel 3: Ausdrucksstärke und Modellkonstruktionen

Kapitel 4: Tableau-Algorithmen

Kapitel 5: Komplexität

Kapitel 6: Effiziente Beschreibungslogiken

Kapitel 7: ABoxen und Anfragebeantwortung

Kapitelüberblick

Kapitel 7: ABoxen und Anfragebeantwortung

Ziel: Schlussfolgerungsprobleme mit Instanzdaten (ABoxen) und Hintergrundwissen (TBoxen) untersuchen

- Grundbegriffe
- Typen von Anfragen
- Konjunktive Anfragen mit ABoxen+TBoxen
- Query Rewriting: Nutzung von Datenbanktechnologie

Plus Überblick aktuelle Forschungsthemen

Literatur für die Vorlesung (Basis)



Franz Baader, Ian Horrocks, Carsten Lutz, Uli Sattler.

An Introduction to Description Logic.

Cambridge University Press, 2017.

In SUUB verfügbar: <https://tinyurl.com/suub-intro-dl-ebook>

<https://tinyurl.com/suub-intro-dl>



Franz Baader, Diego Calvanese, Deborah L. McGuinness,

Daniele Nardi und Peter F. Patel-Schneider (Hrsg.).

The Description Logic Handbook.

2. Auflage, Cambridge University Press, 2007.

In SUUB zur Ausleihe und elektronisch verfügbar.



Franz Baader, Uli Sattler.

An Overview of Tableau Algorithms for Description Logics.

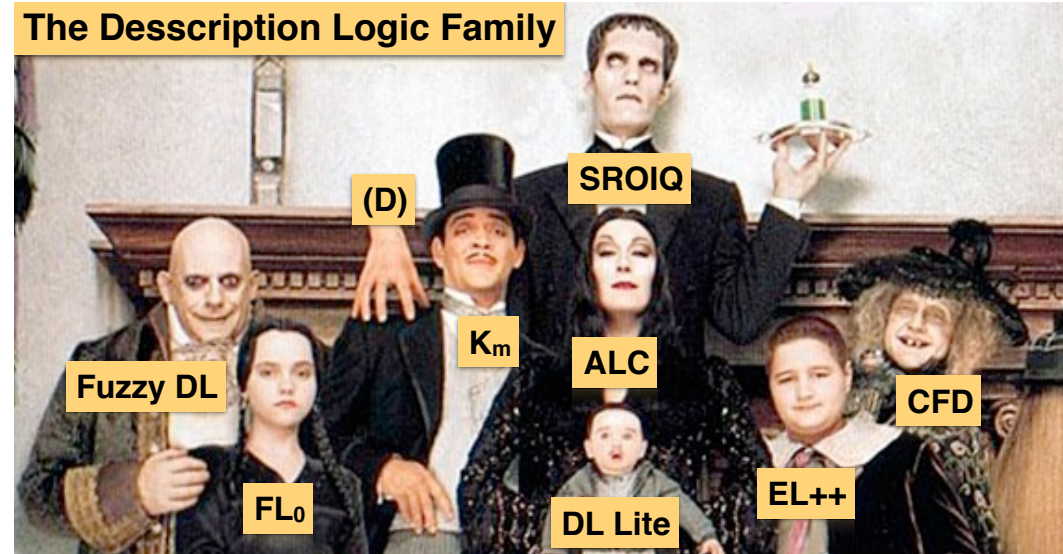
Studia Logica, 69:5–40, 2001. <http://dx.doi.org/10.1023/A:1013882326814>

Links für dieses Kapitel

- Stanford Center for Biomedical Informatics Research.
 Protégé Ontology Editor.
<http://protege.stanford.edu/>
 Der Ontologie-Editor (open source, gute Doku, viele Plugins)

- University of Manchester
 OWL API
<http://owlcs.github.io/owlapi/>
<https://github.com/owlcs/owlapi/wiki> (Wiki + Doku.)
 Die API der Wahl, um mit Ontologien zu arbeiten

The Description Logic Family



... sagt **danke** für Eure Aufmerksamkeit. Bis nächste Woche!