

---

# Beschreibungslogik und Ontologiesprachen

---

Vorlesung im Sommersemester 2010

# Organisatorisches

- Zeit und Ort:

Mo 14-16 MZH1460 und Mi 12-14 MZH1470

jedoch nicht am 10. und 12. Mai!

- Vortragender:

Prof. Dr. Carsten Lutz

Raum 3090

Tel. (218)-64431

clu@uni-bremen.de

- Position im Curriculum: Modulbereich Theorie, Vertiefung

# Organisatorisches

- Form: K4, jeder dritte Termin ist Übung  
Fragen und Diskussion jederzeit erwünscht.
- Voraussetzungen:  
VL „Logik“ hilfreich, aber nicht erforderlich.
- Vorlesungsmaterial:

Folien und Aufgabenblätter auf:

<http://www.informatik.uni-bremen.de/tdki/lehre/ss10/bl/>

Beispiele, Beweise, etc an der Tafel bitte mitschreiben

# Prüfungsmodalitäten

Mündliche Prüfung

oder

Übungen + Fachgespräch:

- Übungsaufgaben jede zweite Woche (Kernaufgaben + Zusatzaufgaben)
- Werden in Kleingruppen (2-3 Pers.) bearbeitet, abgegeben und korrigiert
- Lösungen für Übungsaufgaben werden von Studierenden präsentiert.
- Erste Übung (und Abgabe) am 19.4.

# Literatur

- The Description Logic Handbook, 2nd edition.  
Baader, Calvanese, McGuinness, Nardi, and Patel-Schneider (eds.),  
Cambridge University Press, 2007
- An Overview of Tableau Algorithms for Description Logics.  
Baader und Sattler. *Studia Logica*, 69:5-40, 2001
- Buchkapitel über Komplexitätsresultate in Stud.IP (hoffentlich)
- Proceedings der jährlichen Description Logic Workshops,  
online abrufbar unter <http://dl.kr.org/>

---

# Kapitel 1

## Einleitung

# Wissensrepräsentation

Die Wissensrepräsentation ist zentrales Teilgebiet der KI

Das Ziel ist...

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. [Brachman&Nardi03]

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

# 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.

Wohldefinierte Syntax und Semantik.

Syntax: die Sprache, in der Wissen „aufgeschrieben“ wird  
in dieser VL stets symbolisch und logik-basiert

Semantik: fixiert die Bedeutung des repräsentierten Wissens  
in exakter, eindeutiger Weise

# 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.

Deklarative Semantik: unabhängig von verarbeitender Software

- erlaubt (weitgehende) Anwendungsunabhängigkeit
- wir wollen repräsentieren (=beschreiben), nicht programmieren wie etwa in Prolog
- basiert auf (logischen) Strukturen

# Wissensrepräsentation

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.

Die Natur von Wissen ist schwierige (philosophische) Frage.

Wir beschränken uns auf konzeptuelles Wissen:

- Beschreibung des Konzeptes einer Vorlesung, eines Computers, einer Krankheit, eines Werkstückes, etc.
- Eine „was ist ein XYZ“ Beschreibung

Andere Arten von Wissen z.B. zeitliches Wissen, räumliches Wissen, prozedurales Wissen, Wissen über Wissen, etc.

# Wissensrepräsentation

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.

- Wissensrepräsentation bedeutet Abstraktion!
- Alle Aspekte eines Konzeptes im Detail zu identifizieren ist Aufgabe der Philosophie
- Nicht relevante Aspekte eines Konzeptes sollten nicht repräsentiert werden
- Damit ist eine bestimmte WR nicht unbedingt für alle Anwendungen adäquat (trotzdem deklarativ!)

# Wissensrepräsentation

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.

Intelligente Anwendungen können Informationen interpretieren und neue daraus ableiten.

- klassische Datenbank i.d.S. nicht intelligent:  
eingeegebene Daten = ausgegebene Daten, nur reorganisation
- Schlussfolgern (Reasoning, aus explizit gegebenem Wissen implizites ableiten) erforderlich; **T1.1**
- Anwendungsgebiete: Datenmanagement, intelligente Datenbanken, WorldWideWeb Suche, Benutzeroberflächen, etc.

# Wissensrepräsentation

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.

Schlussfolgerungsprobleme sollten entscheidbar sein und von möglichst geringer (Berechnungs-)Komplexität

Wechselwirkung:

- geringe Komplexität und Entscheidbarkeit vs.
- benötigte Ausdruckstärke

Ideale Balance abhängig von Anwendung und Abstraktionsgrad, also nicht ein WR Formalismus sondern viele!

# Wissensrepräsentation

- Logik erster Stufe (FO)
  - Formale Syntax, deklarative Semantik
  - Hohe, dennoch beschränkte Ausdrucksstärke
  - Sehr hohe Berechnungskomplexität  
(unentscheidbar, rekursiv aufzählbar, Theorembeweisen möglich)
- Aussagenlogik (AL)
  - Formale Syntax, deklarative Semantik
  - sehr beschränkte Ausdrucksstärke
  - relativ geringe Berechnungskomplexität
  - effektive Schlussfolgerungssysteme („SAT-Solver“) verfügbar

BL ist Kompromiss bzgl. Ausdrucksstärke und Komplexität

# Beschreibungslogik

Beschreibungslogik ist ein wichtiges Teilgebiet der WR, neben z.B. nicht-monotonen Logiken, Aktionslogiken, etc.

BLen sind eine Familie von Logiken

- zur Repräsentation konzeptuellen Wissens
- mit entscheidbaren Schlussfolgerungsproblemen
- für die viele effiziente Reasoner zur Verfügung stehen
- die eine wichtige Rolle als Ontologiesprachen spielen

# Beschreibungslogik

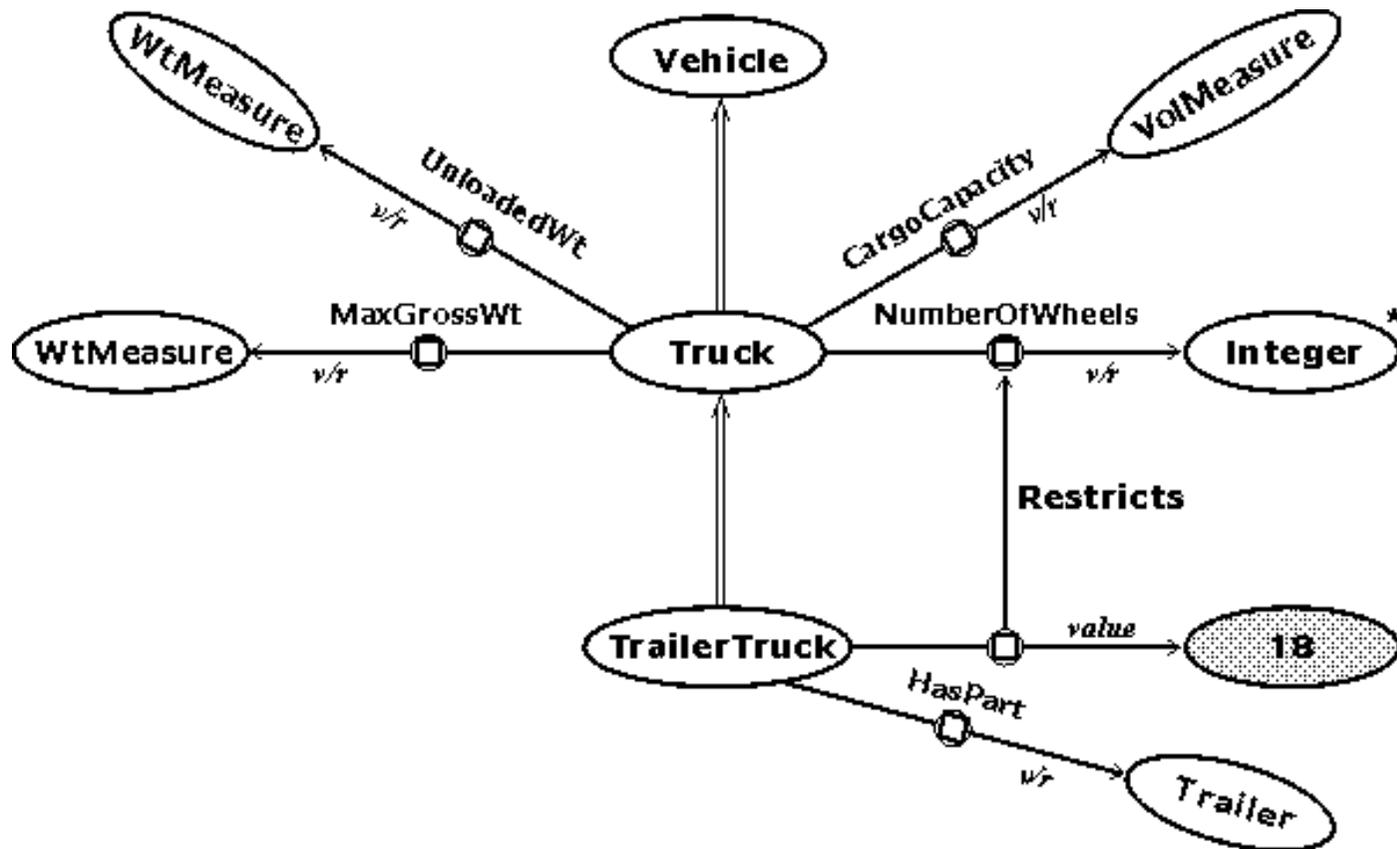
Historisch hervorgegangen sind BLen aus

- structured inheritance networks [Brachman78]
- einem bekannten historischen WR-System namens KL-ONE [Brachman&Schmolze85]

Diese frühen Formalismen

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

# KL-ONE Beispiel



# Beschreibungslogik

Zentrale Elemente der WR mit Beschreibungslogik:

- Konzepte

Konzept beschreibt eine Klasse von Objekten

Mensch, Kurs, Universität, Tafel, Student, etc.

Können durch logische Ausdrücke (Formeln) beschrieben werden:

Mensch  $\sqcap \exists$  hört.Vorlesung

- Rollen

Sind binäre Relationen zwischen Objekten

hört, lehrt, istTeilVon, etc.

können (meist) nicht durch komplexe Ausdrücke beschrieben werden

# Beschreibungslogik

Zentrale Elemente der WR mit Beschreibungslogik:

- 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 WR mit Beschreibungslogik:

- ABoxen (assertionale Boxen)

Beschreiben Individuen (= Objekte) und deren Eigenschaften

z.B.

Student(hans)

Vorlesung  $\sqcap$   $\exists$  hatThema.InformatikThema(blv)

hört(hans, blv)

# Ontologien

Ontologie ist ursprünglich eine Teildisziplin der Philosophie  
(die Lehre von dem, was ist)

Seit einigen Jahren beliebtes Paradigma in der Informatik:

Eine Ontologie ist eine explizite Spezifikation einer  
Konzeptualisierung eines Anwendungsbereiches.

[Gruber92]

Konzeptualisierung  $\approx$  Identifikation von Konzepten, Relationen, etc.  
i.e., konzeptuelles Wissen

Expl. Spezifikation  $\approx$  Beschreibung in formaler Sprache

# Ontologien - 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

Es gibt verschiedene Standards:

- ICD-10: Int. Classification of Diseases and Related Health Problems  
Publiziert von WHO, verwendet z.B. in Deutschland
- SNOMED CT: Standardized Nomenclature of Medicine, Clinical Terms  
Verwendet z.B. in USA, GB, Schweden, Australien,...

# Ontologien - SNOMED

## SNOMED

- von einem internationalen non-profit Konsortium entwickelt (IHTSDO)
- standardisiert etwa 500.000 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 erhalten natürlich keine logischen Ausdrücke sondern einen Katalog von Begriffen.

# Ontologien - SNOMED

## Ontologiefragment

Perikardium  $\sqsubseteq$  Gewebe  $\sqcap$   $\exists$ teilVon.Herz

Perikarditis  $\equiv$  Entzündung  $\sqcap$   $\exists$ ort.Perikardium

Entzündung  $\sqsubseteq$  Krankheit  $\sqcap$   $\exists$ wirktAuf.Gewebe

## Katalog



# Ontologien – NCI Thesaurus

Krebsforschung ist sehr interdisziplinär:

Medizin (verschiedene Fachrichtungen),  
Biologie, Chemie, Genetik, Pharmazeutik, etc.

Dies führt zu

- einer unüberschaubaren Anzahl von Fachbegriffen und Synonymen
- Kommunikationsschwierigkeiten zwischen Menschen und beim Datenaustausch

Das amerikanische National Cancer Institute (NCI) begegnet diesem Problem mit dem NCI Thesaurus.

# Ontologien – NCI Thesaurus

## NCI Thesaurus

- erfasst etwa 50.000 Begriffe, darunter 10.000 Krebserkrankungen und 8.000 Medikamente und Therapien
- wird monatlich aktualisiert
- basiert auf einer Ontologie, in der diese Begriffe formal mittels einer Beschreibungslogik definiert werden

## Krebsforscher können

- textuelle Definitionen der standardisierten Begriffe einsehen
- die standardisierten Begriffe zum Datenaustausch nutzen.

The screenshot displays the NCI Terminology Browser in a Mozilla Firefox browser window. The left pane shows a tree view of medical terms, with 'Cystic Fibrosis' selected. The right pane shows the 'Concept Details' for 'Cystic Fibrosis'.

**NCI Terminology Browser - Mozilla Firefox**  
 http://nciterns.nci.nih.gov/NCIBrowser/GetTree.do?bookmarktag=1&tag=exp

- Genetic Disorder
  - Chromosome Disorder
  - Coagulation Factor Deficiency
  - Cockayne Syndrome
  - Congenital Combined Immunodeficiency
  - Congenital Disorder of Natural Immunity
  - Congenital Leucocyte Abnormality
  - Congenital T-Cell Immunodeficiency
  - Cystic Fibrosis
  - Cystinosis
  - DNA Repair Disorder
  - Down Syndrome
  - Fanconi Syndrome
  - Fibrodysplasia Ossificans Progressiva
  - Granular Corneal Dystrophy
  - Hereditary Elliptocytosis
  - Hereditary Hemorrhagic Telangiectasia
  - Hereditary Neoplastic Syndrome
  - Hyperimmunoglobulin Syndrome
  - Laurence-Moon Syndrome
  - Maple Syrup Urine Disease
  - Dihydrouracil Dehydrogenase Deficiency
  - Trichothiodystrophy
- Hamartoma
- Hyperplasia
- Neoplasm
- Non-Neoplastic Disorder
- Polyp
- Psychiatric Disorder
- Radiation-Related Disorder
- Rare Disorder
- Syndrome
- Finding
- Drug, Food, Chemical or Biomedical Material

**Mozilla Firefox**  
 http://nciterns.nci.nih.gov/NCIBrowser/GetRoleAndProperty.do?bookmarktag=1&conceptname=Cystic\_Fibr

U.S. National Institutes of Health | www.cancer.gov

HELP RESULTS CUSTOMIZE ABOUT BROWSE HIERARCHY LOGOUT

Search  
Go!

Concept Details  
Bookmark this page

Cystic Fibrosis [Printable Page](#) [History](#) [Graph](#)

**Identifiers:**

name	Cystic_Fibrosis
code	C2975

**Relationships to other concepts:**

Disease_Has_Associated_Anatomic_Site	Pancreas
Disease_Has_Primary_Anatomic_Site	Gastrointestinal System
Disease_Has_Associated_Anatomic_Site	Gastrointestinal System

**Information about this concept:**

ALT\_DEFINITION

DEFINITION

NCI-GLOSS|A common hereditary disease in which exocrine (secretory) glands produce abnormally thick mucus. This mucus can cause problems in digestion, breathing, and body cooling.

NCI|A congenital metabolic disorder affecting the exocrine glands, inherited as an autosomal trait. The secretions of exocrine glands are abnormal, resulting in excessively viscous mucus production which causes obstruction of passageways (including pancreatic and bile ducts, intestines, and bronchi). The sweat sodium and chloride content are increased. Symptoms usually appear in childhood and include meconium ileus, poor growth despite good appetite, malabsorption and foul bulky stools, chronic bronchitis with cough, recurrent pneumonia, bronchiectasis, emphysema, clubbing of the fingers, and salt depletion in hot weather. -- 2003

Start | slides | 5 Microsoft Office P... | NCI Terminology Bro... | NCI Terminology Bro... | IrfanView | 14:58

# Ontologien und Beschreibungslogik

Beschreibungslogik wird häufig als Ontologiesprache eingesetzt, ebenso z.B. RDF, F-Logik und FOL.

OWL: the Web Ontology Language:

- Standardisiert vom W3C (World Wide Web Consortium)
- OWL 1.0: 2004    OWL 2.0: Oktober 2009
- OWL 1.0 umfasst 3 Sprachen, OWL 2.0 umfasst 5 Sprachen
- Zugeschnitten aufs Web: XML Syntax etc.
- Viele Tools verfügbar (Editoren, Browser, Reasoner, etc.)

# Struktur Vorlesung

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