

Vorlesung Semantic Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering

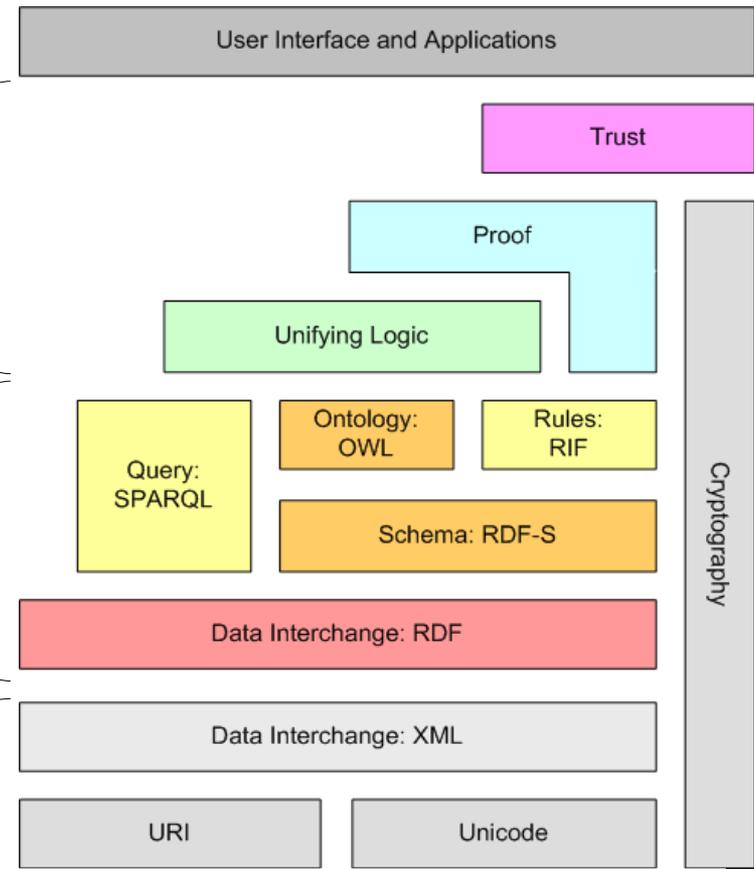
Semantic Web – Aufbau



here be dragons...

Semantic-Web-
Technologie
(Fokus der Vorlesung)

Technische
Grundlagen



Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>

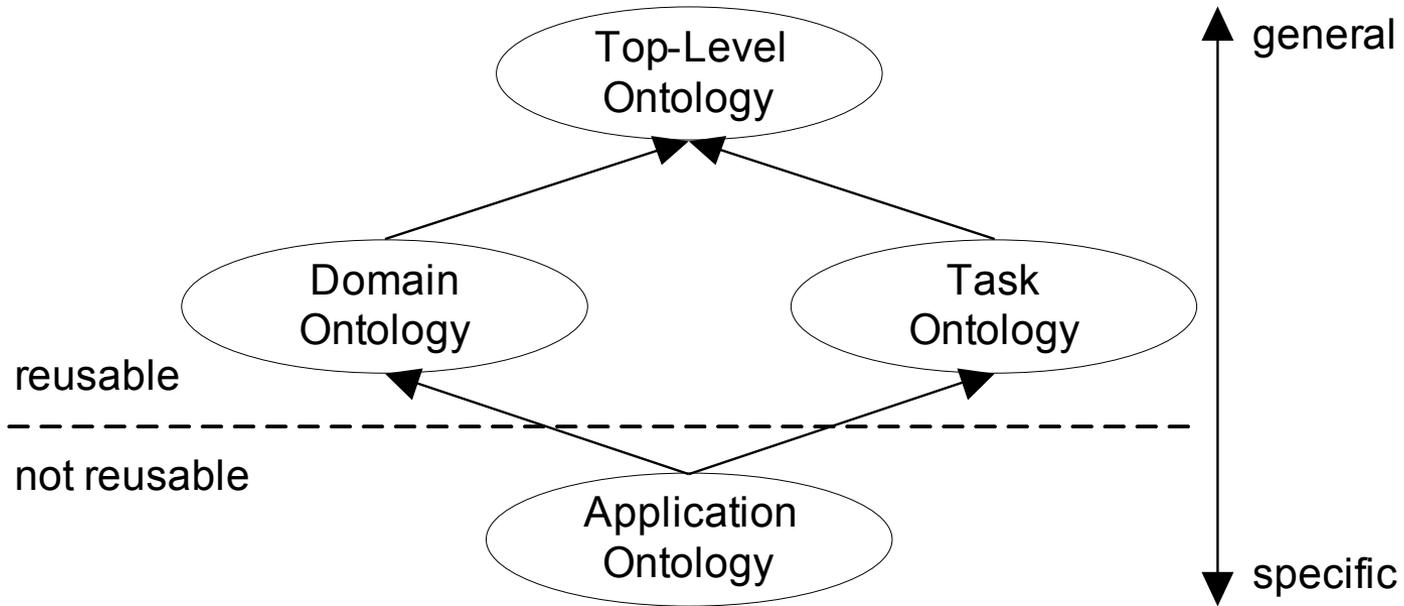
Rückblick: Ontologien



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Ontologien liefern die eigentliche Semantik
 - RDF Schema (leichtgewichtig)
 - OWL (schwergewichtiger)
- ermöglichen Reasoning
 - dazu muss ein relevanter Ausschnitt der Welt modelliert sein
 - und zwar möglichst exakt

Klassifikation von Ontologien

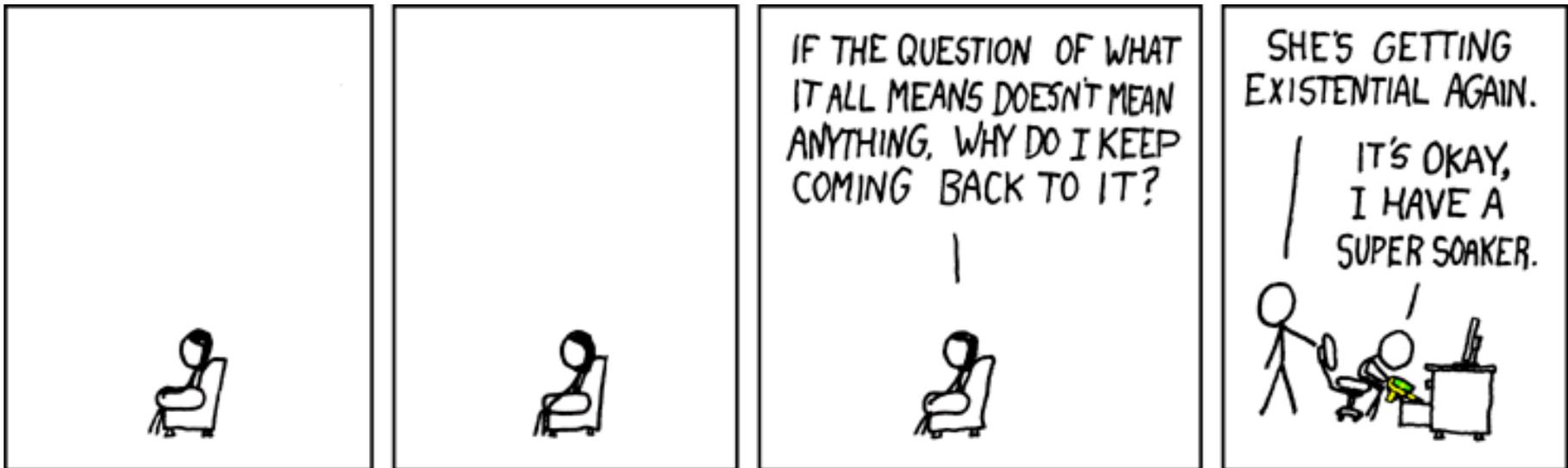


Top-Level-Ontologien

- Top-Level-Ontologien
 - sind domänenunabhängig
 - sind aufgabenunabhängig
 - ...und damit sehr allgemein
- Ziel
 - Wiederverwendung
 - Semantische Klarheit
 - Interoperabilität
 - z.B. vereinfachtes Matching

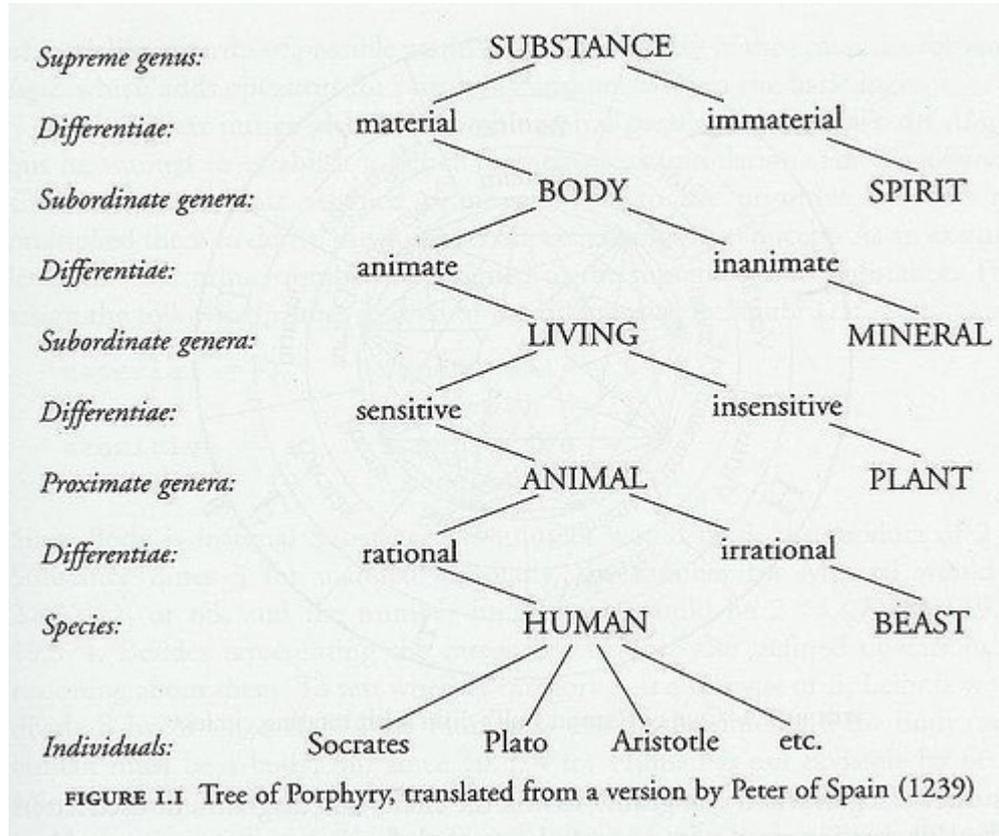
Warnung

- Top-Level-Ontologien haben viel mit Philosophie zu tun



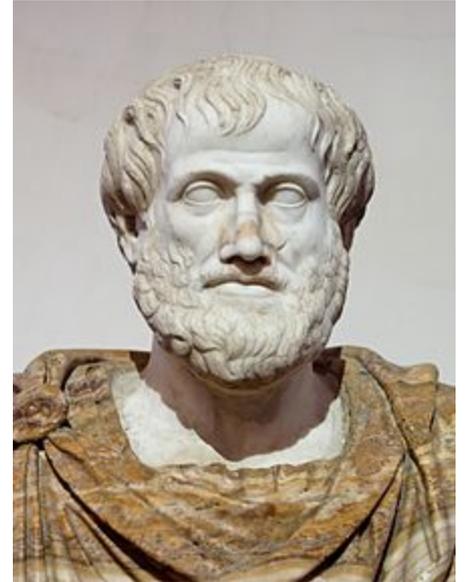
<http://xkcd.com/220/>

Beispiel



Porphyrios, griechischer Philosoph, ca. 233-301

- Eine der ältesten Top-Level-Ontologien
 - Aristoteles (384-322)
- Vier Grundkategorien von Existierendem
 - Zurückgeführt auf Relation zu einem *Zugrundeliegendem* (Subjekt)
 - einfache Vorstellung von *Zugrundeliegendem*: eine Instanz
 - z.B. Sokrates



Das ontologische Quadrat von Aristoteles

- Mögliche Relationen einer Substanz zu einem Zugrundliegendem
- Inhärenz (trennt Substanzielles von Nicht-Substanziellem)
 - Etwas ist Teil von einem Zugrundeliegenden oder
 - Nicht Teil davon, kann aber nicht unabhängig existieren
- Prädikation (trennt Allgemeines von Individuellem)
 - Das Zugrundeliegende hat/trägt eine Eigenschaft

Das ontologische Quadrat von Aristoteles



	ist inhärent	ist nicht inhärent
ist prädiziert	<i>Kategorie II</i> z.B. die Farbe Weiß	<i>Kategorie III</i> z.B.: die Kategorie der Kaffeetassen
ist nicht prädiziert	<i>Kategorie I</i> z.B.: die weiße Farbe einer bestimmten Kaffeetasse	<i>Kategorie IV</i> z.B.: eine bestimmte Kaffeetasse

Grundlegende Unterscheidungen für Top-Level-Ontologien



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Abstrakte vs. konkrete Entitäten
- Abstrakte Entitäten haben keine zeitliche oder räumliche Ausdehnung
 - Zahlen
 - Informationen
- Konkrete Entitäten haben zumindest eine zeitliche Ausdehnung
 - Dinge (Bücher, Tische, ...)
 - Ereignisse (Vorlesungen, Weltmeisterschaften, ...)

Grundlegende Unterscheidungen für Top-Level-Ontologien



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- 3D- vs. 4D-Sicht
- 3D-Sicht
 - Dinge haben Ausdehnung im Raum
 - sind zu jedem Zeitpunkt ihrer Existenz vollständig vorhanden
- 4D-Sicht
 - Dinge haben Ausdehnung in Zeit und Raum
 - sind zu jedem Zeitpunkt ihrer Existenz nur partiell vorhanden
- Tatsächliche vs. mögliche Dinge
 - Ansatz "Tatsächliche Dinge" (actualism): nur existierende Dinge werden in Ontologie gefasst
 - Ansatz "Mögliche Dinge" (possibilism): auch mögliche Dinge werden in Ontologie gefasst



Grundlegende Unterscheidungen für Top-Level-Ontologien



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Kolokation
- Können verschiedene Dinge am selben Platz sein?
- Die Frage sollte eigentlich leicht zu beantworten sein:
 - 3D-Sicht: nein
 - 4D-Sicht: ja, aber nicht zur selben Zeit
- ...ist aber nicht so trivial
 - Beispiel: eine Statue und das Metall, aus dem die Statue ist
 - gibt es überhaupt Statuen?
 - oder nur Material, das wie Statuen geformt ist?
 - ...und sind das dann identische oder unterschiedliche Kategorien?
 - Anderes Beispiel: ein Loch in einem Stück Emmentaler
 - gibt es überhaupt Löcher?
 - oder nur löchrige Objekte?



Die Top-Level-Ontologie von John Sowa



- Eine ältere Top-Level-Ontologie (90er Jahre)
- Drei Unterscheidungen führen zu 12 Kategorien
 - Physical vs. Abstract
 - Dinge die in Raum und Zeit existieren
 - Dinge, auf die das nicht zutrifft
 - Continuant vs. Occurent
 - Dinge, die zu jedem Zeitpunkt ganz existieren
 - Dinge, die zu einem Zeitpunkt nur partiell existieren
 - Independent vs. Relative vs. Mediating
 - Dinge, die allein existieren können
 - Dinge, die andere Dinge für ihre Existenz benötigen
 - "Dritte" Dinge, die zwei andere in Relation setzen

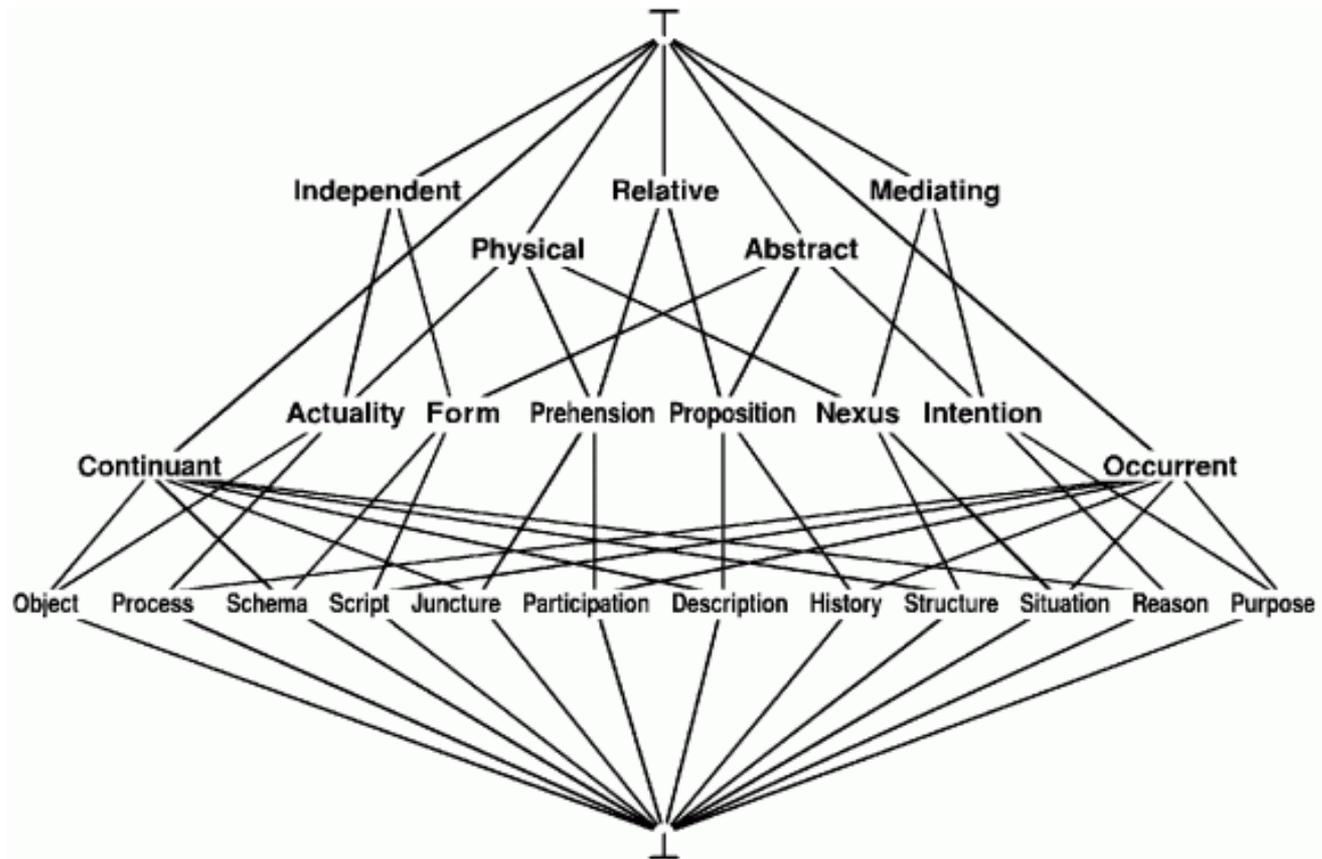
Die Top-Level-Ontologie von John Sowa

- Anhand dieser drei Unterscheidungen entstehen zwölf disjunkte Klassen von Objekten

	Physical		Abstract	
	Continuant	Occurent	Continuant	Occurent
Independent	Object	Process	Schema	Script
Relative	Juncture	Participation	Description	History
Mediating	Structure	Situation	Reason	Purpose

John F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations (1999)

Die Top-Level-Ontologie von John Sowa



John F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations (1999)

- *Descriptive Ontology for Linguistic and Cognitive Engineering*
- Eine der bekanntesten Top-Level-Ontologien
 - entwickelt im WonderWeb-Projekt (EU 2002-2004)
 - starker philosophischer Überbau
- Modularer Aufbau
 - Basisontologie: 37 Klassen, 70 Relationen
 - Alle Module: ~120 Klassen, ~300 Relationen

Grundlegende Unterscheidungen in DOLCE



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Kategorien, Instanzen, Eigenschaften
- Kategorien (particulars): können Instanzen bilden
 - "Stadt", "Universität"
- Instanzen (universals): können keine Instanzen bilden
 - "Darmstadt", "Technische Universität Darmstadt"
- Eigenschaften: beschreiben eine Instanz
 - z.B. die Farbe eines Buches, die Größe eines Menschen
 - sind weder Kategorie noch Instanz
 - können ohne eine Instanz nicht existieren!

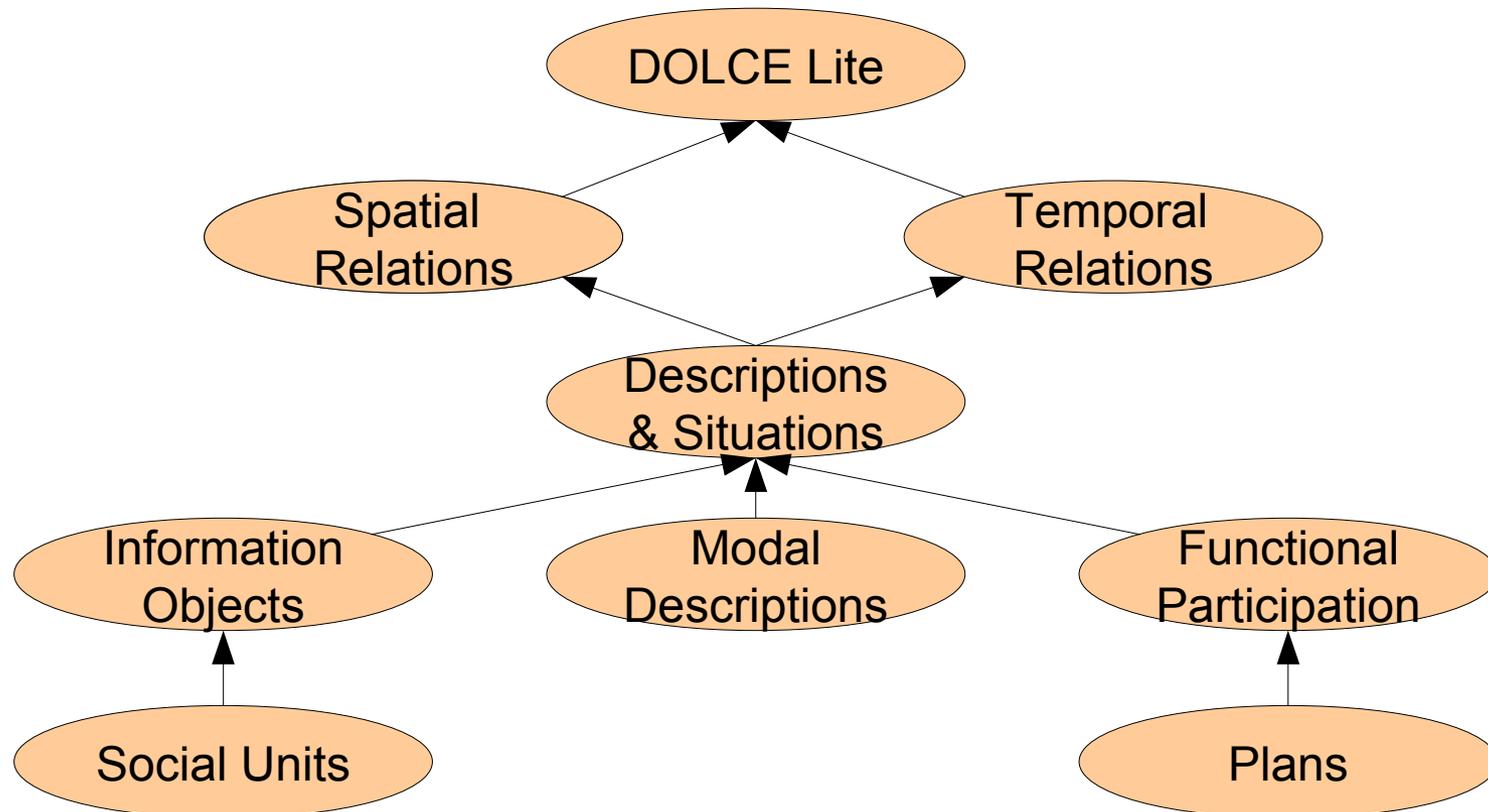
Grundlegende Annahmen von DOLCE



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Eine Top-Level-Ontologie von Kategorien
 - d.h., *particulars*: Stadt, Land, ...
 - für tatsächliche und mögliche Entitäten
- 4D
 - d.h. manche Entitäten haben auch eine zeitliche Ausdehnung
- Kolokation
 - ist erlaubt
 - aber: nicht zwei gleichartige Dinge am selben Ort
 - nicht zwei Statuen
 - aber eine Statue und eine Menge Bronze

DOLCE als Ontologie-Bibliothek



Der "DOLCE-Baukasten"



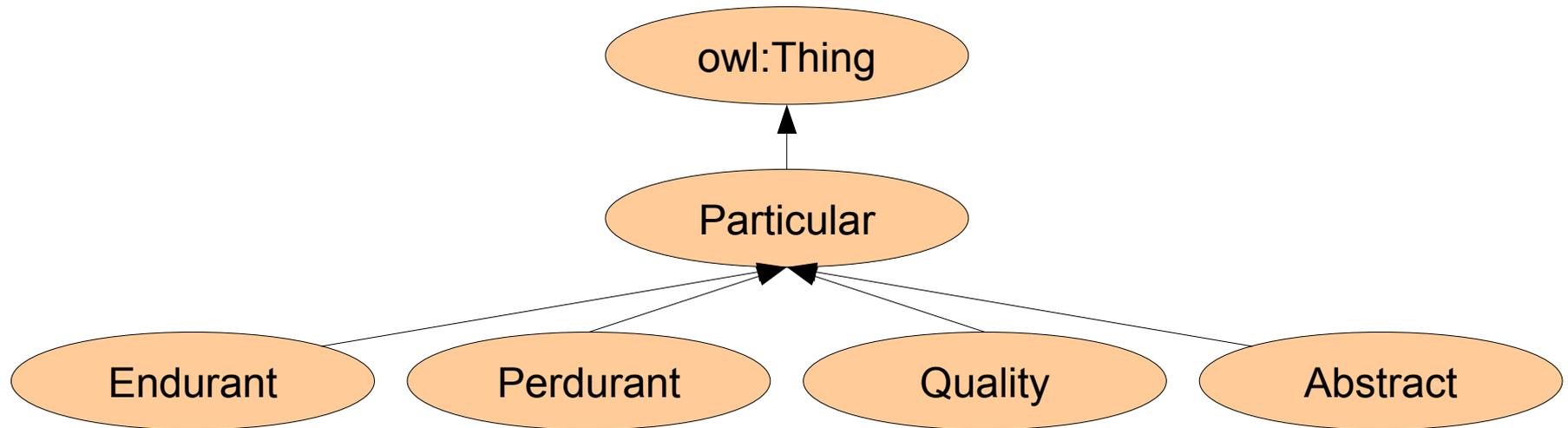
TECHNISCHE
UNIVERSITÄT
DARMSTADT

- DOLCE-Lite: Basisklassen
- Spatial/Temporal Relations
- Descriptions&Situations
 - Kommunikation zwischen Menschen und Systemen
- Information Objects
- Social Units: Menschen, Organisationen...
- Modal Descriptions
 - Mentale Objekte, Versprechungen, Commitments
- Functional Participation
 - Material, Werkzeug, Ergebnis, ...
- Plans



Die vier Basis-Kategorien in DOLCE

- alle vier sind disjunkt



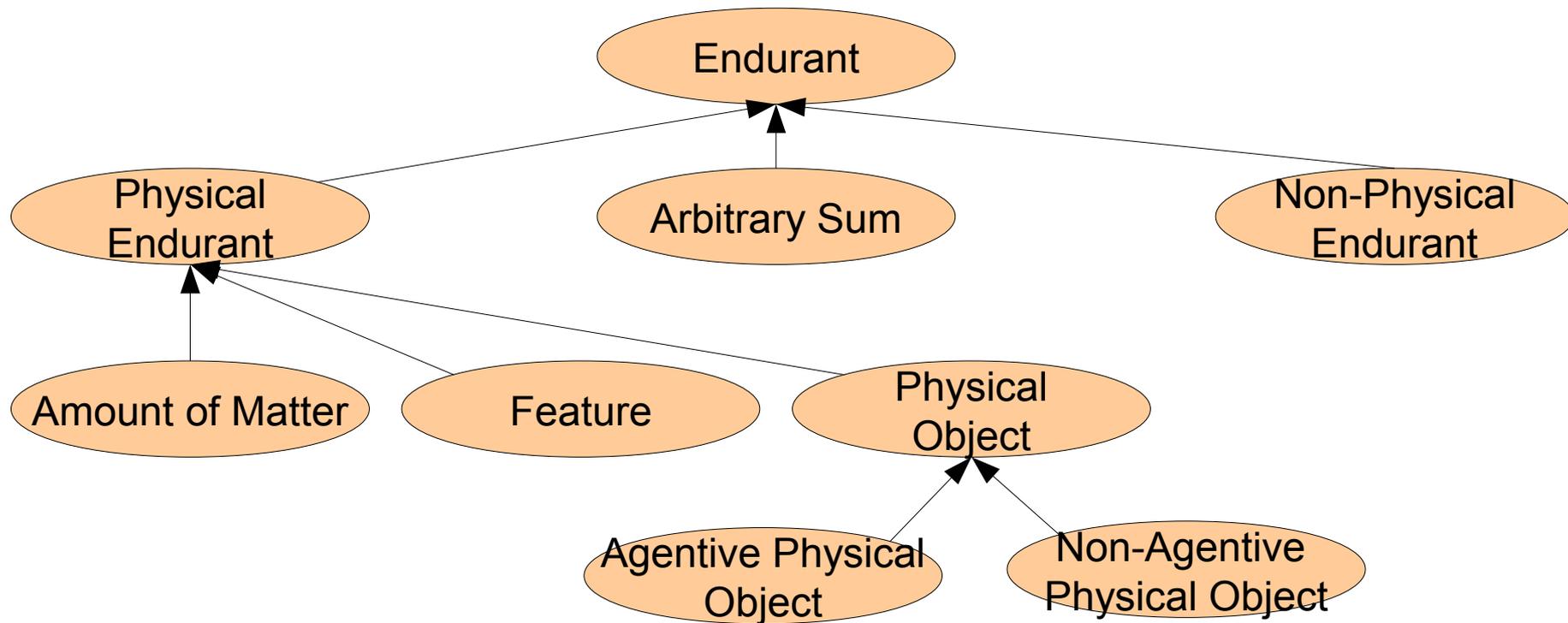
Masolo et al. (2003): *Ontology Library (final)*. WonderWeb Deliverable D18.

Endurants vs. Perdurants



- Endurants existieren in der Zeit
 - Vorstellung: Dinge wie Menschen, Bücher, ...
 - können auch nicht-physikalisch sein: Organisationen, Informationen
 - Sind zu jedem Zeitpunkt ihrer Existenz komplett vorhanden
- Perdurants "passieren" in der Zeit
 - Vorstellung: Ereignisse und Prozesse
 - Sind zu jedem Zeitpunkt ihrer Existenz nur partiell vorhanden
 - d.h., vergangene und zukünftige Phasen existieren zu einem bestimmten Zeitpunkt nicht

Endurants in DOLCE (1)

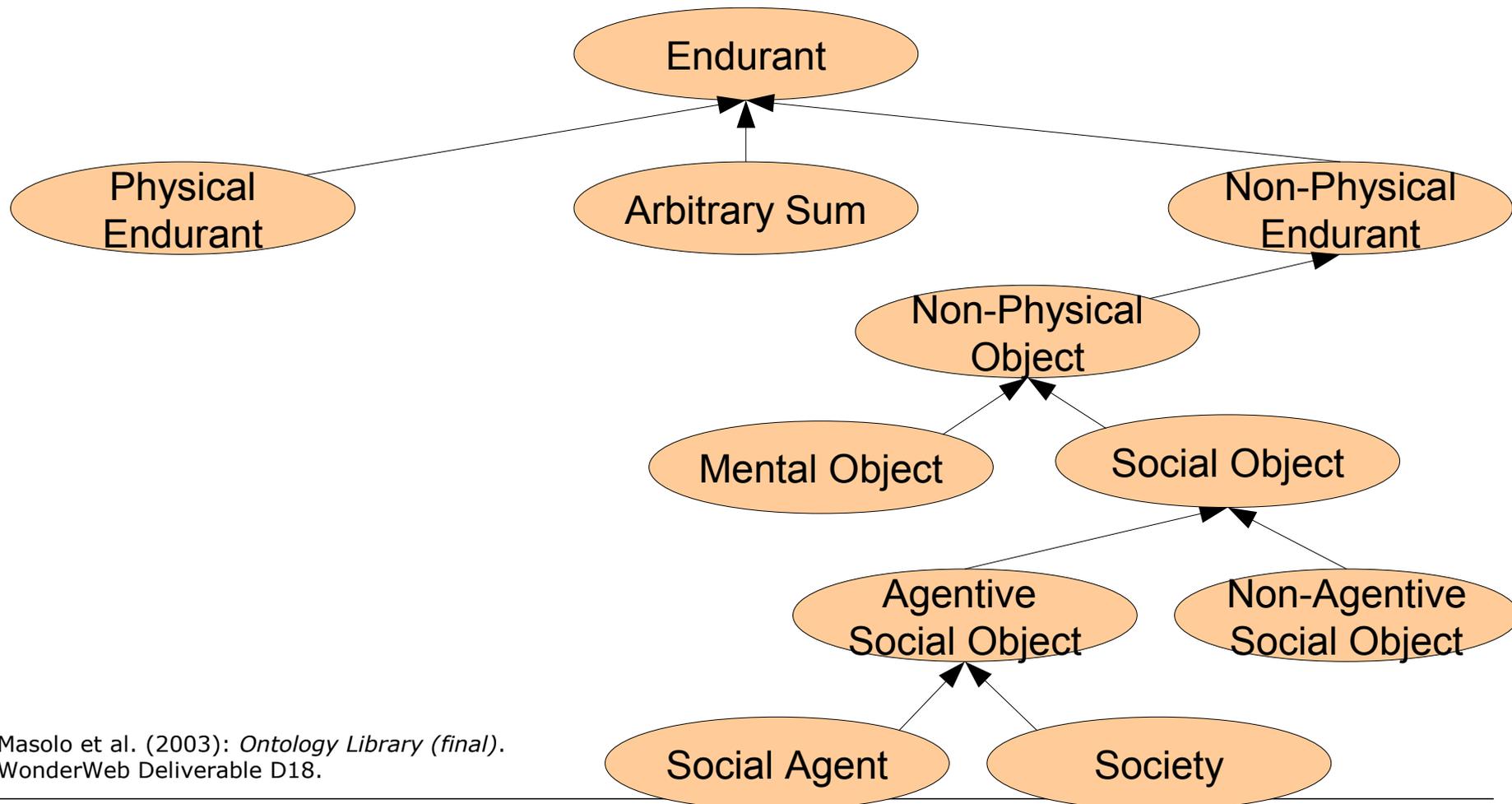


Masolo et al. (2003): *Ontology Library (final)*. WonderWeb Deliverable D18.

Unterscheidung von Endurants

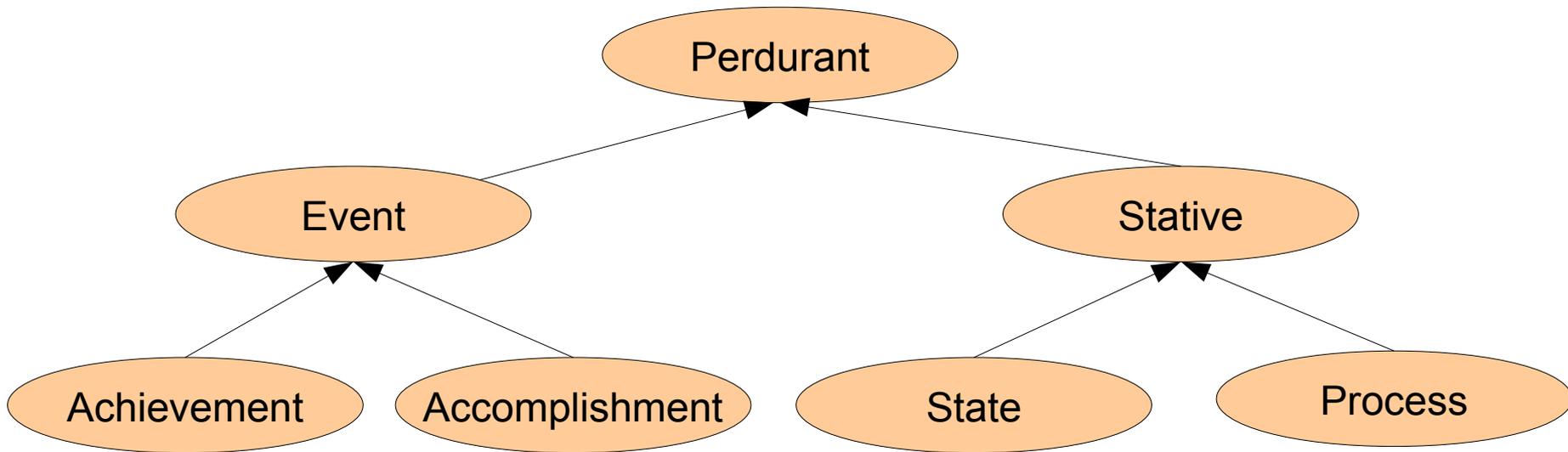
- Amount of Matter vs. Physical Object
 - Amount of Matter ist mereologisch invariant
 - d.h.: ein Teil von einem AoM ist vom selben Typ
 - ein Teil von "eine Menge Erde" ist immer noch "eine Menge Erde"
 - aber ein Teil einer Tasse ist nicht mehr unbedingt eine Tasse
- Features
 - können nicht ohne einen anderen Physical Endurant existieren
 - z.B. Rand, Loch, ...

Endurants in DOLCE (2)



Masolo et al. (2003): *Ontology Library (final)*.
WonderWeb Deliverable D18.

Perdurants in DOLCE



Masolo et al. (2003): *Ontology Library (final)*. WonderWeb Deliverable D18.

Unterscheidung von Perdurants

- Events vs. Statives
 - Die Summe von zwei aufeinander folgenden Statives eines Typs ist ein längeres Stative vom selben Typ
 - die Summe von zwei mal hintereinander "Herumsitzen" ist ein längeres "Herumsitzen"
 - aber die Summe von zwei mal hintereinander "zum Mond fliegen" ist nicht einfach ein längeres "zum Mond fliegen"

Unterscheidung von Perdurants

- Achievement vs. Accomplishment
 - Achievements sind unteilbar ("Erreichen der Stadtgrenze")
 - Accomplishments sind teilbar
- State vs. Process
 - States bestehen nur aus States vom gleichen Typ ("Herumsitzen")
 - Processes können auch aus anderen Processes bestehen
 - z.B. "Studieren", besteht aus "Vorlesung hören", "Übung machen", "Referat halten", ...

Verhältnis von Endurants und Perdurants



- Endurants nehmen an Perdurants Teil
 - entweder aktiv (Lesender und Lesevorgang)
 - oder passiv (Buch und Lesevorgang)
 - Unterscheidung in DOLCE-Lite:
 - Konstante Partizipation (über den gesamten Perdurant hinweg)
 - Temporäre Partizipation
 - Weitere, z.B. in Functional Participation
 - patient (verfolgt keinen eigenen Plan), z.B. Ziel, Thema
 - product (Ergebnis eines Perdurants)
 - use-of (Verwendung), als Ressource oder Werkzeug
 - ...

Verhältnis von Endurants und Perdurants



- Endurants haben nur Endurants als Bestandteile, Perdurants nur Perdurants
 - Bücher bestehen aus Seiten, Einband, ...
 - Ein Lesevorgang besteht aus Wahrnehmungs-, Umblättern-, ...vorgang
- Endurants und Perdurants verändern sich auf verschiedene Weise
 - z.B.: Endurants können zu verschiedenen Zeiten verschiedene Bestandteile haben
 - Seite aus einem Buch herausreißen
 - für Perdurants gilt das so nicht

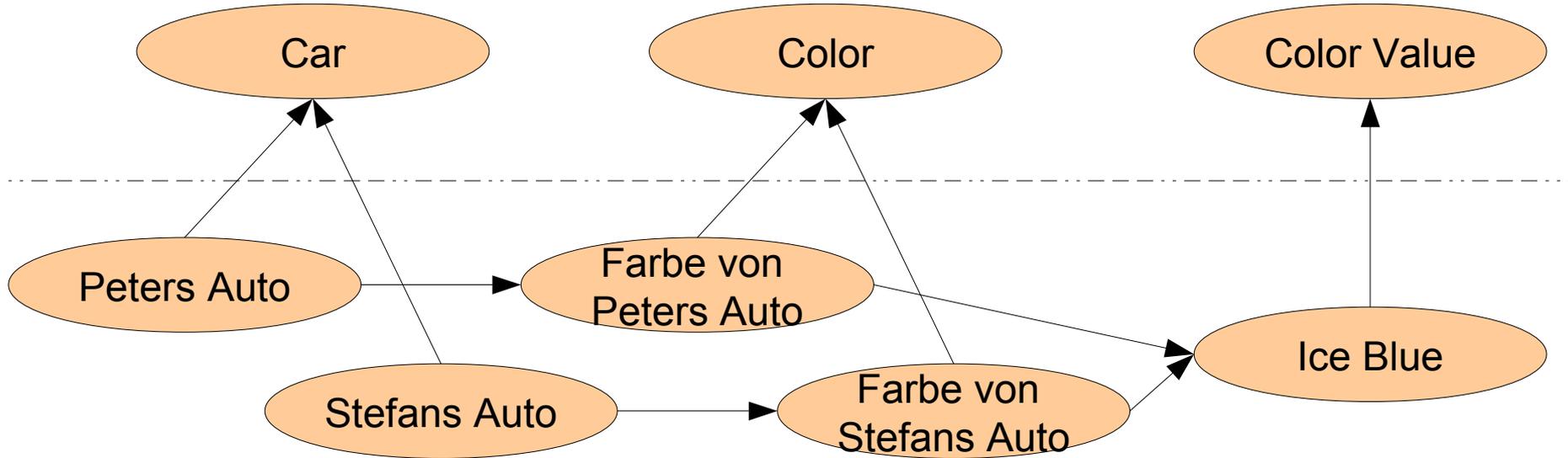
- Grundsätzliche Unterscheidung:
 - Qualität ist eine Eigenschaft einer Entität
 - Der Qualitätsraum (quality space) ist die Menge der möglichen Ausprägungen
 - Wert (quale) einer Qualität ist Position im Qualitätsraum
- Qualitäten benötigen Entitäten
 - im Prinzip alle Particulars möglich
 - "leben" nur so lang wie die zugehörige Entität

Qualitäten

- Beispiel:
 - Farbe ist eine Qualität
 - RGB ist ein Qualitätsraum
- "Zwei Autos haben die exakt gleiche Farbe"
 - Jedes Auto besitzt seine eigene Qualität "Farbe"
 - beide Qualitäten haben denselben Wert im Raum
- Warum hat jedes Auto seine eigene Qualität?

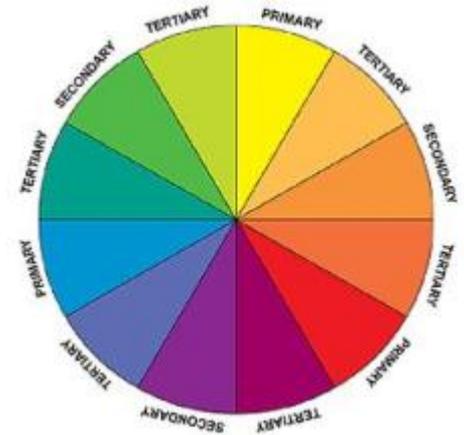


Qualitäten



Qualitäten

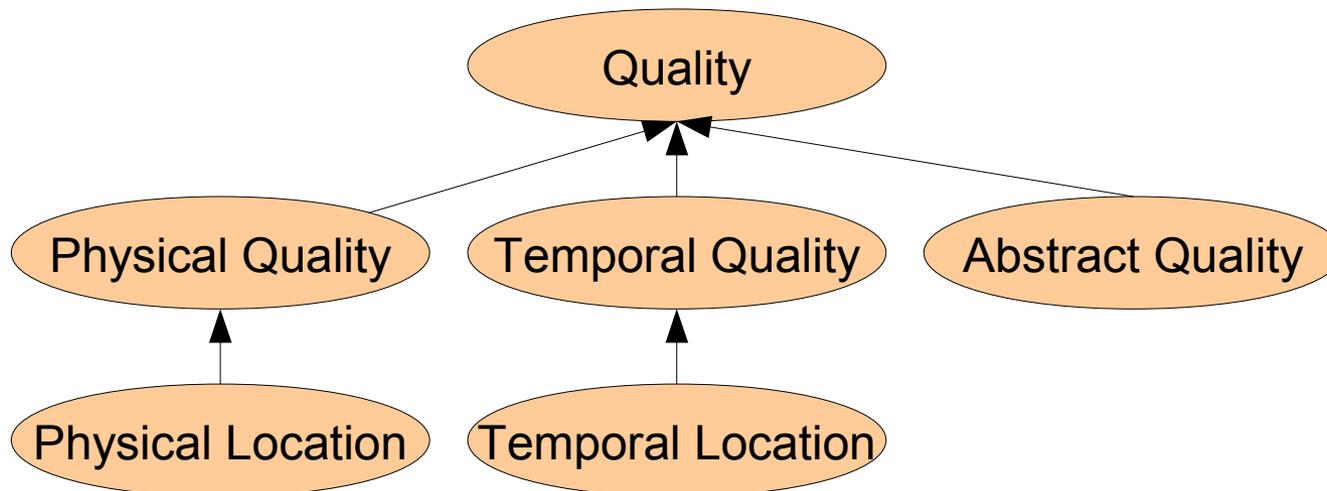
- Beispiel:
 - Farbe ist eine Qualität
 - RGB ist ein Qualitätsraum
- "Zwei Autos haben die exakt gleiche Farbe"
 - Jedes Auto besitzt seine eigene Qualität "Farbe"
 - beide Qualitäten haben denselben Wert im Raum
- Warum hat jedes Auto seine eigene Qualität?
 - Qualitäten "leben" nur so lange wie die Entität, an die sie gebunden sind
 - Ansonsten hätte das zweite Auto keine Farbe mehr, wenn das erste aufhört zu existieren



- Direkte und indirekte Qualitäten
- Direkte Qualitäten lassen sich einem Entitäten direkt zuschreiben
 - z.B.: Position eines Endurants
- Indirekte Qualitäten sind Qualitäten von Entitäten, die einer anderen Entität zugeschrieben sind
 - z.B.: Position eines Perdurants ist gegeben durch Positionen der involvierten Endurants
 - z.B.: Geschwindigkeit eines Endurants ist gegeben durch Geschwindigkeitsqualität des Perdurants ("sich bewegen")

Qualitäten in DOLCE

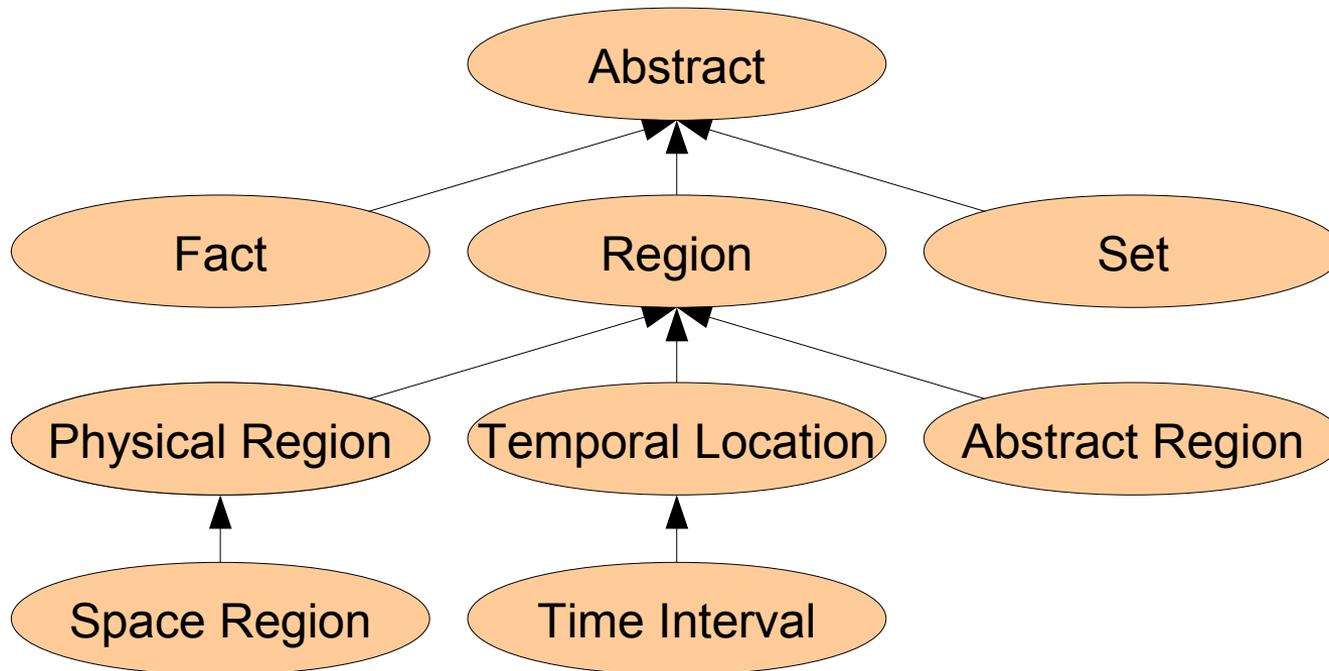
- Physical Quality: Eigenschaften von physikalischen Endurants
- Temporal Quality: Eigenschaften von Perdurants
- Abstract Quality: Eigenschaften von Abstrakta



Masolo et al. (2003): *Ontology Library (final)*. WonderWeb Deliverable D18.

Abstrakta in DOLCE

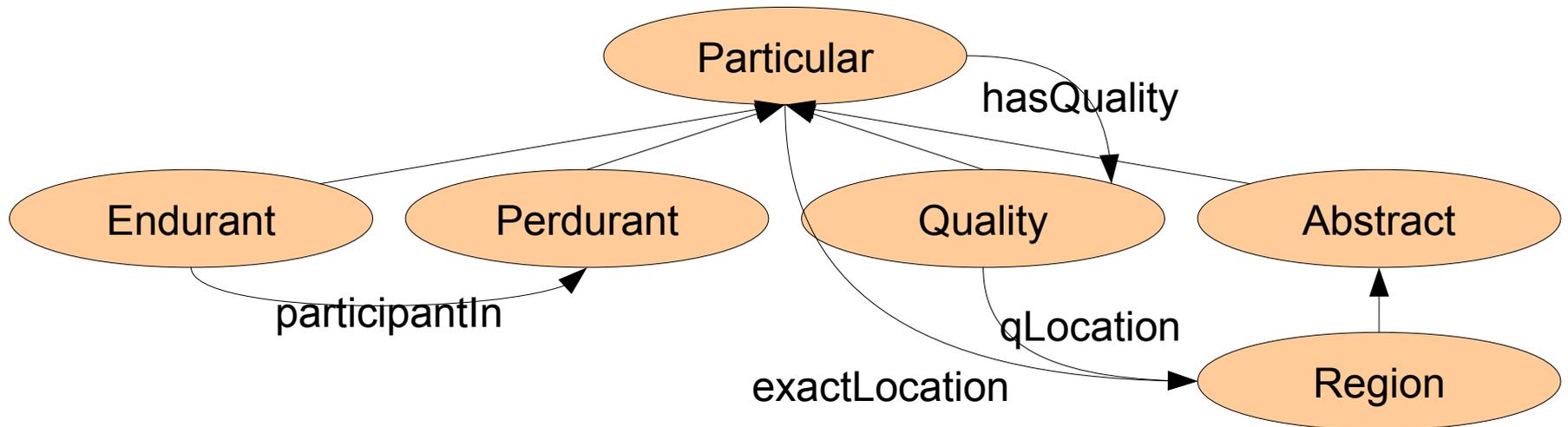
- Alle Dinge, die weder räumliche noch zeitliche Qualitäten besitzen und selbst keine Qualitäten sind



Masolo et al. (2003): *Ontology Library (final)*. WonderWeb Deliverable D18.

Grundlegende Beziehungen in DOLCE

- Beziehungen zwischen ausgewählten Basis-Kategorien



Grundlegende Relationen in DOLCE



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Teil-Ganzes-Beziehung
 - part/partOf
 - properPart/properPartOf
 - temporaryPart/temporaryPartOf
- Particulars haben gemeinsame Teile
 - overlaps
- Particulars haben gemeinsame Eltern
 - sibling-part
- ...

Formalisierung in DOLCE

- Viele der Definitionen auch ausmodelliert
 - soweit in OWL möglich
 - z.B. Teil-Ganzes-Beziehungen:

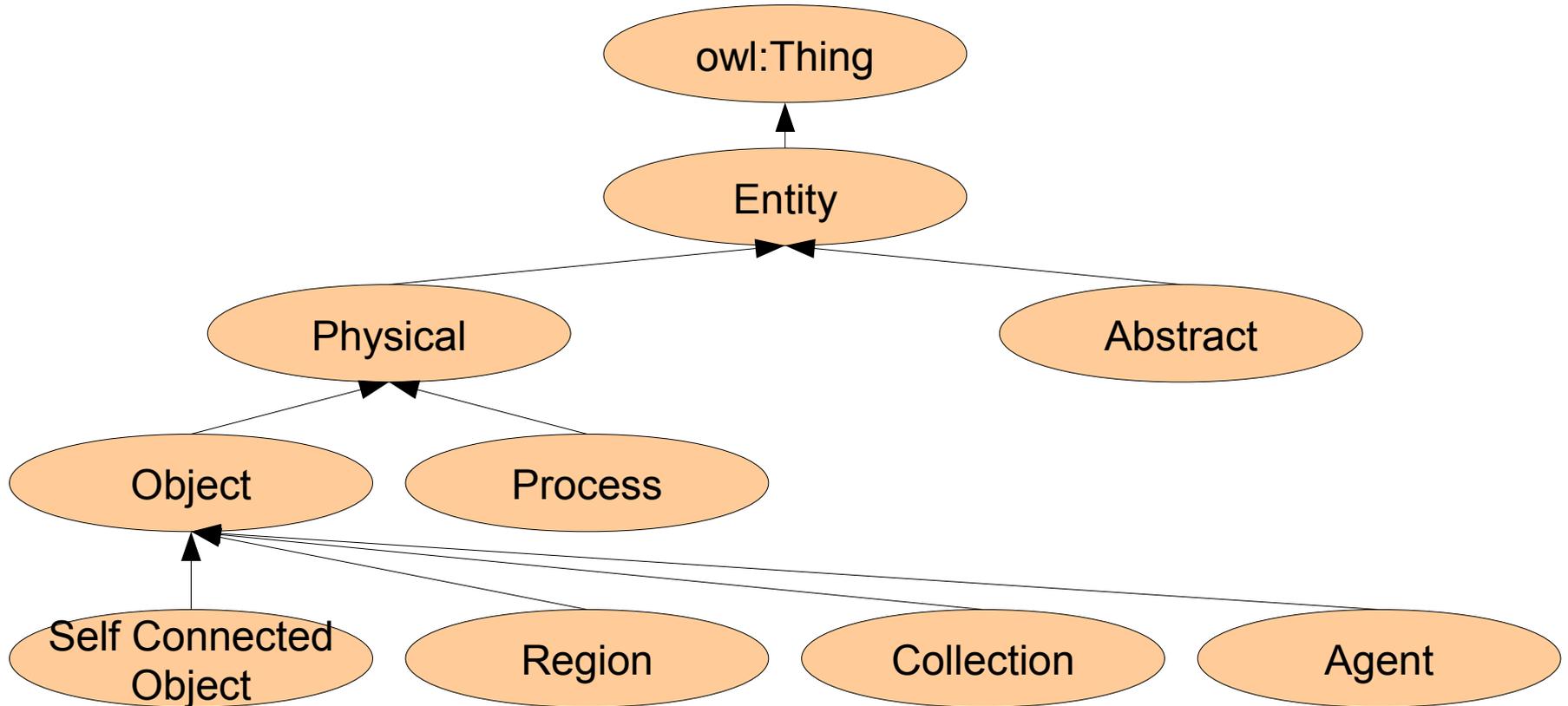
```
:endurant rdfs:subClassOf
  [ a owl:Restriction ;
    owl:onProperty :part ;
    owl:allValuesFrom :endurant ] .
```

- z.B. Definition von physikalischen Endurants:

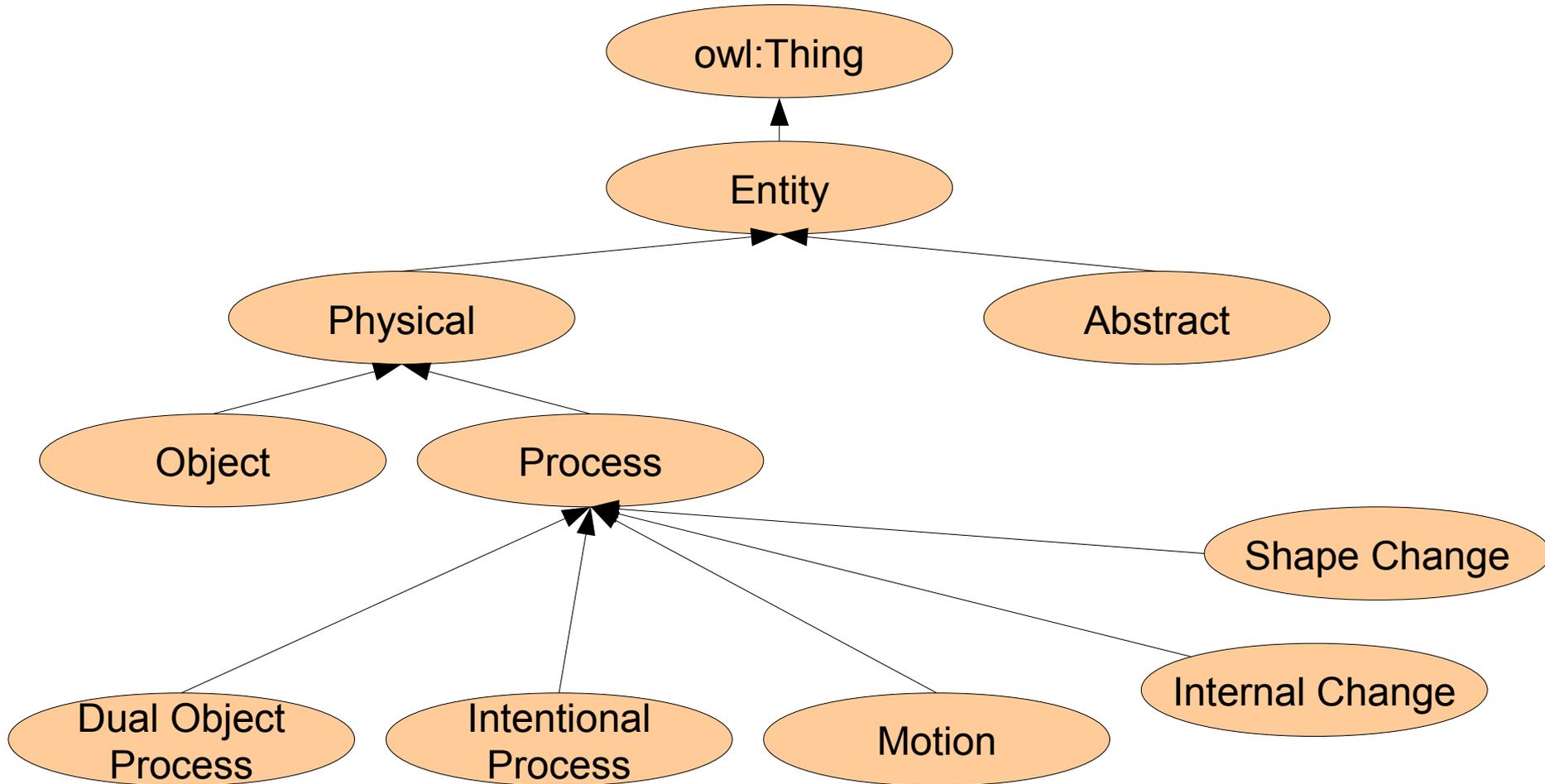
```
:physicalEndurant rdfs:subClassOf
  [ a owl:Restriction ;
    owl:onProperty hasQuality ;
    owl:someValuesFrom :SpatialLocation ] .
```

- Suggested Upper Merged Ontology
- Aus verschiedenen früheren Top-Level-Ontologien zusammengefügt
- ca. 1000 Klassen
- stark formalisiert in KIF (Knowledge Interchange Format)
 - "Übersetzung" in OWL verfügbar

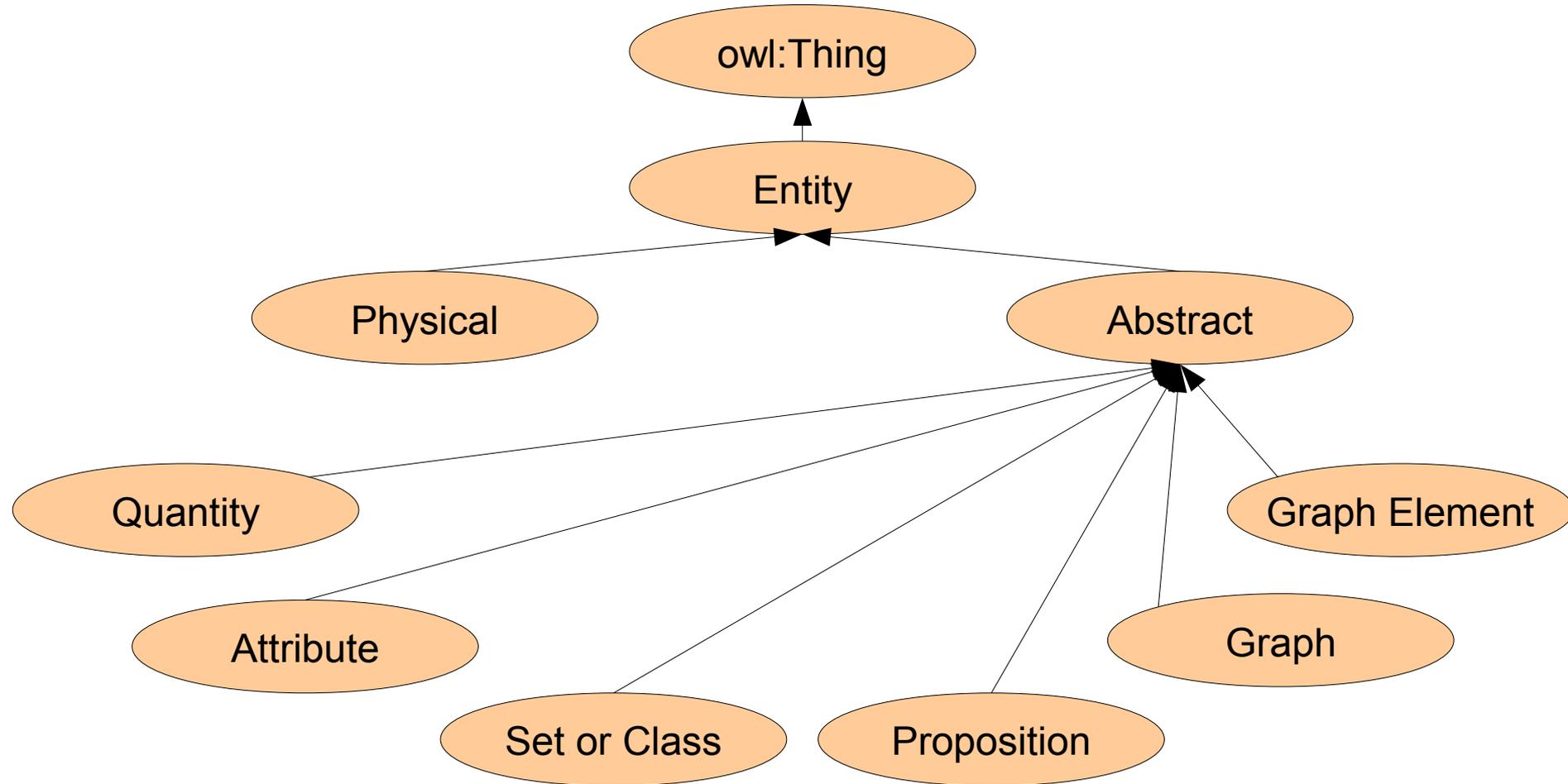
Top-Kategorien in SUMO (1)



Top-Kategorien in SUMO (2)



Top-Kategorien in SUMO (3)



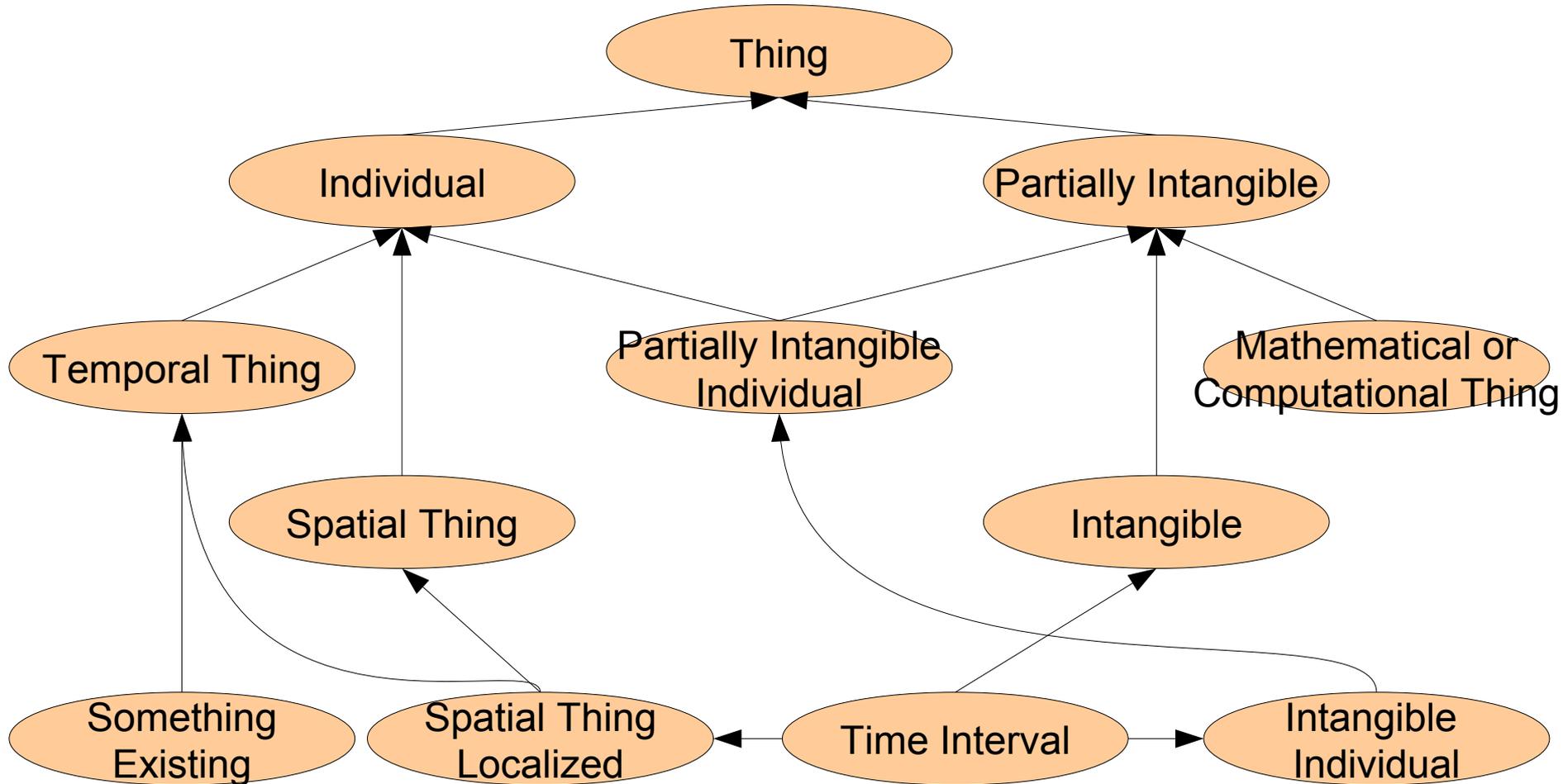
Formalisierungen in SUMO



- Axiomatisierung von Klassen
 - `Entity disjointUnionOf (Physical Abstract) .`
 - `Physical disjointUnionOf (Object Process) .`
 - ...
- Axiomatisierung von Definitionen
 - z.B.: Self-Connected Object
 - `SelfConnectedObject(x)`
 $\Leftrightarrow \forall p1, p2 : x = mereologicalSum(p1, p2)$
 $\rightarrow p1 \text{ connected } p2 .$
- Axiomatisierung von Allgemeinwissen
 - z.B.: Physikalische Objekte existieren zu (mindestens) einem Zeitpunkt an einem Ort:
 - `Physical(x) $\Leftrightarrow \exists t, p : x \text{ located } p \wedge x \text{ time } t .$`

- von *EnCyClopedia*
- formalisiert in eigener Sprache (CycL)
- Top Level und teilweise sehr tiefe allgemeine Ontologie
- ~250.000 Klassen
- Freie Variante OpenCyc
 - Existiert auch in OWL
 - und als Linked Data Endpoint

Basiskategorien von CyC

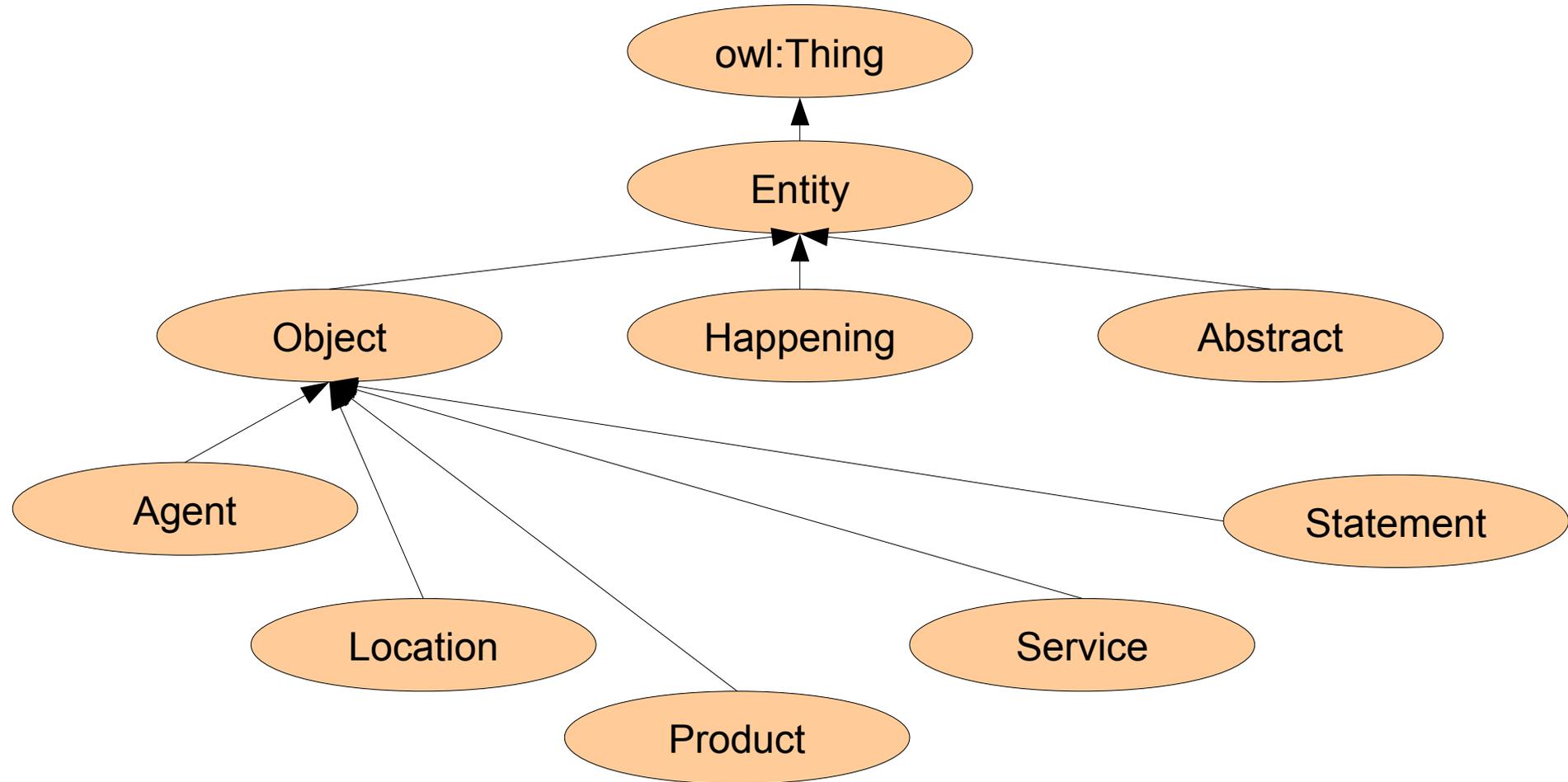


Formalisierungen in CyC

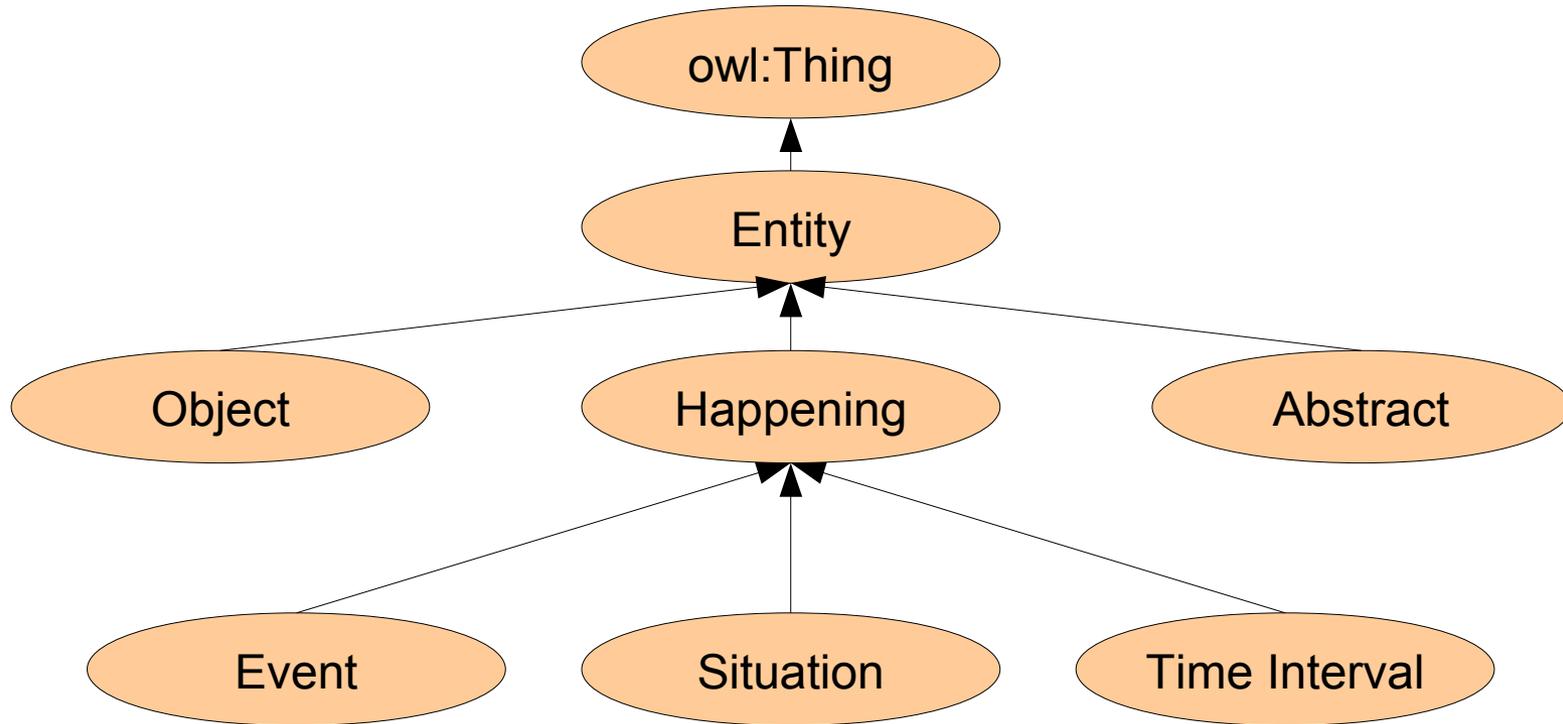
- Hauptsächlich Klassenhierarchie
- Viele Mehrfachvererbungen
- ~2,2 Mio. Axiome
 - Subklassen
 - Alternative Begriffe

- PROTo ONTology
- Vier Ebenen
 - System-Ontologie: System-Kategorien (Entity, Entity Source, Lexical Resource)
 - Top Level: Grundlegende Begriffe
 - Upper Level: Breite allgemeine Ontologie
 - Knowledge Management: Spezialkonzepte für Wissensverarbeitung
 - z.B. Weighted Term, User Profile, ...
- ~300 Klassen, ~100 Relationen

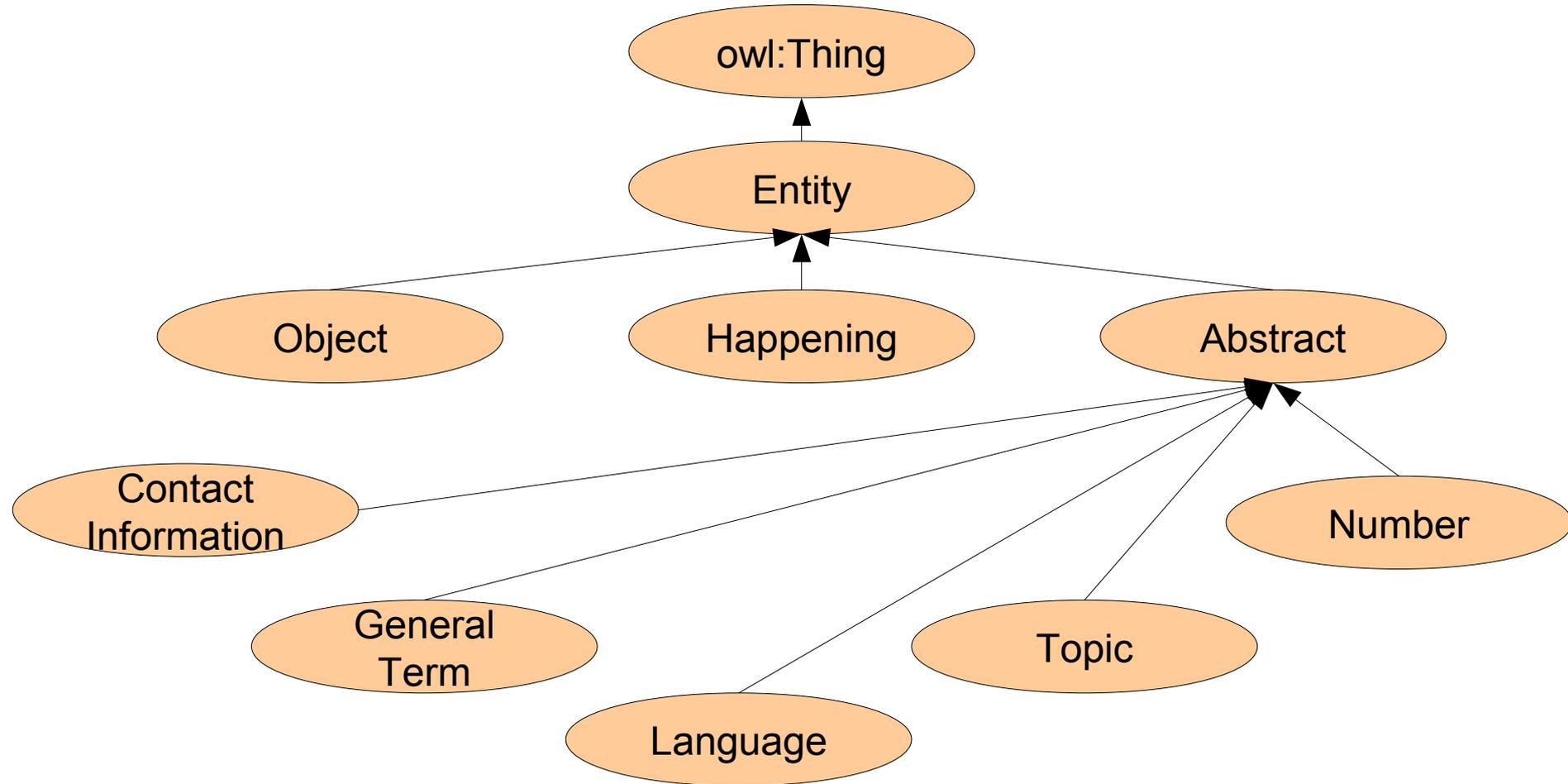
Basiskategorien von PROTON (1)



Basiskategorien von PROTON (2)



Basiskategorien von PROTON (3)



Relationen in PROTON



- Allgemeine Teil-von-Beziehung: part-of
- Zahlreiche spezifische Beziehungen, z.B.
 - Verwandtschaften (Kind, Eltern, ...)
 - Aussagen (Quelle, Urheber, ...)
 - Organisationen (Subunternehmen, Chef, Angestellter von, ...)
 - ...

- Auch Datenattribute
 - Längen-, Breitengrad
 - Anzahl Angestellte
 - ...

Formalisierung in PROTON



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- PROTON ist nur schwach formalisiert
 - Subklassen, aber keine Disjunktheit
 - tw. Domain und Range von Relationen
 - wenige Relationen als transitiv oder invers gekennzeichnet

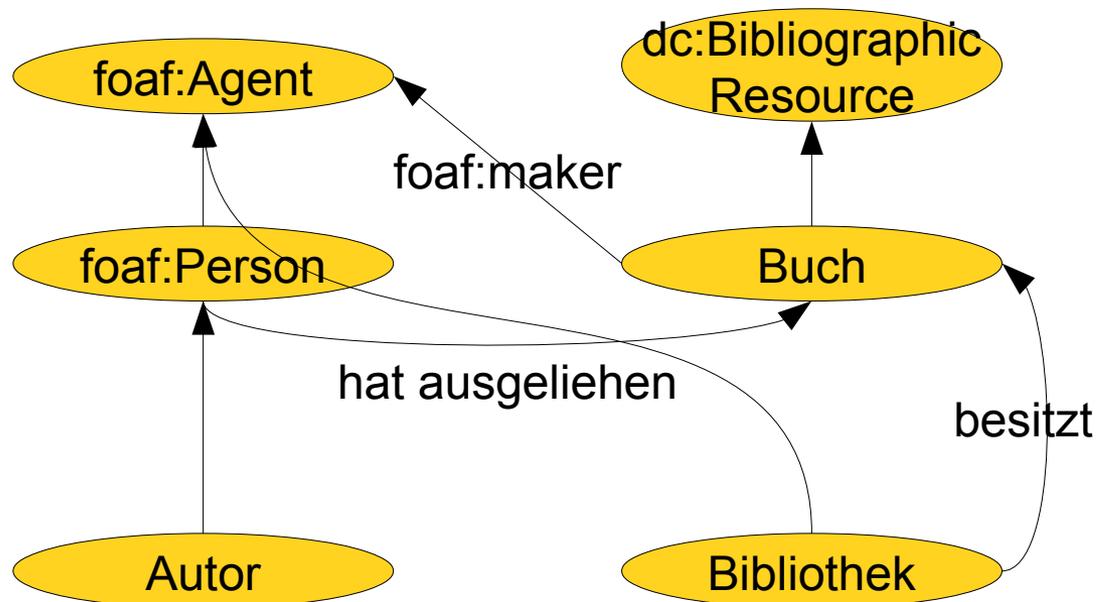


Vergleich

- Größe: CyC >> SUMO > PROTON > DOLCE
- Formalisierung: SUMO > DOLCE > CyC > PROTON
- Teilweise eklatante Unterschiede in der Kategorisierung
- Beispiel: Zeitintervall
 - In DOLCE: Eine Region (Abstract)
 - In SUMO: Eine Quantity (Abstract)
 - In PROTON: Ein Happening (~DOLCE:Perdurant)
 - In CyC: u.a. ein TemporalThing (~DOLCE:Perdurant) und ein IntangibleIndividual (~DOLCE:NonPhysicalEndurant)
- Top-Level-Ontologien in der Regel nicht kompatibel!
- Teilweise Bestrebungen (z.B.: SmartSUMO = DOLCE+SUMO)

Beispiel: Ontologie an Top-Level-Ontologie ausrichten

- Betrachten wir das Beispiel aus der zweiten Übung
 - der Einfachheit halber ohne Datenattribute
- Ziel: Wiederverwendung von Top-Level-Ontologien



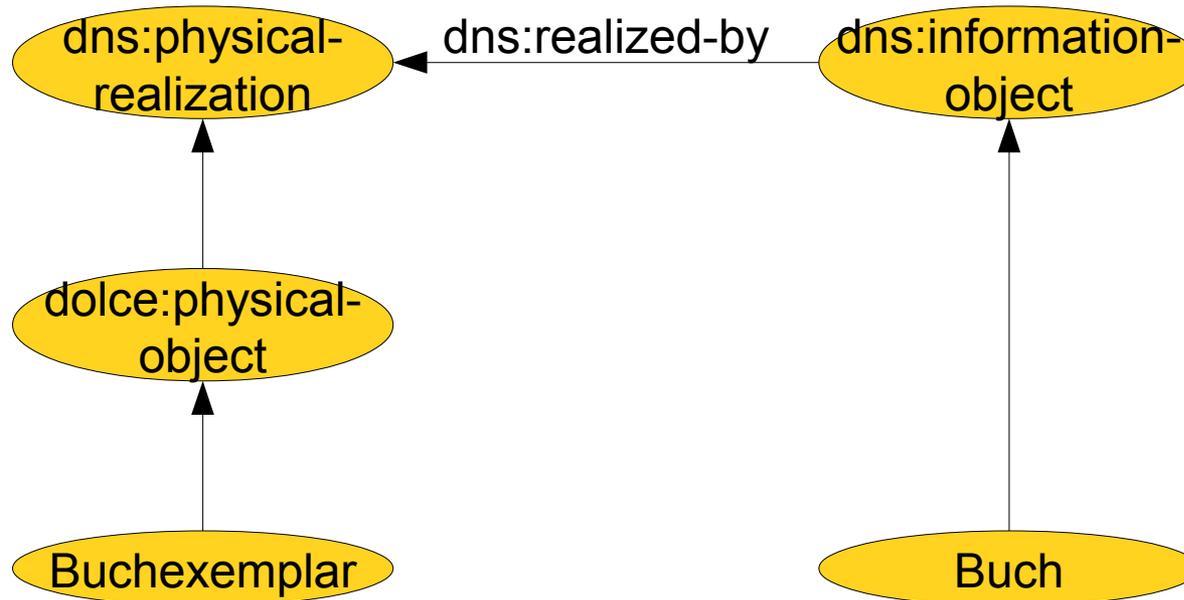
Beispiel: DOLCE-Bibliothek



- Erster Schritt: passende Superklassen suchen
 - Person/Agent \sqsubseteq dns:rational-physical-object
 - Bibliothek \sqsubseteq social-units:organization
 - Was machen wir mit Büchern?
 - Vorschlag 1:
Buch \sqsubseteq dns:information-object (... \sqsubseteq dolce:non-physical-object)
 - Vorschlag 2:
Buch \sqsubseteq dns:non-agentive-physical-object

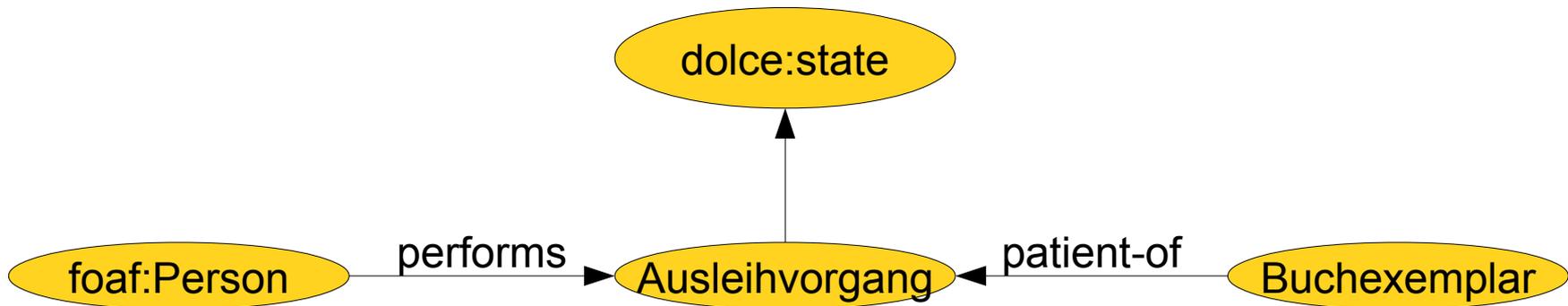
Beispiel: DOLCE-Bibliothek

- Es könnte sein, dass beides nützlich ist:

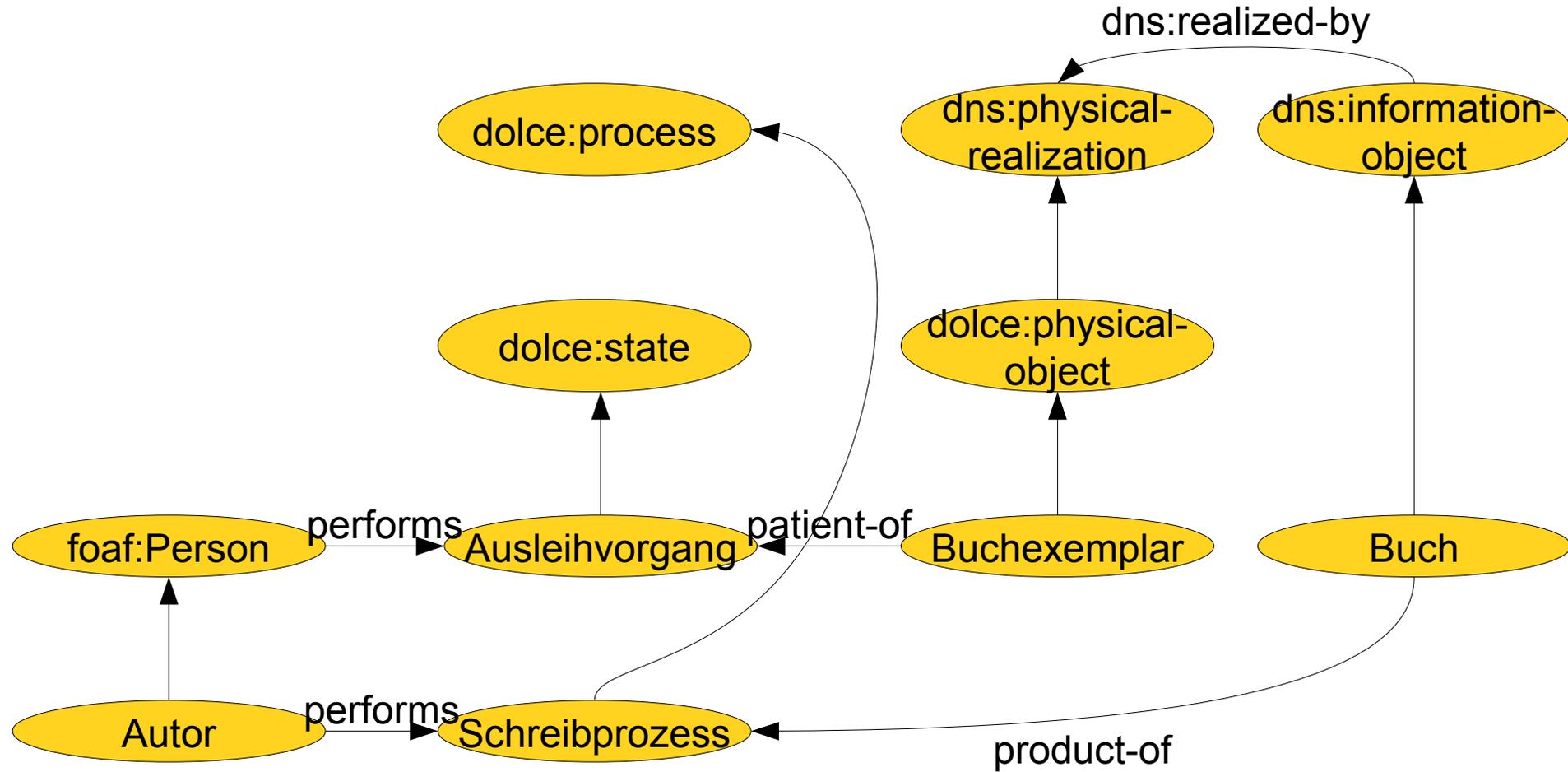


Beispiel: DOLCE-Bibliothek

- Was machen wir mit "hat Ausgeliehen"?
 - Der Ausleihvorgang ist ein Perdurant, an dem Buch und Person partizipieren
 - Vorschlag: ein state (wenn man zwei Ausleihvorgänge summiert, werden sie zu einem längeren)
- Das gilt ähnlich auch für "Autor von"



Beispiel: DOLCE-Bibliothek



Beispiel: DOLCE-Bibliothek



- Was passiert hier?
- Unsere Ontologie wird exakter
 - wir werden gezwungen, einige Dinge genauer zu fassen
 - z.B. Buchexemplar vs. Buch als Informationseinheit
- Unsere Ontologie wird komplexer
 - neue Klassen
- Wir bekommen Formalisierung geschenkt
 - z.B.: Bücher und Personen sind automatisch disjunkt

Beispiel: PROTON

- In PROTON finden wir viele Kategorien schon vor:
 - z.B. Book, Person, Organization (als Superklasse zu Bibliothek)
 - Book ist eine Subklasse von Statement, also nicht das physikalische Buch
 - Genauso Relationen: documentAuthor
- Ausleihe:
 - können wir ähnlich wie im DOLCE-Fall über ein Event modellieren

- Top-Level-Ontologien bieten (domänenunabhängige) Basiskategorien und -relationen
- Beispiele:
 - DOLCE
 - SUMO
 - CyC
 - PROTON
- Unterschiedliche Konzeptualisierungen
- Ontologie an Top-Level-Ontologie ausrichten
 - kann einiges klarer machen
 - Gewinn an Formalisierung
 - ändert meist die Ontologie

Vorlesung Semantic Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering