

Semantic Web Technologies II

SS 2008

30.04.2008

Qualität von Ontologien

Dr. Peter Haase
PD Dr. Pascal Hitzler
Dr. Steffen Lamparter
Denny Vrandečić



Content licensed under Creative Commons
<http://creativecommons.org/licenses/by/2.0/de/>

Übersicht

- Warum Evaluation?
- Qualitätskriterien
- Aspekte und Methoden der Verifikation
- Zusammenfassung: Evaluationsmodell

Was ist eine gute Ontologie?

- Ideen?

Ontology

- “An ontology is the formal specification of a shared conceptualization of a domain of interest.” (Gruber, 93)
- An ontology is a sequence of annotations, axioms, and facts. (W3C, 2004)
- Ontologie als *philosophia prima*

Ontologie

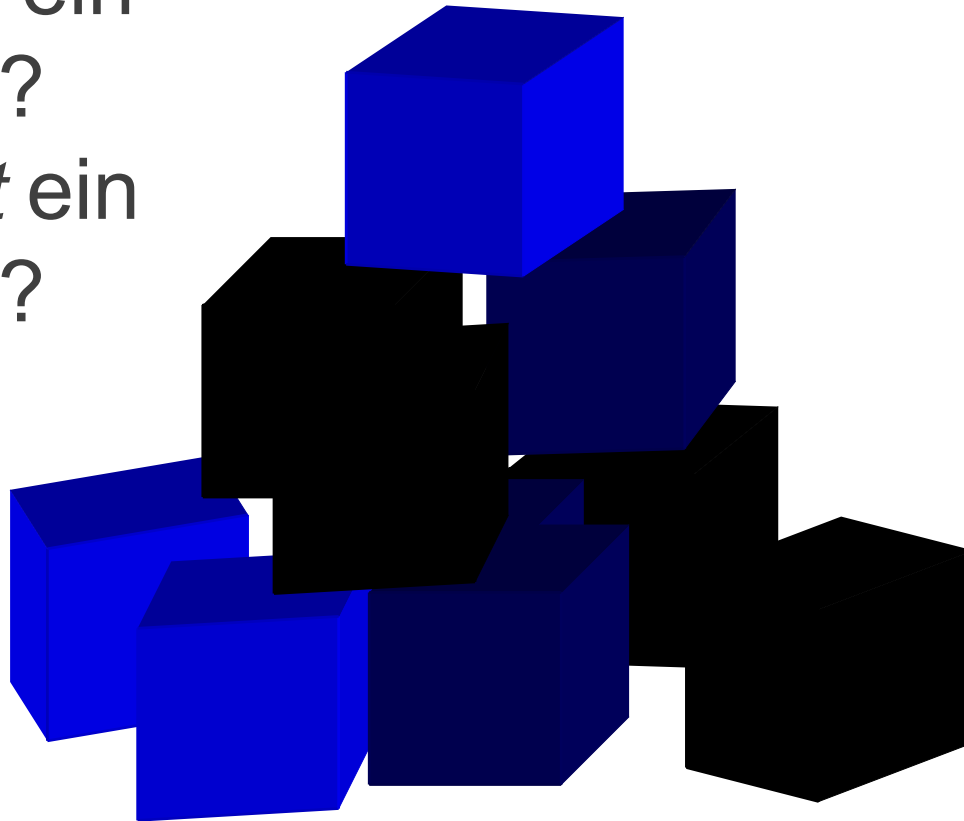
**Man hat überaus viele Ontologien sowohl derer
alten als neuen Philosophen, theils besonders.**

Ontologie, Grund-Wissenschaft, Grund-Lehre, *philosophia prima*, ist derjenige Theil der Welt-Weisheit, darinne die allgemeine Erkenntniß der Dinge abgehandelt wird... Jene ist ein Inbegriff confuser Begriffe... Sie ist von den Schul-Lehrern mit dem größten Eifer getrieben und sehr hoch erhaben, vom Cartesian und den neuern aber verachtet worden... Soll man einen kurzen Inhalt der Ontologie geben, so muß man sagen, daß sie mit zweyerley zu thun habe, erstlich mit den ersten Gründen menschlicher Erkänntniß, und anderns mit denen deutlichen begriffen von denen allgemeinen Eigenschafften, die einem Dinge überhaupt zukommen. Sie ist demnach der erste Theil der Metaphysick, und wird ihr die allgemeine Welt-Lehre gleich nachgesetzt. Denn eben diese hat auch mit allgemeinen Sätzen und Begriffen zu thun, nur mit dem Unterscheid, daß die Ontologie von den Dingen überhaupt ausser ihrer Verknüpfung unter einander handelt: die allgemeine Welt-Lehre hingegen ist mit den Dingen überhaupt, in so ferne sie unter einander verknüpfet sind, sowol in Ansehung der würckenden als End-Ursachen, und in so ferne also alle Dinge unter dem Worte: Welt begriffen sind...

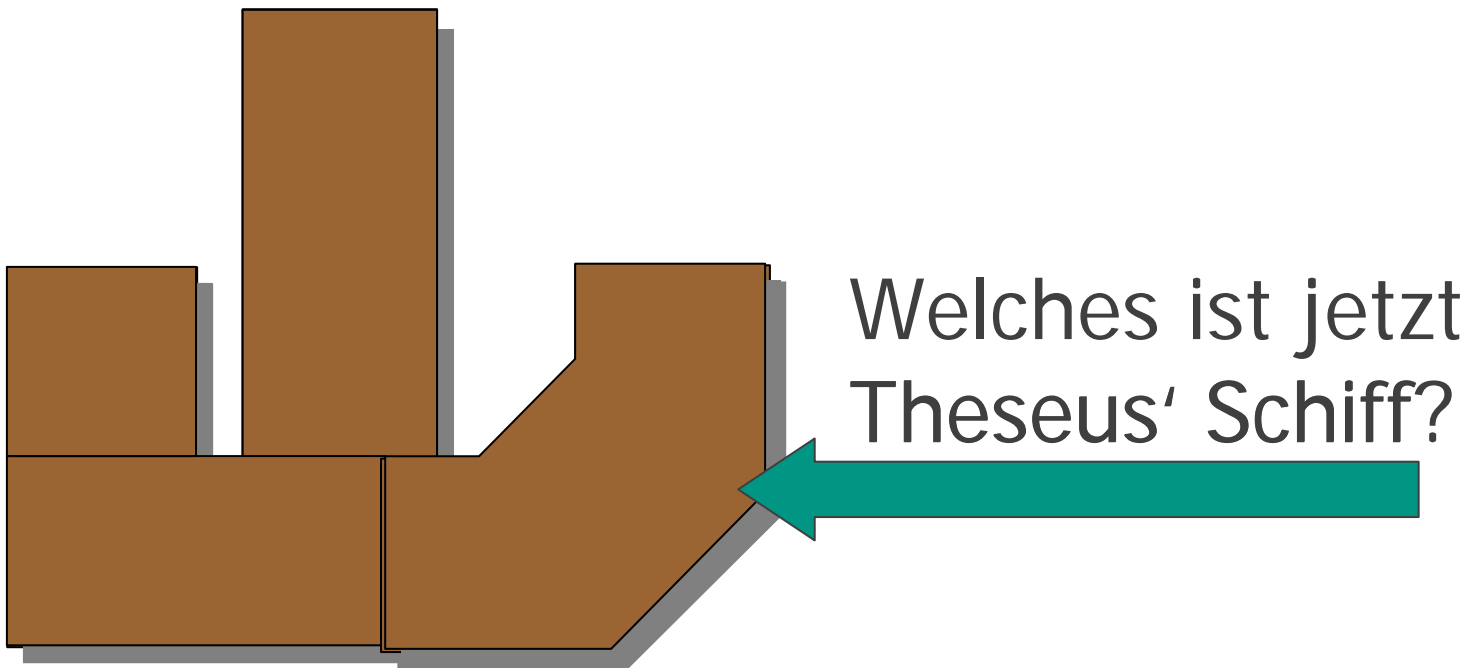
men. Sie ist demnach der erste Theil der Metaphysick, und wird ihr die allgemeine Welt-

Das Haufenproblem

Wann ist ein
Haufen ein
Haufen?
*Was ist ein
Haufen?*



Das Schiff des Theseus



Was hat das mit Ontologien zu tun?

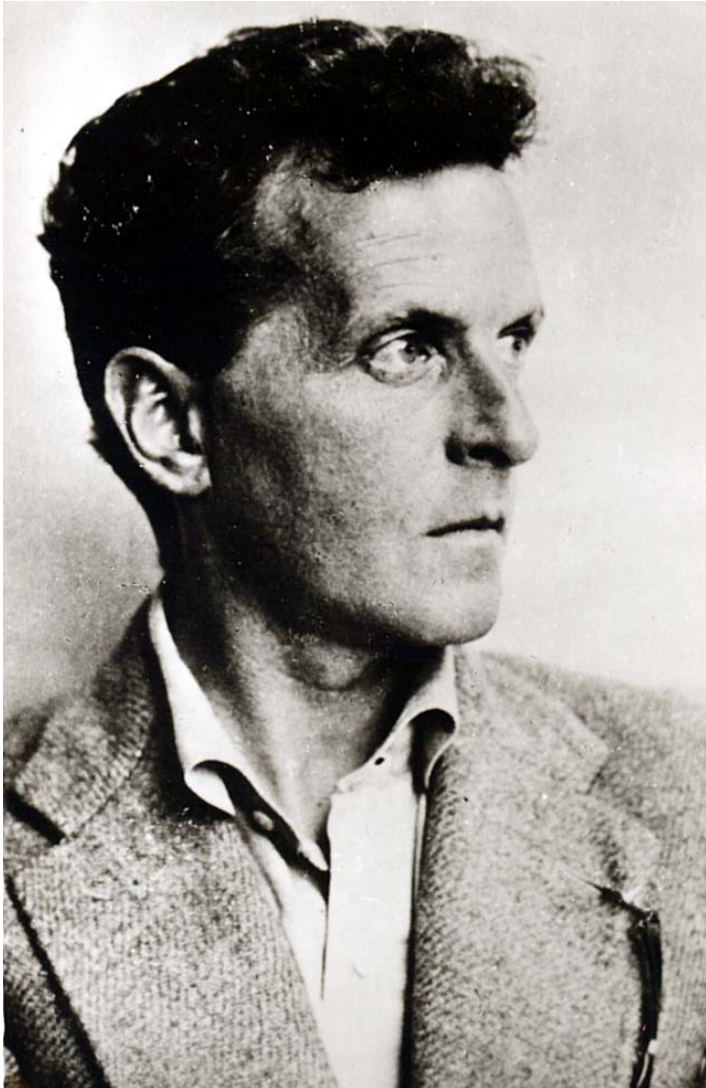
- Grundlegende Probleme
 - Wie definiert man Begriffe?
 - Identität
- Top-Level Ontologien häufig in Zusammenarbeit mit Philosophen erstellt
- Aus Fehlern lernen: unnötige Grundsatzdiskussion vermeiden
 - Universalienstreit

Ludwig Wittgenstein

1889-1951

Österreich, England

Tractatus Logico-Philosophicus



- „Die Welt ist alles, was der Fall ist“
- „Die Welt ist die Gesamtheit der Tatsachen, nicht der Dinge“
- „Wovon man nicht sprechen kann, darüber muss man schweigen“

- *“An ontology is the formal specification of a shared conceptualization of a domain of interest.” (Gruber, 93)*
 - Beschreibung der “Welt” (*domain*)
 - Beschränkung auf formal Spezifizierbares
- An ontology is a sequence of annotations, axioms, and facts. (W3C, 2004)
 - Sammlung von Tatsachen, nicht von Dingen!

Qualitätskriterien

- Accuracy
- Adaptability
- Clarity
- Completeness
- Computational efficiency
- Conciseness
- Consistency
- Organizational fitness

Evaluation

- Verifikation

*Habe ich die Ontologie
richtig erstellt?*

- Spezifikation stellt
Modell bestmöglich dar

- Validierung

*Habe ich die richtige
Ontologie erstellt?*

- Modell entspricht der
Realität

*“An ontology is
the formal specification of a shared conceptualization.”*

Basteln wir eine Ontologie

Probleme

- Wo anfangen?
- Wo aufhören?
- Wie detailliert?
- Wie vollständig?

Wichtigste Frage

- Warum erstelle ich die Ontologie?

Konsequenzen des “Warum”

- Mit welchen Werkzeugen wird auf die Ontologie zugegriffen?
- Welche Benutzer werden mit der Ontologie wie interagieren?
- Welches Wissen enthält die Ontologie?
 - Welche Fragen werden der Ontologie von Mensch oder Maschine gestellt werden, die diese richtig beantworten können soll?

Qualitätskriterien

Qualitätskriterien

- Accuracy
- Adaptability
- Clarity
- Completeness
- Computational efficiency
- Conciseness
- Consistency
- Organizational fitness

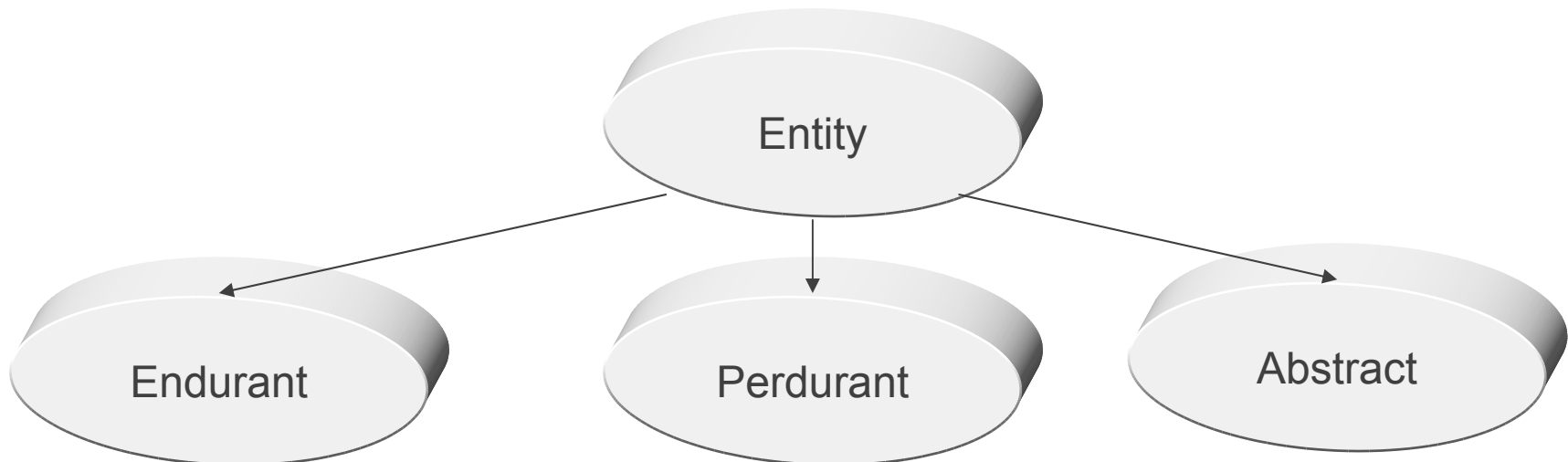
Accuracy

- Genauigkeit / Richtigkeit / Korrektheit
- Entspricht die Formalisierung der Realität?
 - Was ist mit Ontologien zu imaginären Welten?
- Entspricht die Formalisierung der Konzeptualisierung?
 - Was aber, wenn die falsch ist?
 - Tomate \sqsubseteq Gemüse

- Anpassbarkeit / Erweiterbarkeit
- Kann die Ontologie überraschungsfrei erweitert werden?
 - Haben Erweiterungen Nebenwirkungen?
 - Bsp.: Tier = \neg Pflanze
- Verhindern Redundanzen Veränderungen?
 - $A \sqsubseteq B, B \sqsubseteq C, A \sqsubseteq C$: was bedeutet es $A \sqsubseteq C$ zu entfernen?
- Wie flexibel ist die Ontologie bezüglich unerwarteter Aufgaben?

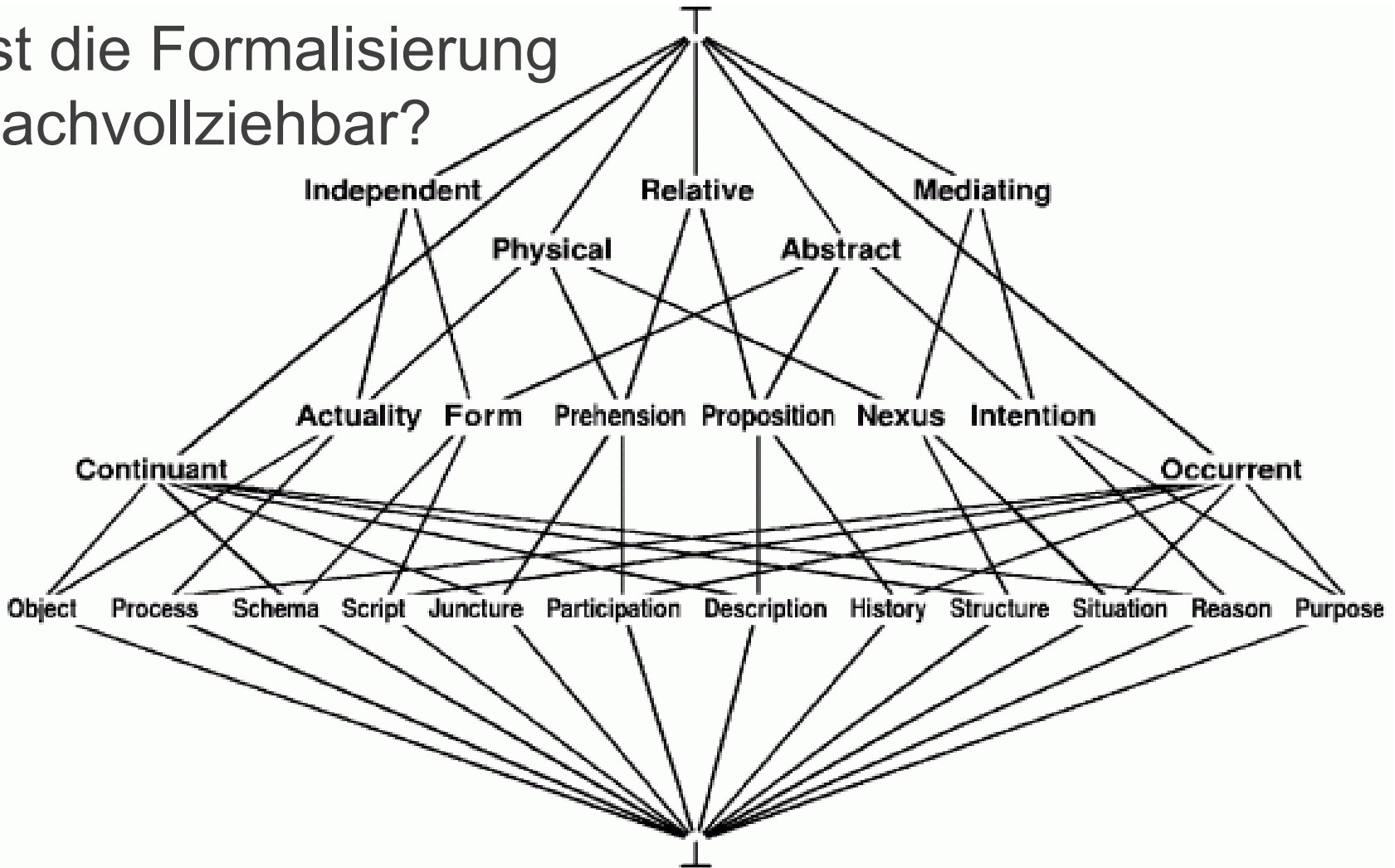
Clarity

- Verständlichkeit / Lesbarkeit
- Sind Label vollständig in relevanten Sprachen vorhanden?
- Sind die Label verständlich?
 - Beispiel: Top-Level von DOLCE



Clarity John Sowa's Top Level Ontology

Ist die Formalisierung nachvollziehbar?



Completeness

- Vollständigkeit
- Sind alle notwendigen Begriffe abgedeckt?
- Ist die Domäne vollständig dargestellt?
- Werden die Competency Questions beantwortet?
- *Nicht:* Ist die Welt vollständig beschrieben? Es geht nur um die Domäne im Rahmen der Aufgabe

Computational Efficiency

- Berechenbarkeit / Effizienz
- Reasoner schließen aus Ontologien
- Konstrukte und Muster sind unterschiedlich teuer auf unterschiedlichen Reasonern
 - z.B. KAON2 sehr schnell für konjunktive Anfragen gegen große ABoxen
 - Nicht alle Reasoner beherrschen die gesamte Sprache

Conciseness

- Knappheit
- Werden redundante Axiome vermieden?
- Beschränkt sich die Ontologie auf die Domäne?
- *Minimal ontological commitment*

Consistency

- Konsistenz / Kohärenz
- Liegen logische Inkonsistenzen vor?
 - Siehe Vorlesung nächste Woche
- Entspricht die Dokumentation der Formalisierung?
- Ist die Ontologie konsistent zu anderen verwendeten Ontologien?

Organizational fitness

- Darf die Ontologie verwendet werden?
 - Lizenzierung?
 - Benötigt die Ontologie eine Akkreditierung?
- Können die verwendeten Werkzeuge mit dieser Ontologie umgehen?
- Wurde ein Ontologiemodellierungsprozess befolgt?
- Ist die Ontologie ausreichend dokumentiert?

Methoden und Aspekte der Verifikation

Aspekte der Verifikation

- **Vokabular**
- **Syntax**
- **Struktur**
- **Semantik**
- **Repräsentation**
- **Kontext**

Vokabular

Sokrates Typ Mensch.

Sokrates Todesjahr 399 v. Chr.

<http://simia.net/id/Sokrates>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://simia.net/id/Mensch.>

<http://simia.net/id/Sokrates>

<http://simia.net/id/Todesjahr>

["-0399"^http://www.w3.org/2001/XMLSchema#gYear.](http://www.w3.org/2001/XMLSchema#gYear)

URI Unique Resource Identifier

- Einerseits zum identifizieren
- Teilweise aber auch als Locator für weitere Informationen
- Üblich: HTTP URIs, die mit dem HTTP Protokoll aufgelöst werden können
- Beispiel: <http://xmlns.com/foaf/0.1/knows>

Welche URI für Sokrates?

<http://simiana.net/id/Sokrates>

~~<http://en.wikipedia.org/wiki/Socrates>~~

<http://dbpedia.org/resource/Socrates>

<http://philosophy.stanford.edu/history/p27>

<http://aifb.uni-karlsruhe.de/id/Socrates>

<http://socrates.gr/foaf.rdf#me>

<http://socrates.gr>

~~http://imdb.com/movie/Terminator_2~~

<http://freebase.com/entity/374fd839a01>

Konkrete Daten

"-0399"^^xsd:gYear

"-399"^^xsd:gYear

"399BC"^^xsd:gYear

"-400"^^xsd:year

"-0399-01-01"^^xsd:date

"-0399"^^xsd:date

"399 BC"^^xsd:string

Datentyp: von Werkzeug
interpretierbar

Datenwert:
syntaktisch korrekt

Vokabular

- Verständliche URIs
- Gebräuchliche URIs wiederverwenden
- Konsistente URIs
- URIs für „Linked Data“
- Labels für alle URIs
- Konsistente Sprache und Stil für Labels

Aspekte der Verifikation

- Vokabular
- **Syntax**
- Struktur
- Semantik
- Repräsentation
- Kontext

Verschiedene Syntaxen

- OWL RDF/XML Serialization
- OWL/XML Presentation Syntax
- OWL Abstract Syntax
- NTriples
- N3
- Turtle
- KAON2 Syntax
- Manchester Syntax
- OWL 2 Functional Syntax

- Reihenfolge der Axiome
 - Zusammengehörige Axiome gruppieren
 - Sinnvolle Reihenfolge für Verständnis
- Ontologische Entwurfsmuster
- Abkürzungen bei der Syntax
 - Gewohnte Abkürzungen wiederverwenden
- Verbreitete Syntaxen verwenden
- Einzug sinnvoll nutzen

Aspekte der Verifikation

- Vokabular
- Syntax
- **Struktur**
- Semantik
- Repräsentation
- Kontext

- OWL Ontologien werden durch RDF-Graphen beschrieben
 - Manche Syntaxen direkt (RDF/XML, N3, ...)
 - Andere immer übersetzbar (AS, KAON2, ...)
- Graphmaße direkt anwendbar
- Häufigste Metriken in der Literatur
- Über 40 beschriebene Metriken

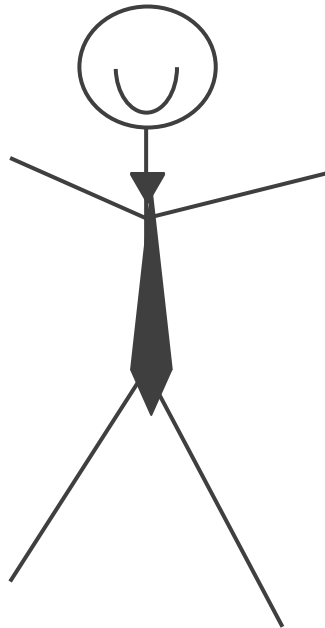
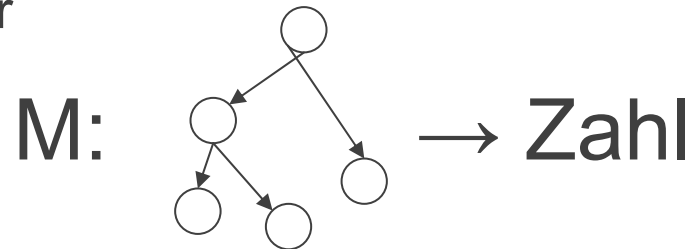
- Siehe auch nächste Woche zu Mustern

Aspekte der Verifikation

- Vokabular
- Syntax
- Struktur
- **Semantik**
- Repräsentation
- Kontext

Wie misst man Semantik?

- Syntaktische und strukturelle Eigenschaften leicht messbar
- Semantik nicht direkt messbar



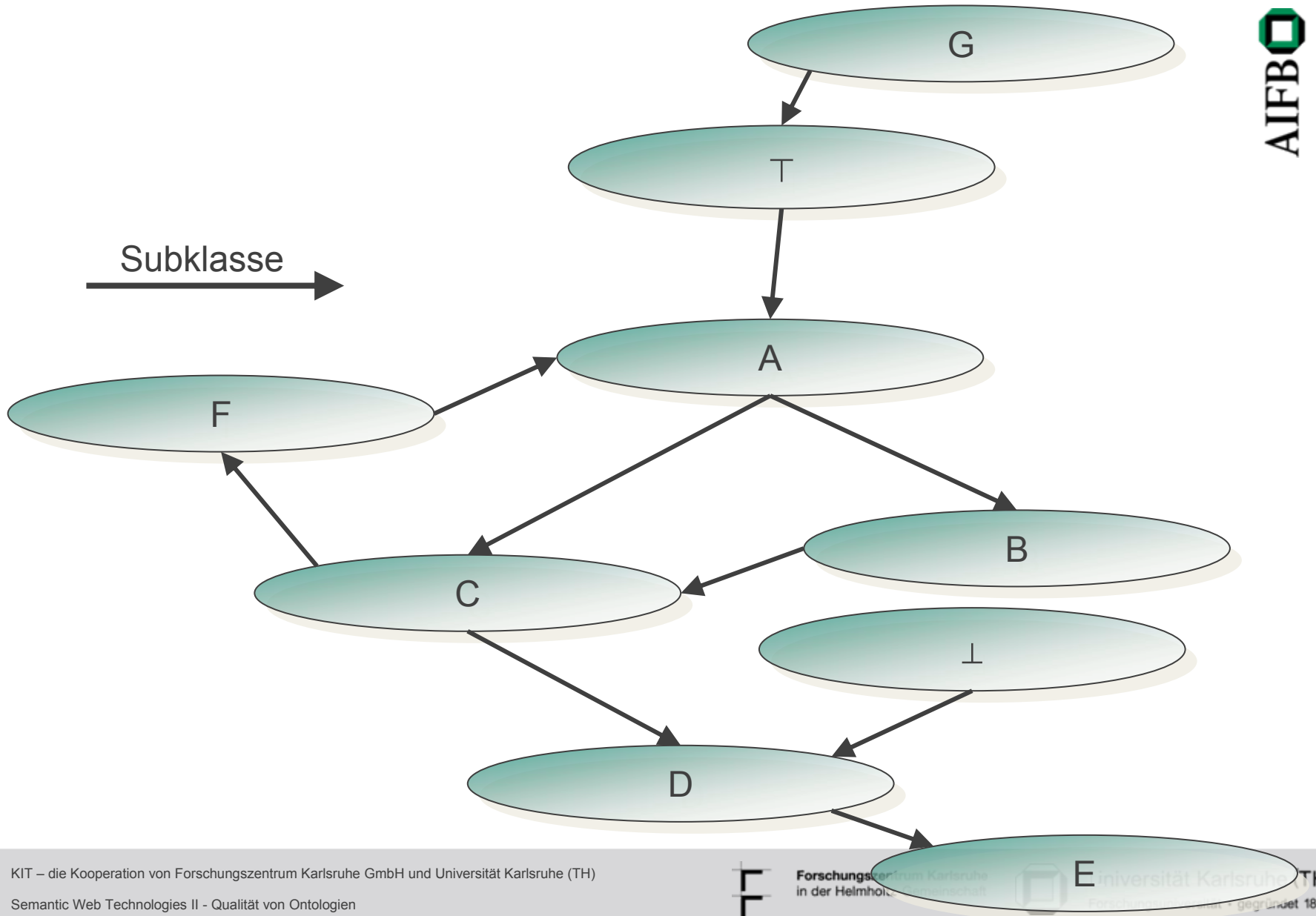
Metriken

- 2,9 Megabyte
- 29.352 Zeilen
- 2.217 Entitäten
- 2.156 Individuen
- 22 Klassen
- 20.680 Axiome
- 10.570 Beziehungsinstanzen
- 6.888 Attributsinstanzen
- 0 Subklassenaxiome

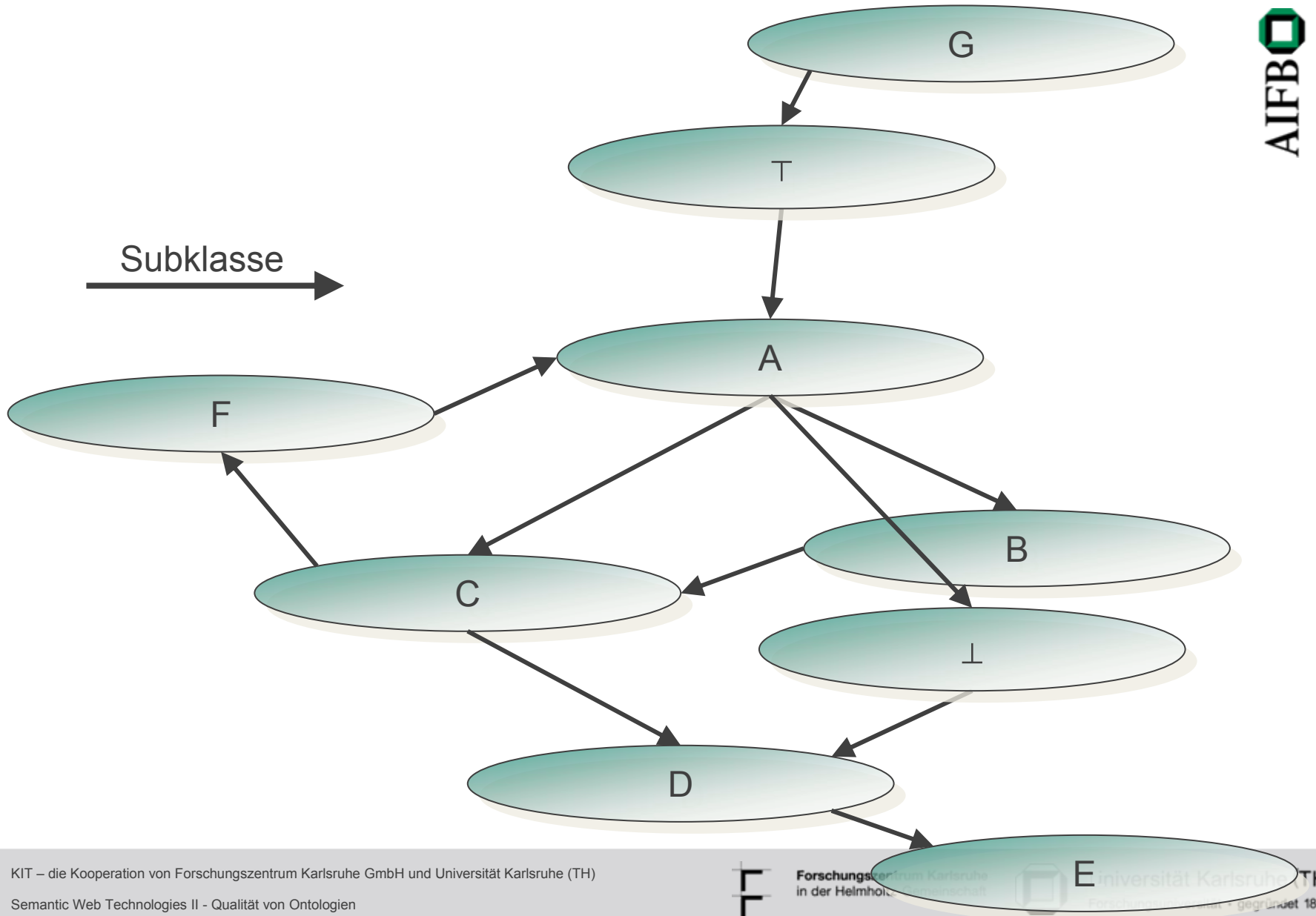
Wie misst man Semantik?

- Syntaktische und strukturelle Eigenschaften leicht messbar
- Semantik nicht direkt messbar
- Stattdessen:
 - Strukturelle Transformationen
 - Semantikerhaltend
 - Bestimmte semantische Eigenschaften für strukturelle Metriken erfassbar

Tiefe der Klassenhierarchie



Tiefe der Klassenhierarchie



Normalisierung

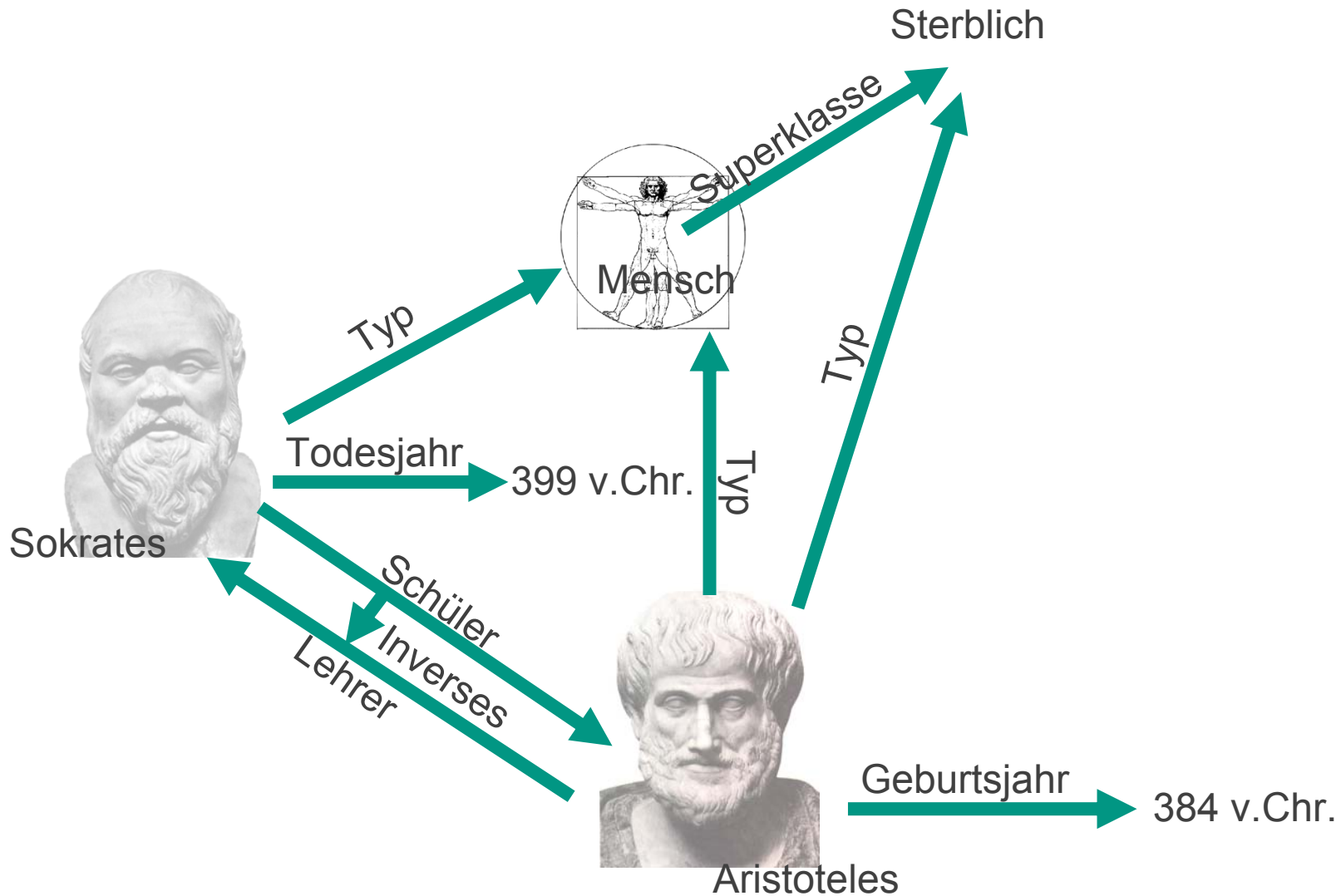
1. Alle Klassen benennen
2. Alle Individuen benennen
3. Subsumptionshierarchie materialisieren und Namen vereinigen
4. Instanzen werden in die spezifischste Klasse oder Relation runtergeschoben
5. Relationsinstanzen bezüglich Transitivität und Inversen normalisieren

- Strukturelle Metriken meist irreführend
 - „Anzahl der Klassennamen“ nicht „Anzahl der Klassen“
- Normalisierung ermöglicht es Metriken für semantische Eigenschaften zu definieren
- Am sinnvollsten bei aggregierten Webontologien

Aspekte der Verifikation

- Vokabular
- Syntax
- Struktur
- Semantik
- **Repräsentation**
- Kontext

Repräsentation



Repräsentation

ex:Sokrates ex:Schueler ex:Aristoteles.
ex:Schueler owl:inverseOf ex:Lehrer.

ex:Aristoteles ex:Lehrer ex:Sokrates.
ex:Lehrer owl:inverseOf ex:Schueler.

ex:Socrates ex:Pupil ex:Aristotle.
ex:Pupil owl:inverseOf ex:Teacher.
ex:Socrates owl:sameAs ex:Sokrates.
ex:Aristotle owl:sameAs ex:Aristoteles.
ex:Pupil owl:equivalentProperty ex:Schueler.
ex:Teacher owl:equivalentProperty ex:Lehrer.

Repräsentation

- Wie repräsentiert die Struktur die Semantik?
- Redundanz vermeiden
- Diskrepanz zwischen Semantik und Struktur oft Fehlerhinweis
 - 20 Klassennamen aber nur 3 Klassen?

Aspekte der Verifikation

- Vokabular
- Syntax
- Struktur
- Semantik
- Repräsentation
- **Kontext**

- Nutzung der Ontologie
- Welche Fragen werden an die Ontologie gestellt? (*Competency questions*)
- Stärkere Formalisierung (in Modulen)
- Kann über Standards hinausgehen
 - CBox für weitere Constraints
 - Autoepistemische Operatoren
 - Regelbasierte Integritätsbedingungen
 - Siehe auch nächste Woche

Competency questions

- CQs beschreiben welche Fragen die Ontologie beantworten können muss
- CQs sollten *vor* dem eigentlichen Ontology engineering definiert werden
- Formale CQs und Antworten ermöglichen automatische Tests
 - CQs müssen *immer* formalisierbar sein

Competency questions

- Darf Sebastian das Tiramisu essen?
- Welcher Wein passt zur Forelle Blau?
- Wie viele Portionen Hummer Napoleon können noch zubereitet werden?
- Wie viele BE hat das Tagesgericht 2?
- Welche der Desserts sind mit Schokolade aus Ecuador?

```
select ?gericht where {  
  ?gericht rdf:type food:Dessert .  
  ?gericht food:ingredient ?zutat .  
  ?zutat rdf:type food:Chocolate .  
  ?zutat food:origin food:Ecuador .  
}
```

Stärkere Formalisierung

Beispiel regelbasierte Integritätsbedingung:

X Schüler Y \wedge

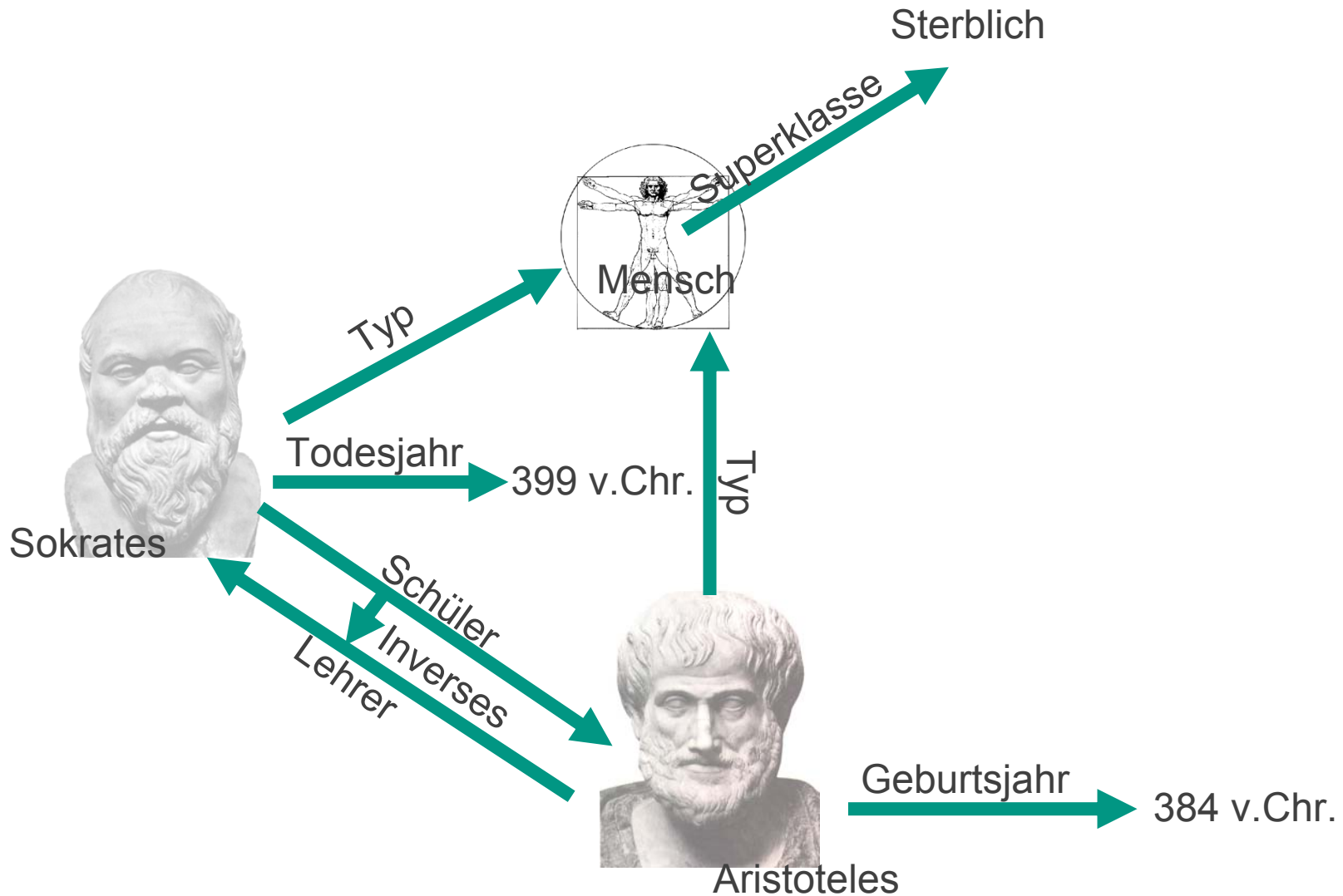
X Todesjahr T \wedge

Y Geburtsjahr G \wedge

T vor G

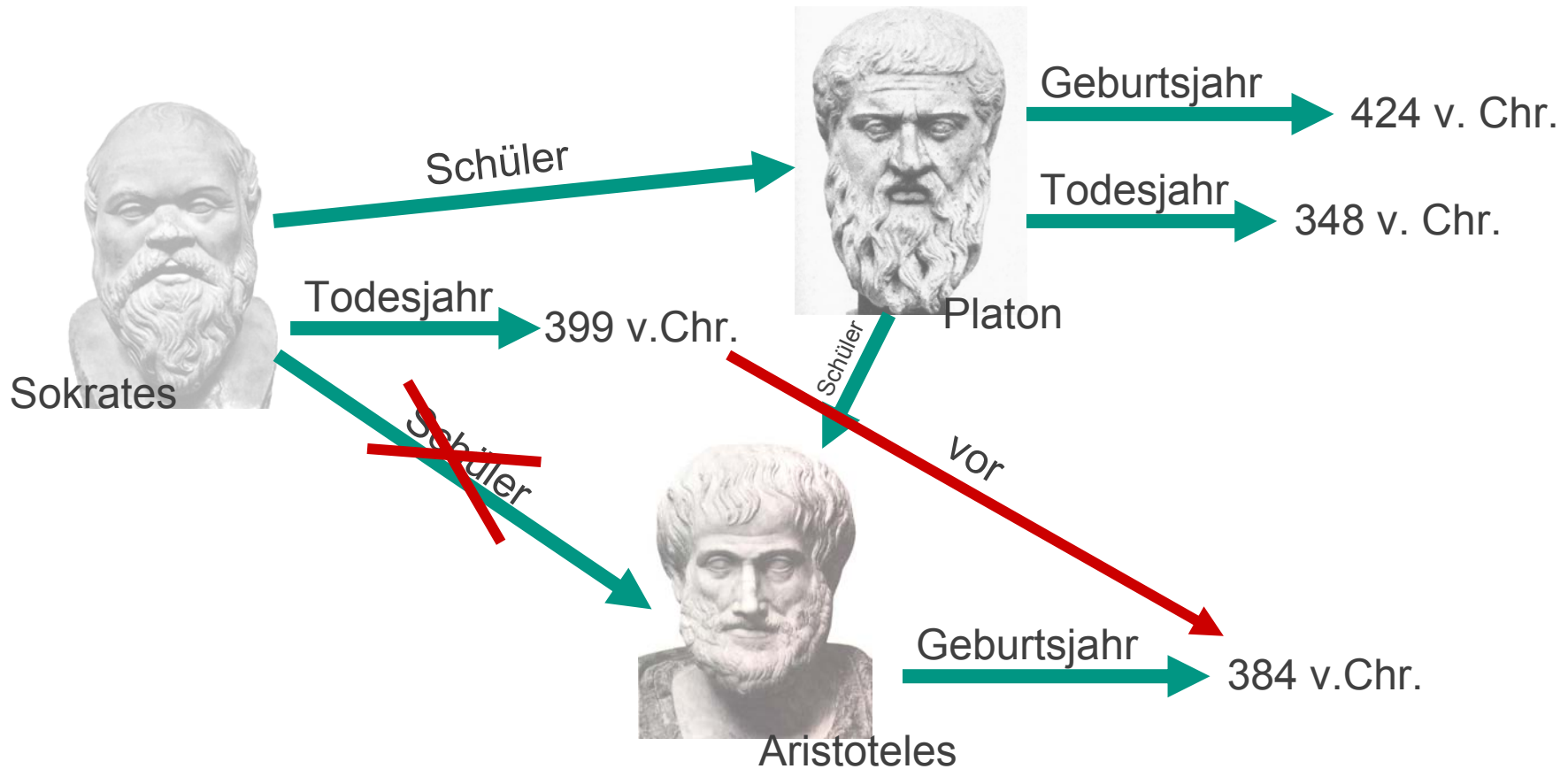
→ Fehler.

Ontologie



Ontologie

$X \text{ Schüler } Y \wedge X \text{ Todesjahr } T \wedge Y \text{ Geburtsjahr } G \wedge T \text{ vor } G \rightarrow \text{ Fehler.}$

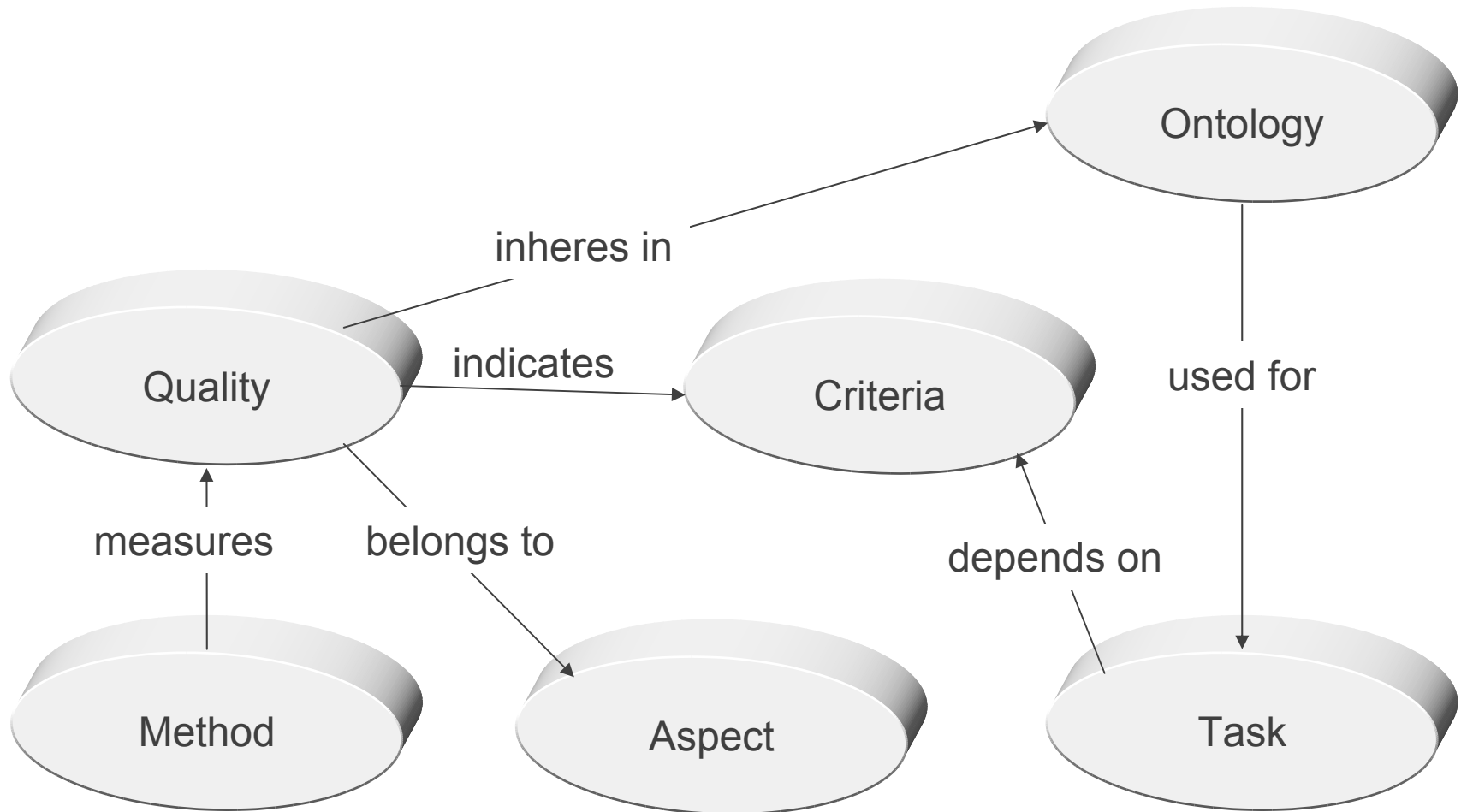


Zusammenfassung: Evaluationsmodell

Kriterien und Aspekte

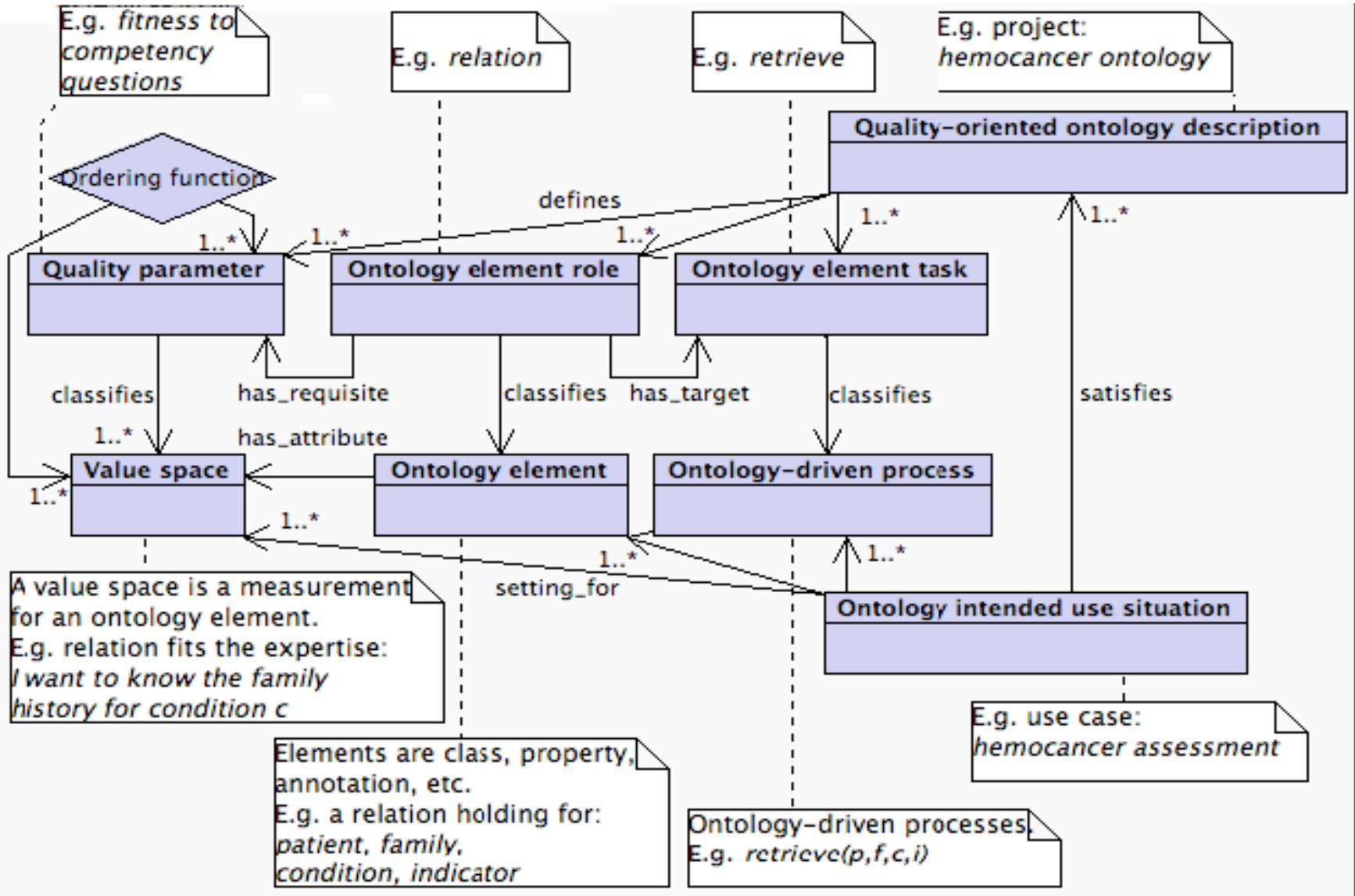
- Accuracy
- Adaptability
- Clarity
- Completeness
- Computational efficiency
- Conciseness
- Consistency
- Organizational fitness
- Vokabular
- Syntax
- Struktur
- Semantik
- Repräsentation
- Kontext

Evaluationsmodell



Evaluationsmodell

(Gangemi et al. 2006, Modeling Ontology Evaluation and Validation)



- Montag Übung
- Mittwoch 2. Vorlesung zum Thema
Ontologiequalität
 - Muster
 - Expressive Constraints
 - Reparieren von inkonsistenten Ontologien
 - Selbstreparierende ontologische Systeme