

# Referat Ontologieevaluation

## Gliederung

### THEORIE

- 1) Prozess der Ontologieentwicklung
- 2) Warum werden Ontologien evaluiert?
- 3) Ansätze der Ontologieevaluation
  - 3.1) Level der Evaluation
    - 3.1.1) Vokabular und Daten
    - 3.1.2) Relationen
    - 3.1.3) Kontext
    - 3.1.4) Syntax
    - 3.1.5) Struktur, Architektur, Design
  - 3.2) Methoden der Ontologieevaluation
    - 3.2.1) Golden Standard
    - 3.2.2) Benutzung (Application and task)
    - 3.2.3) „Data-driven“
    - 3.2.4) Bewertung anhand von definierten Kriterien und Standards (von Menschen durchgeführt)
- 4) Taxonomie-Fehler und Designanomalien
  - 4.1) Inkonsistenz
  - 4.2) Unvollständigkeit
  - 4.3) Redundanz
  - 4.4) Designanomalien
- 5.) Erweiterte Taxonomie-Fehler
  - 5.1) Semantische Inkonsistenz
  - 5.2) erweiterte Unvollständigkeit
  - 5.3) erweiterte Redundanz
- 6.) PRAXIS Fallbeispiel : NCI Thesaurus

## 1) Prozess der Ontologieentwicklung

- Themenauswahl: Welches Problem ist zu lösen?
- Was gehört zu dem Problem? Klassen, Instanzen, Beziehungen
- Umsetzung in Ontologieeditor
- **Evaluation der Ontologie**
- Anwendung und Weiterentwicklung

## 2) Warum werden Ontologien evaluiert?

Forschung auf diesem Gebiet seit ca. 1994, neuer Aufschwung durch die Entwicklung des Semantic Web

- Nutzer: viele Ontologien zu einem Themengebiet: Welche Ontologie passt am besten zu den Anforderungen?
- Konstrukteure: Hinweise für den Gestaltungsprozess und Verfeinerungen der bereits existierenden Ontologien
- (semi-) automatische Ontologieerstellungsprogramme: benötigen ebenfalls Kriterien für die Erstellung, um diese zu optimieren

## 3) Ansätze der Ontologieevaluation

### 3.1) Level der Evaluation

Häufig werden die Ontologien nicht als Ganzes evaluiert, es werden von den verschiedenen Methoden eher einzelne Teilaspekte untersucht:

- Vokabular und Daten
- Relationen
- Kontext
- Syntax
- Struktur, Architektur, Design

### 3.2) Methoden der Ontologieevaluation

Welche Methode verwendet wird ist davon abhängig, welche Ontologien evaluiert werden sollen und zu welchem Zweck. Folgende Methoden werden kurz näher vorgestellt:

- Golden Standard
- Benutzung (Application and task)
- „Data-driven“
- Bewertung anhand von definierten Kriterien und Standards (von Menschen durchgeführt)

Level	Approach to evaluation			
	Golden standard	Application-based	Data-driven	Assessment by humans
Lexical, vocabulary, concept, data	x	x	x	x
Hierarchy, taxonomy	x	x	x	x
Other semantic relations	x	x	x	x
Context, application		x		x
Syntactic	x <sup>1</sup>			x
Structure, architecture, design				x

#### 4) Taxonomie-Fehler und Designanomalien

##### 4.1) Inkonsistenz

- drei Formen von Inkonsistenz-Fehlern:
  - „Circulatory“-Fehler treten auf, wenn eine Klasse für sich als Unterklasse oder Oberklasse auf einem beliebigen Level der Ontologie-Hierarchie definiert ist
  - Distanz von 0,1 oder n
  - „Circulatory“-Fehler der Distanz 0: *OddNumber* Benennung wird als Subklasse von *NaturalNumber* festgelegt und *NaturalNumber* als Unterklasse von *OddNumber* definiert
  - Abtrennungs-Fehler: abhängig vom Typ der Dekomposition der Oberklasse in Unterklassen
- zwei Arten:
  - einfache Instanzen und Klassen in disjunkter Dekomposition und Abtrennung
  - äußere Instanzen in vollständiger Dekomposition und Abtrennung
- Semantische Inkonsistenz-Fehler: treten auf, wenn eine falsche Klassenhierarchie aufgestellt wird (wenn z.B. eine Benennung als eine

Unterklasse einer Benennung definiert wird, diese allerdings nicht wirklich dazu passt)

- Benennung *Seaplane* als Unterklasse von *AirPlane*

#### 4.2) Unvollständigkeit

- Unvollständige Benennung einer Klassifikation: tritt auf, wenn Benennungen übersehen und dadurch nicht zugeordnet werden

-Abtrennungs-Fehler: Auslassen wichtiger Axiome oder Informationen über die Klassifikation einer Benennung

-zwei Arten:

- disjunktes Auslassen von Wissen: Benennung wird vielen Unterklassen zugeordnet, aber disjunkte Wissensaxiome zwischen ihnen werden ausgelassen

- vollständiges Auslassen von Wissen: vollständige Bedingung wird bei der Dekomposition der Benennung in Unterklassen nicht berücksichtigt

#### 4.3) Redundanz

Redundanz entsteht, wenn sich eine Information mehrfach aus den Beziehungen, Klassen und Instanzen einer Ontologie schlussfolgern lässt. Dabei kann in folgende Redundanztypen unterschieden werden:

- Redundanzen bei den Beziehungen von UnterklasseVon:

-> wenn Klassen mehr als eine UnterklasseVon-Beziehung haben.

Unterteilung in direkt und indirekt:

Direkt: Eine UnterklasseVon-Verbindung besteht zwischen derselben Bezugs- und Zielklasse.

Indirekt: UnterklasseVon-Beziehung besteht zwischen Klasse und indirekter Oberklasse jeglicher Ebene.

=> Beispiel:

*BeachLocation* ist als Unterklasse von *Location* und *Place* definiert, darüber hinaus ist *Location* als *UnterklasseVon Place* definiert, es besteht also eine indirekte *UnterklasseVon*-Beziehung zwischen *BeachLocation* und *Place*, die Redundanz erzeugt. (Grafische Darstellungen für alle Beispiele geplant)

- Redundanzen bei den Beziehungen von UntereigenschaftVon:  
-> können bei der Bildung von Eigenschaftshierarchien entstehen.
- Redundanzen bei den Beziehungen von InstanzVon:

=> Beispiel: Die Instanz *Swat* (Region in Pakistan) wird als *InstanzVon Location* und *Place* definiert und *Location* als Unterklasse von *Place*. Die *InstanzVon*-Verbindung zwischen *Swat* und *Place* verursacht eine Redundanz, da *Swat* eine indirekte Instanz von *Place* und *Place* eine Oberklasse von *Location* ist.

- Identische formale Definition von Klassen, Eigenschaften und Instanzen:
  - > wenn unterschiedliche (oder dieselben) Namen zweier Klassen, Eigenschaften oder Instanzen definiert, aber sie dieselben formalen Definitionen erhalten

#### 4.4) Design-Anomalien

- Anhäufung von Eigenschaften:
  - > wenn wiederholt eine Anhäufung von Eigenschaften in unterschiedlichen Klassendefinitionen benutzt wird
- Vererbungskette:
  - > wenn die Hierarchie lang ist und die Klassen keinerlei Beschreibungen enthalten, spricht man von einer Vererbungskette. Diese sollte in Unterhierarchien aufgebrochen werden
- „Lazy Concept“:
  - > wenn eine Klasse oder Eigenschaft der Taxonomie keine Instanzen hat und in der Anwendung niemals auftaucht. Diese sollten durch speziellere oder allgemeinere Klassen oder Eigenschaften ersetzt werden, die solche Instanzen enthalten und auch genutzt werden.
- „Lonely disjoints“:
  - > wenn zwei disjunkte Geschwisterteile kreiert werden und ein Teil in der Klassenhierarchie verschoben wird, ohne dass das disjunkte Axiom zwischen ihnen gelöscht wird.

### 5.) Erweiterte Taxonomie-Fehler

#### 5.1) Semantische Inkonsistenz

Unterteilbar in drei Unterklassen:

- Schwächerer Gegenstandsbereich durch Unterklassen-Fehler
  - > Wenn Subklassen einen größeren Gegenstandsbereich enthalten als ihre Oberklassen
  - => Beispiel: Oberklasse=*Staff*, Unterklassen=*UniversityMember*, *AcademicStaff*, *AdminStaff* und *LabStaff*. *UniversityMember* ist als Unterklasse zu allgemein. Eine Unterklasse sollte immer durch die Oberklasse spezifiziert werden und das Gegenstandsgebiet einengen.

- Übertretung des Gegenstandsbereichs durch Unterklassen-Fehler: Unterklassen sollten alle Eigenschaften ihrer Elternklassen besitzen und im eigenen Bereich keine dieser Eigenschaften verletzen. Verletzung des Gegenstandsbereiches der Oberklasse: die Unterklasse besitzt zusätzliche Eigenschaften, die die Oberklasse nicht besitzt, und diese Eigenschaften verletzen einige Eigenschaften der Oberklasse.  
=> Beispiel: *VegetarischePizza* ist eine Unterklasse von *Pizza*, *ChinesischePizza* und *ItalienischePizza* sind Unterklassen von *VegetarischePizza*, obwohl diese beiden Pizzen mit Fleisch belegt sind.
- Disjunkter Gegenstandsbereich durch Unterklasse-Fehler:  
-> wenn disjunkte Gegenstandsbereichsklassen als Unterklassen von Klassen definiert werden, die einen anderen Gegenstandsbereich beinhalten  
=> Beispiel: *Getränk* und *Burger* als Unterklasse der Oberklasse *Essbares*.

## 5.2) Erweiterung der Unvollständigkeit

- Auslassung von „Functional Property“ für „single valued property“ bei OWL:  
Wenn es nur ein Wert für ein Subjekt gibt, muss diese Eigenschaft als functional erklärt werden.  
=> Beispiel: *hatBlutgruppe* als eine Eigenschaft zwischen *Person* und *Blutgruppe*. Jedes Subjekt *Person* gehört zu nur einer *Blutgruppe*, demnach sollte die Eigenschaft *hatBlutgruppe* als funktional gekennzeichnet werden, damit eine Person nur mit einer Blutgruppe assoziiert werden kann.
- Auslassung von „Inverse-Functional Property“ für „unique valued property“:  
=> Beispiel: *National\_SecurityNo* gehört zu *Person*, da es diese eindeutig identifiziert.
- Auslassung von „Sufficient Knowledge“:  
Erklärung von sufficient und necessary description

## 5.3) Erweiterung der Redundanz

- Redundanz von disjunkten Beziehungen:  
Erklärung von direkter und indirekter Disjunktheit

## **PRAXIS**

### **6. Fallbeispiel : NCI Thesaurus**

- Thesaurus aus dem Gebiet der Krebsforschung
- keine Ontologie im eigentlichen Sinne; besitzt aber einige Eigenschaften einer Ontologie
- ist ein kontrolliertes Vokabular, das Beziehungen zwischen Konzepten in einer formalisierten Struktur darstellt

### **EVALUATION**

- Konformität mit terminologischen Standards (ISO)
- Konformität mit den Prinzipien der Ontologierstellung

#### **Methode:**

qualitative Untersuchung der Online Version des Thesaurus und der OWL Datei

Untersuchung der Einträge auf Verletzungen der Standards  
Bewertung anhand von definierten Kriterien und Standards (von Menschen durchgeführt);

Untersuchung von Teilaspekten, wie Struktur, Vokabular und Relationen

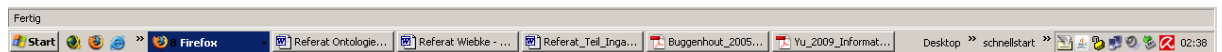
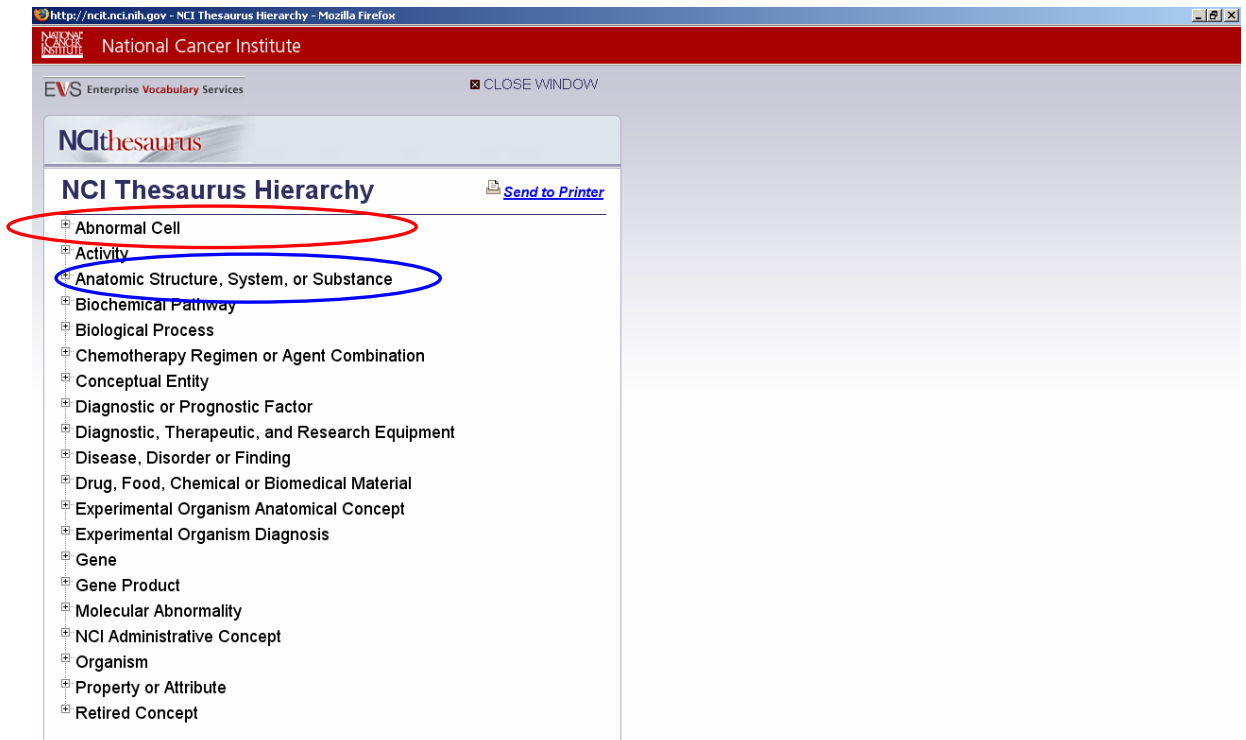
#### **Ergebnisse**

- Der Thesaurus verletzt sowohl die terminologischen Standards als auch die Standards der Ontologierstellung in vielfältiger Weise

#### **Beispiele für Verletzungen der Kriterien**

1. Auswahl der Oberklassen ist fragwürdig und wird nicht erklärt
2. Inkonsistenzen bei der Klassenbildung : Klassen und Definitionen stimmen nicht überein
3. Falsche Synonymbildung
4. die „Definitionen“ sind oft keine Definitionen, sondern Beschreibungen eines Begriffes
5. Inkonsistenzen beim Gebrauch der OWL Relationen.

# 1. Auswahl der Oberklassen ist fragwürdig und wird nicht erklärt



„Abnormal Cell“ ist eine Klasse auf der obersten Hierarchieebene, während die Klasse „Cell“ auf der dritten Hierarchieebene als Unterklasse zu Microanatomic Structure konzipiert ist, die wiederum eine Unterklasse zu Anatomic Structure System or Substance ist.

„Abnormal Cell“ ist im Kontext des NCIT keine Zelle und die Klassen stehen in keiner Beziehung zueinander.



## 2. Inkonsistenzen bei der Klassenbildung : Klassen und Definitionen stimmen nicht überein

NCI Thesaurus - Mozilla Firefox

http://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI%20Thesaurus&code=C54266

ontologie im web

rights.info News Informationsplattform... URHG News Immaterblog.de

NCI Thesaurus NCIT.pdf (applicati... DBV-ontengineer... Ontology Engineeri... NCI Metathesaurus NCI Thesaurus NCI Metathesaurus NCI Thesaurus

NCIthesaurus

feeling Search ?

Exact Match Begins With Contains

Quick Links

Home | View Term Hierarchy | Subsets Suggestion | Help

Emotion (Code C54266)

Suggest changes to this concept

Terms & Properties Relationships Synonym Details View All

View in Hierarchy View History

**Terms and Properties**

**Definition:** That aspect of consciousness devoted to affect or feeling; a strong feeling, aroused mental state, or intense state of drive or unrest directed toward a definite object, with physiological, somatic, and behavioral components.

**Preferred Name:** Emotion

**NCI Thesaurus Code:** C54266

**NCI Metathesaurus CUI:** C0013987 ([see NCI Metathesaurus info](#))

**Synonyms & Abbreviations:** ([see Synonym Details](#))

Emotional

http://ncit.nci.nih.gov/ncitbrowser

Start Firefox Referat Ontologie... Referat Wiebke... Referat\_Tel\_Inga... Buggenhout\_2005... YU\_2009\_Informat... Desktop » schnellstart » 02:58

Die Klasse „Emotion“ wird definiert als „strong feeling“, während die Klasse „Feelings“ als „emotional Sensation“ definiert wird. Hier treten terminologische Probleme auf (eine Klasse/Konzept definiert eine andere Klasse/Konzept). Außerdem resultiert aus der terminologischen Unsicherheit eine Diskrepanz zwischen Klassenbildung und Definition. Eigentlich müssten die Klassen „Feelings“ und „Emotion“ Synonyme sein.

Feelings (Code C77965)

Terms & Properties Relationships Synonym Details View All

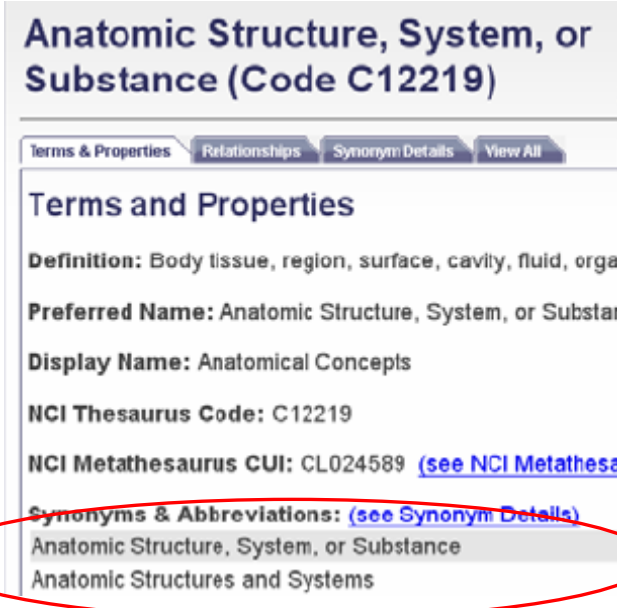
**Terms and Properties** View All

**Definition:** Emotional sensations

**Preferred Name:** Feelings

**NCI Thesaurus Code:** C77965

### 3. Falsche Synonymbildung



**Anatomic Structure, System, or Substance (Code C12219)**

Terms & Properties Relationships Synonym Details View All

**Terms and Properties**

**Definition:** Body tissue, region, surface, cavity, fluid, organ

**Preferred Name:** Anatomic Structure, System, or Substance

**Display Name:** Anatomical Concepts

**NCI Thesaurus Code:** C12219

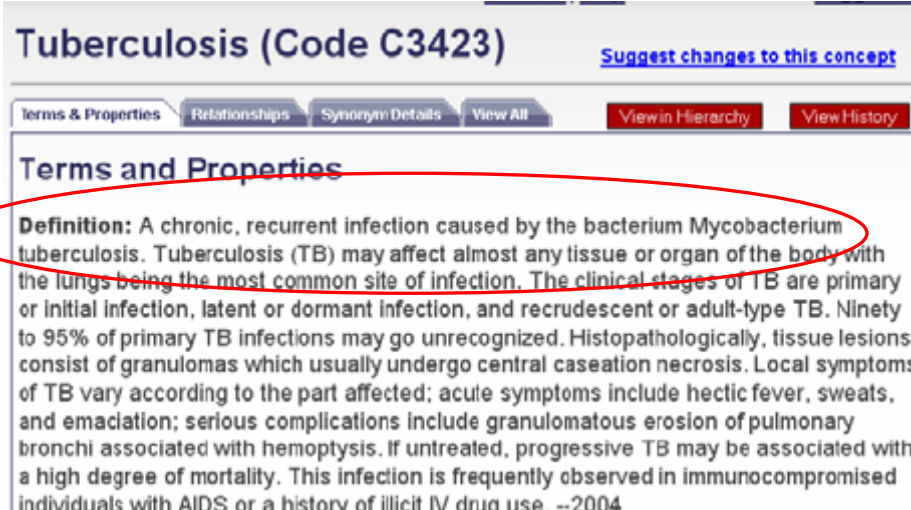
**NCI Metathesaurus CUI:** CL024589 ([see NCI Metathesaurus](#))

**Synonyms & Abbreviations:** ([see Synonym Details](#))

- Anatomic Structure, System, or Substance
- Anatomic Structures and Systems

Da der Begriff „Substance“ weder eine Struktur noch ein System ist dürfen die beiden Begriffe nicht synonym verwendet werden.

### 4. die „Definitionen“ sind oft keine Definitionen, sondern Beschreibungen eines Begriffes



**Tuberculosis (Code C3423)** [Suggest changes to this concept](#)

Terms & Properties Relationships Synonym Details View All [View in Hierarchy](#) [View History](#)

**Terms and Properties**

**Definition:** A chronic, recurrent infection caused by the bacterium *Mycobacterium tuberculosis*. Tuberculosis (TB) may affect almost any tissue or organ of the body with the lungs being the most common site of infection. The clinical stages of TB are primary or initial infection, latent or dormant infection, and recrudescence or adult-type TB. Ninety to 95% of primary TB infections may go unrecognized. Histopathologically, tissue lesions consist of granulomas which usually undergo central caseation necrosis. Local symptoms of TB vary according to the part affected; acute symptoms include hectic fever, sweats, and emaciation; serious complications include granulomatous erosion of pulmonary bronchi associated with hemoptysis. If untreated, progressive TB may be associated with a high degree of mortality. This infection is frequently observed in immunocompromised individuals with AIDS or a history of illicit IV drug use. --2004

Der erste Satz hätte als Definition genügt. Alle weiteren Informationen hätten besser in eine Scope Note gepasst