

# Ontologien mit RDF Schema

Markus Stocker

11. Juni 2018

# Rekapitulation

- Was ist RDFS? Wofür verwendet man die Sprache?
- Was ist eine Klasse?
- Was ist `rdfs:subClassOf`?
- Was ermöglicht `rdfs:range`?

# Übersicht

- Was ist Ontologie?
- Bestandteile und Aufbau
- Ontologien mit RDFS
- Ontologien entwickeln
- Protégé

# Was ist Ontologie?

- Eine Disziplin der Philosophie
- Die Lehre vom Sein
- Befasst sich mit den Grundstrukturen der Wirklichkeit
- Die fundamentalen Klassen und Beziehungen existierender Dinge
- Aristoteles 10 Kategorien: Substanz, Quantität, Relation, ...

# Ontologie in der Informatik

- Teil der Wissensrepräsentation im Gebiet Künstliche Intelligenz
- Beschreibung von “Wissen” über eine Domäne
- Begriffe und Beziehungen eines Gegenstandsbereiches
- Beispiel: Erdähnliche und Gas Planeten mit Satelliten
- Eine maschinenverarbeitbare (formale) Beschreibung
- Beispiel: `TerrestrialPlanet rdfs:subClassOf Planet`

# Ontologie in der Informatik

- Konsensbildung zwischen mehreren Parteien (z.B. in Projekte)
- Dienen dem Austausch von Wissen zwischen Anwendungen
- Interoperabilität verbessern

# Bestandteile einer Ontologie

- Klassen (*concepts, classes*)
  - ▶ Mengen der Dinge mit gleichen Eigenschaften
  - ▶ Auch als Begriffe bezeichnet
  - ▶ Beispiel: Die Menge der Planeten
- Relationen (*relations, properties*)
  - ▶ Dienen der Beschreibung von Eigenschaften
  - ▶ Beziehungen zwischen Klassen und deren Instanzen
  - ▶ Beispiel: Die Eigenschaft Radius, Beziehung zwischen Planet und Wert
- Instanzen (*instances, individuals*)
  - ▶ Elemente einer oder mehrerer Klassen
  - ▶ Auch als Objekte bezeichnet
  - ▶ Beispiel: Der Planet Erde als Instanz der Klasse der Planeten

# Aufbau einer Ontologie

- Schema (*terminological box*): Vokabular, Klassen und Relationen
- Inhalt (*assertional box*): Instanzen/Individuum, Aussagen, Daten
- Dabei verwenden Inhalte das Schema
- Beispiel

```
ex:Planet rdf:type rdfs:Class  
ex:earth rdf:type ex:Planet
```

- Das erste Tripel ist Schema
- Das zweite Tripel ist Inhalt
- `ex:Planet` ist eine Klasse
- `ex:earth` ist eine Instanz (Individuum)



# Aufbau einer Ontologie: Schema

- Das Schema besteht aus Axiome
- Diese gelten als wahr
- Dienen zur Repräsentation von abstraktem Wissen
- Beispiele

```
ex:TerrestrialPlanet rdfs:subClassOf ex:Planet  
ex:radius rdfs:domain ex:Planet
```

# Aufbau einer Ontologie: Inhalte

- Die Inhalte bestehen aus Aussagen
- Dienen zur Repräsentation von konkretem Wissen
- Beispiele

```
ex:earth ex:radius "6371"  
ex:earth rdf:type ex:Planet
```

# Ontologien mit RDFS

- RDFS ist eine Sprache mittels der man Ontologien erstellen kann
- RDFS Dokument ist eine Spezifikation von Wissen über eine Domäne
- Wobei die Spezifikation maschinenverarbeitbar ist
- Die wichtigsten Bestandteile der Sprache haben wir bereits gesehen
- Insb. `rdfs:Class`, `rdfs:subClassOf`, `rdfs:subPropertyOf`

# Ontologien Entwickeln

- Ontologien entwickeln (*ontology engineering*) ist eine eigene Disziplin
- Über einfache Beispiele hinaus, ist die Aufgabe generell komplex
- Schwierigkeit nicht primär im Erlernen von Technologie (z.B. RDFS)
- Schwieriger ist Wissen aus Köpfen in Dokumente zu formalisieren

# Ontologien Entwickeln

- Der zweck der Ontologie spielt eine wichtige Rolle
- Oft bestimmt der Zweck welche Aspekte man formalisiert
- Entwicklung von Ontologien also strukturiert angehen
- Insbesondere auch Anforderungsanalyse durchführen
- Ähnlich wie bei der Entwicklung von Software

# Ontologien Entwickeln

- Zuerst schauen ob es bereits eine entsprechende Ontologie gibt
- Die wiederverwendet werden kann, auch nur Teilweise
- Eventuell gibt es eine übergeordnete Ontologie
- Welche abstrakteres Wissen formalisiert
- Auf einer solchen Aufbauen ist oft hilfreich

# Anforderungsanalyse

- Benötigt die Anwendung eine Ontologie, semantische Repräsentation?
- Oder genügt eine klassische Datenbank (u.U. bereits vorhanden)?
- Ist die Toolunterstützung adäquat für mein Projekt?
- Wie sind die Tools lizenziert und wie ausgereift sind sie?

# Anforderungsanalyse

- Welche Domäne wird modelliert?
- Welche Aspekte müssen erfasst werden?
- Wie detailliert soll Wissen beschrieben werden?
- Welche Tätigkeiten soll die Ontologie unterstützen?



# Ontologie Erstellen

- Übersetzung von Wissen in eine maschinenlesbare Form
- Es gibt mehrere Arten von Quellen für Wissen
- Z.B. menschliche Köpfe, Bücher, Web, Datenbanken
- Einige Quellen eignen sich besser als andere für Formalisierung

# Wissensquelle: Mensch

- Experten halten viel Wissen in Köpfen
- Die meisten können Wissen aber nicht formalisieren
- Weil sie sich mit RDFS nicht auskennen
- Zusammenarbeit mit “Wissensingenieure” nötig
- Diese führen Interviews durch und erfassen Expertenwissen
- Experten können nicht immer dinge genau definieren
- In der Kommunikation können Missverständnisse entstehen



Search or jump to...

Pull requests Issues Marketplace Explore



EnvironmentOntology / envo

Watch

17

★ Star

24

Fork

13

&lt;&gt; Code

Issues 310

Pull requests 2

Projects 12

Wiki

Insights

## New particle formation event #602

Edit

New issue



markusstocker opened this issue on Apr 1 · 24 comments



markusstocker commented on Apr 1



I had a meeting with the research community in Helsinki, including @mazaidan. It was pointed out that "formation of particles in an atmosphere" really should be called "new particle formation event". The researchers felt very strongly about having *new* and *event* in the concept. It was also pointed out that a process is not observed, rather the event is observed. To my understanding, the event consists of many processes. The point is, the correct naming of these terms isn't settled and I like to open this issue to try getting this straight, and I like to directly involve the research community (also Pauli et al. but not sure if they are on GitHub, @mazaidan can you check?).



cmungall commented on Apr 1

Member



we should name the classes using terms used by the research community.

So far ENVO hasn't really recognized any distinction between events and processes at the ontological level. A process is simply something that happens. Connections between smaller units and larger units would be achieved using relations like part-of. But we'd be open to other ways of modeling, especially if this fits better with other ways of modeling. We also try to represent the underlying process rather than it's observation.

A crucial point for the ontology is the level of granularity we're talking about. Does a single instance of particle formation generate a single particle? Or are we talking about assemblages of these? We have somewhat fudged this in analogous biological process ontologies, with the formal definition of metabolic or signaling processes being in terms of individual molecules, but with the classes used to represent statistical aggregates of these.



markusstocker commented on Apr 2



An instance of a new particle formation event generates a polydisperse aerosol, so multiple particles of different diameter sizes. During the event (which lasts several hours), the aerosol changes, overall the particles get bigger in size. What scientists describe are events, e.g. the duration, particle growth rate. I fully agree we need to model this how the research community conceptualizes. I hope to get some involved here.

Assignees

pbuttigieg

Labels

enhancement

in editor's version

interesting-discussion

Projects

cloudAtlas (awaiting triage)

meteo (awaiting triage)

Milestone

No milestone

Notifications

Unsubscribe

You're receiving notifications because you were mentioned.

5 participants



## Wissensquelle: Bücher (Unstrukturiert)

- Text ist einfach für Menschen zu verarbeiten
- Formales Wissen aus Text (automatisiert) zu extrahieren ist schwierig
- Natürliche Sprache ist oft mehrdeutig, kontextabhängig, usw.
- Der Prozess leidet meist unter Informationsverlust
- Obwohl hier viel geforscht wurde, sind Resultate eher bescheiden

## Wissensquelle: Web (Semistrukturiert)

- Gewisse Inhalte kommen Strukturiert daher
- Beispiel: Wikipedia Artikel enthalten auch strukturierte Information
- Solche Inhalte kann man einfacher formalisieren
- Struktur gibt es auch zwischen Inhalten
- Beispiel: Explizite Verlinkung zwischen Seiten
- Weist auf eine Verbundenheit hin, wobei die Art oft unklar bleibt
- Metadaten sind oft Strukturiert
- Beispiel: Exif in Bilder
- Bildinhalt unbekannt aber Zugang auf Dateiname/-grösse/-typ, Zeit

## Wissensquelle: Datenbanken (Strukturiert)

- Datenbankinhalte können meist nach RDF übersetzt werden
- Damit erhält man auf einfache Weise semantische Aussagen
- Schemainformation kann man zusätzlich verwenden
- Um terminologisches Wissen aufzubauen
- Andere Ontologien sind auch Quellen für bereits formalisiertes Wissen

# Ontologien Bearbeiten: Protégé

- Protégé ist ein Programm mit dem man Ontologien editieren kann
- Frei erhältlich unter <https://protege.stanford.edu>
- Als Lokalinstallation oder im Web verfügbar
- Web Version unter <https://webprotege.stanford.edu/>
- Wir benutzen Protégé um eine kleine Ontologie zu entwickeln

## Class Hierarchy

Create Delete Watch Search

▼ owl:Thing

● Solar System Planet

Create

Delete

Tags...

Merge Into...

Prune branch to root

Prune all branches to root

Clear pruning

Show IRI

Show Direct Link

Refresh Tree

## Class: Solar System Planet

## IRI

## Annotations

rdfs:label



Solar System Planet

en



Enter property

Enter value

lang

## Classes

● owl:Thing



Enter a class name


## Relationships

Enter property

Enter value

lang



Class Hierarchy ✕Create Delete Watch Search 

- owl:Thing
  - Planet Atmosphere
  - Solar System Planet
    - Gas Giant
    - Terrestrial Planet**
    - Solar System Satellite

Individuals by Class: Terrestrial Planet ✕

Create Delete

Enter search string to filter list

- Earth**
- Mars

Individual: Earth ✕

## IRI

## Types

☒ Terrestrial Planet ▶ ✕

## Relationships

☒ rdfs:label▶  Earthen ✕

lang

## Same As

## Property Hierarchy

Create Delete Watch Search

[Object Properties](#) Data Properties Annotation Properties

owl:topObjectProperty

has satellite

## Property: has satellite

## IRI

## Annotations

rdfs:label

has satellite

en

## Domain

Solar System Planet

## Range

Solar System Satellite

## Property Hierarchy

Create Delete Watch Search

Object Properties Data Properties Annotation Properties

▼ owl:topDataProperty

▼ has physical characteristic

has mass

has equatorial radius

## Property: has physical characteristic

## IRI

## Annotations

rdfs:label

has physical characteristic

en

**Functional** (Single valued)

## Domain



Solar System Planet

## Range



xsd:decimal

Create New Project

Project name

Owner

Last opened

Last modified

bim-108-01-lecture-12



Markus Stocker

Less than one minute ago

6 minutes ago



Open

Open in new window

Download

Move to trash

☒ Owned by Me

☒ Shared with Me

☐ Trash

Sort by Last Opened ▼

# Zusammenfassung

- Mit Ontologien kann man Wissen formalisieren
- Mit RDFS können einfache Ontologien erstellt werden
- Die Sprache hat eine relativ kleine Expressivität
- Man kann nur “einfaches Wissen” ausdrücken
- Es gibt noch weitere ausdruckstärkere Sprachen (OWL)
- Die Entwicklung von Ontologien ist generell komplex
- Die Aufgabe muss somit strukturiert angegangen werden