TECHNISCHE UNIVERSITÄT DRESDEN

SEMANTIC COMPUTING

Lecture 12: Ontology Learning: Introduction

Dagmar Gromann

International Center For Computational Logic

TU Dresden, 19 December 2018

Overview

- Defining ontology
- Introduction to ontology learning

What is an ontology?

What is an ontology?

- **O**ntology (no plural = uncountable): philosophical study of being (existence, reality, categories of beings, etc.)
- **o**ntology (ontologies = countable): “formal, explicit specification of shared conceptualizations.’ (Studer et al. 1998: 186)
- computational artifact designed with a purpose in mind
- represented in a formal language to allow for the processing, reusing, and sharing of knowledge among humans and machines

Studer, Rudi, Benjamins, Richard V., and Fensel, Dieter (1998), ’Knowledge Engineering: Principles and Methods’, Data & Knowledge Engineering, 25 (1-2), 161-198.

Specification? Formal?

A formal, explicit specification of shared conceptualizations; ideally an ontology:

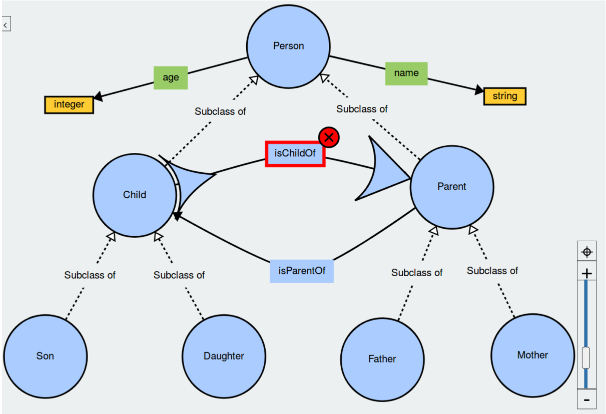
- is a model of (some aspect of) the world
- captures a shared understanding of the domain of interest, a shared conceptualization
- defines a vocabulary relevant to the domain and interpreted the same way by different users
- specifies the meaning of the vocabulary in an explicit manner and often in a formal specification language

Two main parts:

- structure of the model = set of **axioms**
- particular objects and situations = set of **instances**

Axioms

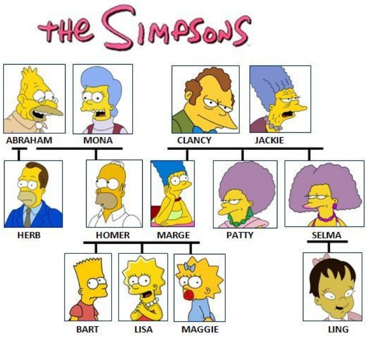
Describe the structure of a model, such as:



Source: http://editor.visualdataweb.org/

Instances

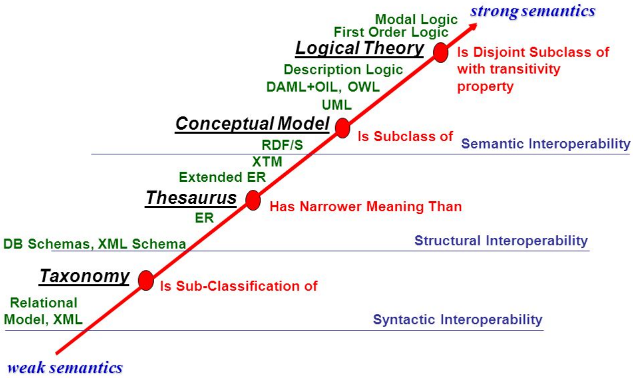
Describe some particular situation in the world, also called individuals, such as:



Source: `https://www.pinterest.com/pin/374784000224934195/`

Ontology Spectrum

From weak semantics (less expressive) to strong semantics (very expressive):



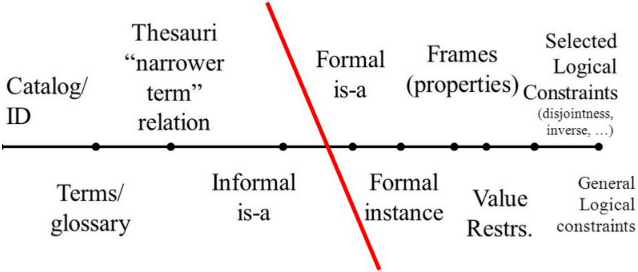
Specification

From weak to strong semantics:

- List of terms
- Taxonomies (hierarchical relations)
- Database schemas
- Entity Relationship (ER) models
- Thesauri, informal hierarchies
- Resource Description Framework (RDF)
- Unified Modeling Language (UML)
- Logic Programming
- Web Ontology Language (OWL), Description Logics
- First-order logic, Modal logics

Formal specification

“Formal” refers to the language in which classes, properties, and relations are expressed and express that it is possible to conclude implications that follow from represented facts (inference).



Example

Ontology structure (Tbox):

Woman ≡ Yellow ⊓ Female

Mother ≡ Woman ⊓ has_child Male ⊓...

Ontology data (Abox):

Marge: Mother

We now know that Marge is a Woman, which means she is Yellow and Female (inference) and we can use that to gather more complex information with specific queries.

An ontology represents a set of possible worlds (models). We do not know which of these models describes the real world, which makes our knoweldge incomplete.

Applications and Uses

Applications:

- Semantic Web and Databases
- Artificial Intelligence and Knowledge Representation
- Specific domains: esp. Biomedicine, Medicine
- Libraries and Cataloging
- Geospatial mapping
- many more (also industrial applications)...

Uses:

- data annotation (web pages, databases, etc.)
- querying large knowledge bases (emphasis on plural - can be used to unite knowledge from different knowledge bases)
- formal properties allow for automated inference mechanisms
- many more...

Ontology Learning

Ontology Learning

Ontology learning

refers to a set of methods and techniques for:

- the development of new ontologies
- extension or adaptation of already existing ontologies

Why ontology learning?

- ontologies allow sharing and reusing of knowledge among people and computer systems
- ontology engineering is expensive wrt to time and resources
- (partial) automation to support/automate ontology engineering process
- can also be referred to as ontology generation, ontology extraction or ontology mining

Terminological or Assertional Knowledge?

Bees are insects that produce honey. They have six legs. Bees live only in beehives - or just hives. Maya and Flip are bees. Maya, in particular, is a notable bee. Maya and Flip are friends.

Ontology learning

Terminological Knowledge about concepts, their definitions, and relations among them. Examples are :

- bees are insects;
- bees produce honey;
- bees have six legs;
- bees live in beehives.

Terminological Knowledge is independent of a particular state of affairs.

Ontology population

Assertional Knowledge about individuals, their mutual relations and relations with concepts. Examples are:

- Maya is a bee;
- Flip is a bee;
- Maya and Flip are friends.

Assertional Knowledge reflects a particular state of affairs.

Source: Rospocher, M. (2018). Learning Expressive Ontological Concept Descriptions via Neural Networks. SemDeep-4.

Types of Ontology Learning

- Ontology learning from natural language text
  - (semi-) automated generation of (lightweight) ontologies by using text processing methods
- Linked Data mining
  - detecting patterns in Resource Description graphs via statistical schema induction or statistical relational learning
- Concept learning in Description Logics and OWL
  - learning schema axioms from existing ontologies and instances
- Crowdsourcing ontologies
  - use a crowdsourcing platform to let humans evaluate or generate ontological relations

Ontology Learning from Text

Ontology Learning from Text

refers to a set of methods that (semi-)automatically identify terms, concepts, relations, and potentially also axioms from textual information in order to generate, improve, or extend an ontology with them.

Methods that can be used towards this end:

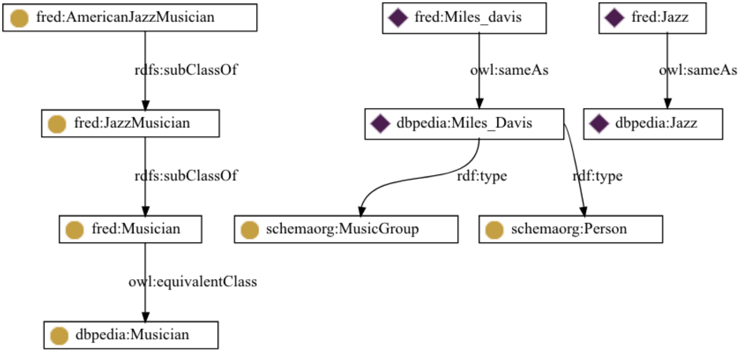
- NLP pipeline
- Information extraction
- Data and text mining
- Machine learning and deep learning

Buitelaar, P., Cimiano, P., and Magnini, B. (2005). Ontology learning from text: An overview. Ontology learning from text: Methods, evaluation and applications, 123, 3-12.

Example: Fred

Input: “Miles Davis was an American jazz musician.”

Output:



Source: `http://wit.istc.cnr.it/stlab-tools/fred/demo/`

NLP for Ontology Learning

- Tokenization
- Morphological analysis (stemming, lemmatization)
- Part-of-Speech (POS) tagging
- Syntactic Parsing
- Relation Extraction
- Semantic Parsing
- ...

Machine Learning for Ontology Learning

- Association Rule Mining
- (Hierarchical) Clustering
- Classification, e.g. of new concepts into existing concept hierarchy
- Conceptual Clustering, e.g. Formal Concept Analysis
- Neural network-based approaches

Ontology Learning Layer Cake

Individual tasks of ontology learning have been organized into the Ontology Learning Layer Cake:

Image Source: Petrucci, G. (2015). Information extraction for learning expressive ontologies. In European Semantic Web Conference, Springer, Cham, 740-750

Term Extraction

- **Goal**: Identify all relevant terms in a given text corpus
- **Term**: word or expression that has a precise meaning in some uses or domain of interest (“the” vs. “disease”)
- **Common Methods**:
  - Automated Term Extraction using frequency based metrics from information extraction such as Term Frequency- Inverse Document Frequency (TF-IDF)
  - using linguistic features, e.g. POS tags
  - clustering to separate terms from other types of words
  - etc.

Synonym Extraction

- **Goal**: Group words together with a similar meaning
- **Common Methods**:
  - Text mining
  - WordNet-based algorithms
  - Distributional and distributed vector representations, e.g. word embeddings
  - etc.

Concept Learning

- **Goal**: induce concepts in an intensional way (intension vs. extension)
- **Methods**:
  - Clustering combined with Latent Semantic Analysis (LSA)
  - Creating concepts out of synonym sets (e.g. WordNet synsets)
  - Use existing concepts to associate newly extracted terms
  - etc.

Concept Hierarchy Learning

- **Goal**: learn taxonomic relations over concepts
- **Methods**:
  - Lexico-syntactic patterns (e.g. “A dog is a mammal” => Dog $\sqsubseteq$ Mammal)
  - Syntactic features and a binary classifier to predict which nouns are in a taxonomic relation
  - Hierarchical clustering
  - etc.

Relation Extraction

- **Goal**: learn a set of all relations among concepts and individuals (instances)
- **Methods**:
  - Linguistic patterns combined with verb-based relation detection or graph theory
  - Dependency parsing and WordNet information to find mereological relations (part-of)
  - RNN-based Encoder-Decoder System with Attention
  - etc.

Axiom Learning

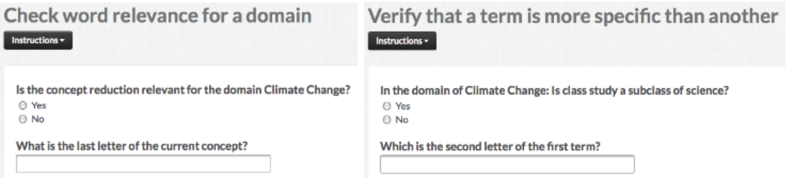
- **Goal**: learning the whole structure of an ontology
- **Methods**:
  - Lexico-syntactic patterns applied to natural language definitions
  - Lexical resources to learn disjointness (an individual cannot be an instance of two classes that are disjoint)
  - Discourse Representation Theory (DRT)
  - Ontology Design Patterns (kind of templates to reuse in the engineering process that can be used for automation)
  - Recently: deep learning approaches
  - etc.

Challenges

- **Heterogeneity**: data on the web or in a corpus frequently differ or even contradict each other
- **Uncertainty**: low-quality data and/or imperfect methods for learning
- **Scalability**: methods should be able to handle huge corpora
- **Dynamics**: methods should be able to continue learning and updating as knowledge evolves/changes
- **Reasoning**: knowledge in ontologies needs to be formal and consistent to enable reasoning

Improvement with Human Involvement

One popular way of improving automated ontology learning has been by involving humans in form of crowdsourcing: a large number of non-experts is asked to complete straightforward questions that verify or falsify learned contents.



Source: Hanika, F., Wohlgenannt, G. and Sabou, M. (2014). The ucomp protégé plugin: Crowdsourcing enabled ontology engineering. In International Conference on Knowledge Engineering and Knowledge Management. Springer, Cham, pp. 181-196

Review of Lecture 12

- What is an ontology?
- What is the difference between axiom and instance?
- What do we mean by formal ontology?
- Where can an ontology be applied or used?
- What is ontology learning? What is ontology population?
- Which types of ontology learning can be differentiated?
- Which tasks do we need to consider in ontology learning?
- Which methods do we have to learn axioms from text?
- What is particularly challenging in ontology learning from text?
- How can we crowdsource ontology learning?