Einführung in die Computerlinguistik

Semantik 2: Ontologie

Dienstag, 8. Januar 2008

Hans-Ulrich Krieger

Language Technology Lab
German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
krieger@dfki.de

Overview

- what is an ontology?
- thesauri
- WordNet
- FrameNet
- SUMO/MILO
- description logics
- focus: RDF & OWL

What is an Ontology?

Ontology [Greek]: most fundamental branch of general metaphysics, dealing with the study of existence (science of being)

term coined by Aristotle (384BC–322BC)

discipline can be subdivided into

- formal ontology (or universal science)
- material ontology

Formal Ontology

question: what are the truth-determining foundations of general metaphysics, i.e., what are the most general rules directing our decisions, leading to more specialized rules (e.g., in medicine): *first principles*

- Law of Identity
  $A = A$: an axiom in most logics

- Law of Excluded Middle
  either P or $\neg P$

- Law of Non-Contradiction
  proof by contradiction: $(\neg P \Rightarrow (R \wedge \neg R)) \Rightarrow P$

Material Ontology

what are the fundamental categories of being? (Aristotle)
more general view: find out what entities and what types of entities exist!

similar to the idea of first principles: start with *Being* (does not need any definition), and add further subcategories, such as *Substance*

what does it mean for an entity to be member of a certain category?
sharing prototypical values for category-specific properties!

Reappearance of the Wheel

Aristotle's theory of categories and classification "reappears" in many scientific disciplines:

- biology
- ...
- CL, AI, CS, LT, ...
  - (computational) linguistics
  - artificial intelligence
  - computer science
  - information science, lexicography, semantic web, ...

What is an Ontology, Again

being more liberate (Tom Gruber, 1993):

> A conceptualization is an *abstract*, simplified view of the world that we wish to represent for some purpose. ... An **ontology** is an *explicit specification* of a conceptualization. ... When the knowledge of a domain is represented in a *declarative formalism*, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the *representational vacabulary* with which a knowledge-based program represents knowledge.

What is an Ontology, Again, cont.

i.e., an ontology is a description of objects (concepts & individuals) and relationships between objects

1+is-a relation: **taxonomy**; 1+2: **thesaurus**

1. categories/concepts/classes/types
   Human, Object
2. relations/roles/properties/attributes/slots
   owns $\subseteq$ Human $\times$ Object
3. individuals/instances/entities
   peter, bmw
4. instantiated relations/ground atoms/relation instances
   owns(peter, bmw)

what is missing here? **semantics!** (not covered here!)

Why are we interested in Ontologies?

- pure epistemological aspects—no practical interest in running systems

- very practical aspects
  - enabling knowledge resources for the Semantic Web
  - help applications to perform better by taking content/meaning into account
  - precise source of information for humans & computer programs

Application Areas

- query expansion in information retrieval & QA
- DB access & ontology retrieval
- word sense disambiguation
- language-specific inferences on lexical semantic representation
- ontology population through information extraction
- general inferences dealing with world knowledge

Examples

- thesauri
- WordNet
- FrameNet
- SUMO/MILO
- description logics & OWL

Merriam-Webster Online Thesaurus

Word: human

Function: adjective

Text: relating to or characteristic of human beings (it's human nature to care about what people think of us)

Synonyms: mortal, natural

Related Words: anthropoid, hominid, humanlike, humanoid

Near Antonyms: angelic (or angelical), divine, godlike, superhuman, supernatural; immortal, omnipotent, omniscient; animal, beastly, bestial, brute; inhuman, robotic

Antonyms: nonhuman

Merriam-Webster Online Thesaurus, cont.

Word: human

Function: noun

Text: a member of the human race (humans are the only mammals not endowed with a natural defense against the elements, such as fur or a thick hide)

Synonyms: being, bird, body, creature, customer, devil, guy, head, individual, life, man, mortal, party, person, scout, sort, soul, specimen, thing, wight

Related Words: hominid, homo, humanoid; brother, fellow, fellowman, neighbor; celebrity, personage, personality, self, somebody

Near Antonyms: animal, beast, brute

WordNet—Hypernyms of Human

WN *hierarchically* organizes nouns, verbs, adjectives, and adverbs into synonym sets which refer to lexical concepts (155,287 unique strings & 117,659 synsets in WordNet 3.0)

Sense 1/noun: *a human being*

```
person, individual, someone, somebody, mortal, human, soul
       => organism, being
           => living thing, animate thing
               => object, physical object
                   => entity
       => causal agent, cause, causal agency
           => entity
```

WordNet—Hypernyms of Human, cont.

Sense 2/noun: *any living or extinct member of the family Hominidae*

```
homo, man, human being, human
       => hominid
           => primate
               => placental, placental mammal, eutherian, eutherian mammal
                   => mammal
                       => vertebrate, craniate
                           => chordate
                               => animal, animate being, beast, brute, ....
                                   => organism, being
                                       => living thing, animate thing
                                           => object, physical object
                                               => entity
```

Relations we are interested in w.r.t. Concept C

- **synonyms** concepts having the same meaning as C
- **antonyms** concepts that do not share any properties with C
- **hypernyms** concepts that are more general than C
- **hyponyms** concepts that are more specific than C
- [**holonyms** concepts that contain C as a part]
- [**meronyms** concepts that are part of C]

FrameNet—Human, Again

FN lists semantic and syntactic combinatory possibilities (valences) of each word in each of its senses ($>$ 10,000 lexical units; $\approx$ 800 hierarchical semantic frames)

two lexical units for *human*: `human_being.n` and `human.n`

but semantic frame is `People`

several "subclasses" of `People`, e.g., `People_by_age`

binary relations, connecting frames: `Inherits_From`, `Uses`, ...

example: `People_by_age Inherits_From People` ("specialization")

`People_by_age Uses Age` ("properties")

SUMO & MILO

Suggested Upper Merged Ontology: very basic concepts & axioms (similar upper ontologies: DOLCE, PROTON)

higher-order LISPish specification language SUO-KIF

```
(instance instance BinaryPredicate)
(subrelation immediateInstance instance)
(instance immediateInstance AsymmetricRelation)

(=> (immediateInstance ?ENTITY ?CLASS)
    (not (exists (?SUBCLASS)
           (and (subclass ?SUBCLASS ?CLASS)
                (not (equal ?SUBCLASS ?CLASS))
                (instance ?ENTITY ?SUBCLASS)))))
```

SUMO & MILO, cont.

MId-Level Ontology: bridges between the abstract content of SUMO and various domain ontologies

all ontologies together: 20,000 terms and 60,000 axioms

partial inference support via Vampire

```
(subclass HumanSlave Human)

(=> (instance ?SLAVE HumanSlave)
    (exists (?PERSON)
      (and (instance ?PERSON Human)
           (not (equal ?PERSON ?SLAVE))
           (possesses ?PERSON ?SLAVE))))
```

SUMO & MILO—That Human Thing, Again

mappings of concepts to WordNet lexicon
example *human*: found the two senses from WordNet

```
(partition Human Man Woman)
(subclass Human CognitiveAgent)
(subclass Human Hominid)
(subclass Man Human)

(<=> (attribute ?PERSON Unemployed)
     (and (instance ?PERSON Human)
          (forall (?ORG)
            (not (employs ?ORG ?PERSON)))))
```

Description Logics

family of logic-based knowledge representation formalisms

descendants of semantic networks and KL-ONE

model-theoretic semantics (decidable 2-variable FOL fragment)

sound & complete decision procedures

complex expressions through concept-forming constructors

$$\texttt{HumanSlave} \equiv \texttt{Human} \sqcap \exists\, \texttt{possesses}^{-1}.\, (\texttt{Human} \sqcap \neg\, \texttt{Slave})$$

$$\texttt{DoctorFamily} \equiv \texttt{Family} \sqcap \exists\, \texttt{hasMember}.\, (\texttt{Person} \sqcap \forall\, \texttt{hasChild}.\, (\texttt{Doctor} \sqcup \exists\, \texttt{hasChild.Doctor}))$$

Recap: What is an Ontology, Again and Again ...

similarities between examples indicate that

- I take a liberal stance here what an ontology is
- we always construct ontologies when conceptualizing a domain
  1. classes
  2. distinguished sub/super relationship, plus other relations
  3. instances
  4. instantiated relationships
- but: formal ontology languages must address
  - semantics: well-defined (yes)
  - decidability: sound (!), complete (?), terminating (?)
  - tractability: average-case problems (yes .. no)

RDF & OWL

The Semantic Web Vision

(syntactic) Web made possible through established standards:
TCP/IP, HTTP, HTML, ...

1st generation: mostly handwritten HTML pages

2nd generation: very often machine-generated active pages

next generation (we're just here!): resources should be more accessible to automated processes

- to be achieved via semantic markup
- metadata annotations, describing content/function

coincides with Tim Berners-Lee's vision of a Semantic Web

Semantic Web & Ontologies

semantic markup must be meaningful to automated processes

ontologies will play a key role here

- source of precisely defined terms (vocabulary)
- can be shared across applications and humans

increased formality facilitates machine understanding

very important: standards!

long road:
XML, URI, **RDF**, **RDFS**, DAML & OIL, **OWL**, SWRL, ...

RDF: Resource Description Framework

- general-purpose language for representing information
- provides a lightweight ontology system
- enabling technology for the Semantic Web
- XML exchange syntax (but also N3, N-Triples)
- RDF data model: triple
- idea: everything can be represented as a triple

RDF, cont.

- triple: $\langle$ subject, predicate, object $\rangle$
- subject, predicate, object: URIs or XSD literals
  (or again triples: reification)
- URI = Uniform Resource Identifier ($\approx$ virtual Web identifier)
  e.g., http://www.w3.org/2002/07/owl#intersectionOf
- XSD (XML Schema Datatypes) literals are typed
  e.g., "2.4"^^xsd:decimal

RDFS: RDF Schema

- describes how to use RDF to describe RDF vocabularies
- defines other built-in RDF vocabulary (domain, subClassOf)
- RDF/RDFS semantics via axiomatic triples and entailment rules (Hayes 2004, ter Horst 2005), e.g.,
  - $\langle$rdf:type, rdf:type, rdf:Property$\rangle$
    $\langle$rdfs:domain, rdfs:domain, rdf:Property$\rangle$
  - $\langle ?c_1$, owl:equivalentClass, $?c_2\rangle \Rightarrow \langle ?c_2$, owl:equivalentClass, $?c_1\rangle$
    $\langle ?i$, rdf:type, $?c_2\rangle, \langle ?c_2$, rdfs:subClassOf, $?c_1\rangle \Rightarrow \langle ?i$, rdf:type, $?c_1\rangle$

OWL

decidable instance of the description logics family (FOL fragment)

well-founded set-theoretical semantics

outcome of the DAML+OIL W3C standardization

de facto standard today to specify ontologies

RDFS-based syntax and ontological primitives
   e.g., `rdfs:subClassOf`

fine-grained, more complex means as in RDFS
   e.g., `owl:intersectionOf`

uses XML/RDF exchange syntax

ontology is a set of axioms describing classes and properties

OWL Class Constructors

Constructor	DL Syntax	Example
`Thing`, `Nothing`	$\top, \bot$	
`intersectionOf`	$C_1 \sqcap \ldots \sqcap C_n$	Human $\sqcap$ Male
`unionOf`	$C_1 \sqcup \ldots \sqcup C_n$	Doctor $\sqcup$ Lawyer
`complementOf`	$\neg C$	$\neg$Male
`oneOf`	$\{x_1, \ldots, x_n\}$	{john, mary}
`someValuesFrom`	$\exists P \,.\, C$	$\exists$hasChild . Lawyer
`allValuesFrom`	$\forall P \,.\, C$	$\forall$hasChild . Doctor
`maxCardinality`	$\leq n P$	≤ 1 hasChild
`minCardinality`	$\geq n P$	≥ 2 hasChild

OWL Semantics

model theory relates expressions to interpretations

$$\mathcal{I} = \langle \mathcal{U}, \cdot^{\mathcal{I}} \rangle$$

i.e., model theory is the glue between the syntax and semantics of a (first-order) language;
domain of interpretation: $\mathcal{U} = \top^{\mathcal{I}}$, interpretation function: $\cdot^{\mathcal{I}}$

- classes: *subsets* of $\mathcal{U}$
- properties: *subsets* of $\mathcal{U} \times \mathcal{U}$
- instances: *elements* of $\mathcal{U}$
- instantiated properties: *elements* of $\mathcal{U} \times \mathcal{U}$

OWL Semantics, cont.

extend interpretation function $\cdot^{\mathcal{I}}$ to concept expressions

- $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
- $(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\neg C)^{\mathcal{I}} = \mathcal{U} \setminus C^{\mathcal{I}}$
- $(\{x_1, \ldots, x_n\})^{\mathcal{I}} = \{x_1^{\mathcal{I}}, \ldots, x_n^{\mathcal{I}}\}$
- $(\exists P \,.\, C)^{\mathcal{I}} = \{x \mid \exists y \,.\, (x, y) \in P^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
- $(\forall P \,.\, C)^{\mathcal{I}} = \{x \mid \forall y \,.\, (x, y) \in P^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}\}$
- $(\leq n P)^{\mathcal{I}} = \{x \mid \#\{y \mid (x, y) \in P^{\mathcal{I}}\} \leq n\}$
- $(\geq n P)^{\mathcal{I}} = \{x \mid \#\{y \mid (x, y) \in P^{\mathcal{I}}\} \geq n\}$

OWL Axioms

Axiom	DL Syntax	Example
`subClassOf`	$C \sqsubseteq D$	HumanSlave $\sqsubseteq$ Human
`equivalentClass`	$C \equiv D$	Human $\equiv$ Man $\sqcup$ Woman
`disjointWith`	$C \sqsubseteq \neg D$	Man $\sqsubseteq \neg$Woman
`sameAs`	$\{x\} \equiv \{y\}$	$\{$president_bush$\} \equiv \{$g_w_bush$\}$
`differentFrom`	$\{x\} \sqsubseteq \neg\{y\}$	$\{$John$\} \sqsubseteq \neg\{$Peter$\}$
`subPropertyOf`	$P \sqsubseteq Q$	hasDaughter $\sqsubseteq$ hasChild
`equivalentProperty`	$P \equiv Q$	cost $\equiv$ price
`inverseOf`	$P \equiv Q^-$	hasCeo $\equiv$ ceoOf$^-$
`SymmetricProperty`	$P \equiv P^-$	distance $\equiv$ distance$^-$
`TransitiveProperty`	$P^+ \sqsubseteq P$	hasPart$^+ \sqsubseteq$ hasPart
`domain`	$\top \sqsubseteq \forall P^- . C$	$\top \sqsubseteq \forall$`hasCeo`$^-$. `Company`
`range`	$\top \sqsubseteq \forall P . C$	$\top \sqsubseteq \forall$`hasCeo` . `Person`
concept assertion	$x : C$	`alice` : `Woman`
role assertion	$(x, y) : P$	(`alice`, `boris`) : `hasChild`

OWL Axioms: How to Read $\sqsubseteq$ and $\equiv$

- $C \sqsubseteq D$ **iff** $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
- $C \equiv D$ **iff** $C^{\mathcal{I}} = D^{\mathcal{I}}$
- examples
  - $\mathsf{Human} \equiv \mathsf{Man} \sqcup \mathsf{Woman}$ **iff**
    $\mathsf{Human}^{\mathcal{I}} = \mathsf{Man}^{\mathcal{I}} \sqcup \mathsf{Woman}^{\mathcal{I}}$
  - $\{\mathsf{president\_bush}\} \equiv \{\mathsf{g\_w\_bush}\}$ **iff**
    $\{\mathsf{president\_bush}\}^{\mathcal{I}} = \{\mathsf{g\_w\_bush}\}^{\mathcal{I}}$ **iff**
    $\{\mathsf{president\_bush}^{\mathcal{I}}\} = \{\mathsf{g\_w\_bush}^{\mathcal{I}}\}$

Open-World Semantics & Non-Unique Name Assumption

OWL must allow for distributed information (Semantic Web!); information can be added incrementally: monotonicity; i.e., new information can NOT retract old; old can NOT be deleted

open-world assumption
what can NOT proven to be true is NOT believed to be false

example ontology:
$\{$Woman(alice), hasChild(alice, doris), hasChild(alice, boris)$\}$
question: $\{\text{alice}\} \sqsubseteq \leq 2$ hasChild vs. $\{\text{alice}\} \sqsubseteq \geq 2$ hasChild
at most: **don't know** at least: **yeeees, but ...**

non-unique name assumption
individuals sharing different names need not be different/might be equal

Übungsblatt