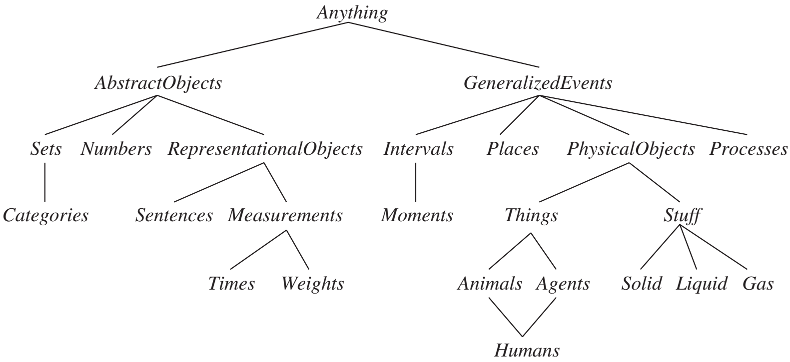
Knowledge Representation

Jihoon Yang

Machine Learning Research Laboratory
Department of Computer Science & Engineering
Sogang University

Ontology ≒ 용어

- Everything in the world is organized into a hierarchy of categories
- Ontology: vocabulary for the domain knowledge
- Ontological engineering: representation of various ontology/abstract concepts (e.g. Events, Time, Physical Objects, Beliefs)



Some KBs or Ontologies

There are some routes of building very larger KBs or ontologies:
– CYC: creating the ontology and writing axioms from cyclopedia (1990)
– DBpedia: importing categories, attributes and values from Wikipedia (2007)
– TextRunner: building by reading a large corpus of Web pages (2008)
– OpenMind: building by volunteers who proposed facts and commonsense knowledge in English (2002)
– Knowledge Graph (KG, previous Freebase): building by Google and holding more that 70 billion facts (2012)
– Wikidata, Linking Open Data (LOD), YAGO etc.

Other data sources (also known Deep Web): MusicBrainz, DrugBank etc.

Production Systems (among earliest FC systems)

Production (rule-based) systems:

– *working memory*: a knowledge base

– *rule memory*: a set of inference rules with form

$p_1 \wedge \cdots \wedge p_n \Rightarrow act_1 \wedge \cdots \wedge act_m$

where p_i are literals, and act_j are actions to take when the p_i are all satisfied – forward chaining

– *match phase*: in each cycle, the system computes the subset of rules whose left-hand side is satisfied by the current contents of the working memory

– *conflict resolution phase*: the system decides which of the rules should be executed

– *act phase*: in each cycle, the system executes the action(s) in the chosen rule(s)

Production Systems

Inefficient forward chaining unification match algorithm:

E.g., If there are $w = 100$ elements in working memory and $r = 200$ rules each with $n = 5$ elements in the left-hand side, and solving a problem requires $c = 1000$ cycles, then the naive match algorithm must perform $wrnc = 10^8$ unifications

Rete algorithm of OPS5:

E.g., rule memory

$$A(x) \wedge B(x) \wedge C(y) \Rightarrow addD(x)$$
$$A(x) \wedge B(x) \wedge D(y) \Rightarrow addE(x)$$
$$A(x) \wedge B(x) \wedge E(y) \Rightarrow addDeleteA(x)$$

and working memory

$$\{A(1), A(2), B(2), B(3), B(4), C(5)\}$$

Expert Systems

So-called expert systems are production systems

MYCIN (Stanford): aided physicians in treating bacterial infections
– approximately 500 rules for recognizing about 100 causes of infection

E.g., IF

the type of x is primary bacteremia

the suspected entry point of x is the gastrointestinal tract

the site of the culture of x is one of the sterile sites

$THEN$

there is evidence that x is bacteroides

– certainty factors: numbers from $[0, 1]$ attached to conclusions to rank order

Recently, IBM Watson Health systems
(Watson for Oncology, Watson Genomics, Clinical Trial Matching)

Categories and Objects (in FOL)

- Organization of **objects** into **categories** is a vital part of KR
- Category: can be represented by
  - Predicates: $Basketball(b)$
  - Objects: *Basketballs*
- Subcategory (or subclass or subset): organized into *taxonomic hierarchy (taxonomy)*; $Subset(Basketballs, Balls)$
- Categories organize knowledge through inheritance
- FOL states facts about categories, either by relating objects to categories or by quantifying over their members
  - An object is a member of a category.
    $BB_9 \in Basketballs$
  - A category is a subclass of another category.
    $Basketballs \subset Balls$
  - All members of a category have some properties.
    $(x \in Basketballs) \Rightarrow Spherical(x)$
- Physical composition (e.g. *PartOf, BunchOf*)
- Measurements (e.g. *Length, Inches, Centimeters, Pounds, \$, Hours*)
- *Stuff* and *Things* are the most general substance and object categories, respectively

Events

- Time
  - *Begin, End, Time, Date, Seconds*
  - *Moment (Duration)*
  - *Interval (Meet, Before, After, During, Overlap, Starts, Finishes, Equals)*

- Fluents and objects
  - Fluent: aspects of the world that change
  - Physical object is a chunk of space-time, and can be viewed as a generalized event

$T(Equals(President(USA), G.Washington), Begin(AD1790), End(AD179$

- Event calculus
  - An approach to describing sequential occurrence of events (or actions)
  - Objects of event calculus: events, fluents, time points

$E_1 \in Flyings \wedge Flyer(E_1, Shankar) \wedge Origin(E_1, SF) \wedge Destination(E_1, DC)$

Semantic Networks

- Capable of representing individual objects, categories of objects, and relations among objects
- Ovals represent object or category names, which are connected with labeled links
- Single/double-boxed link is used to assert properties of every/some member of a category

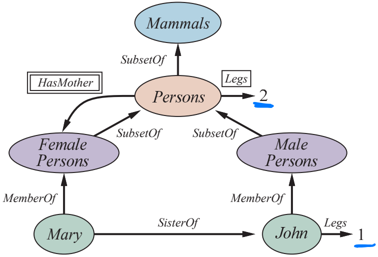


Figure 10.4 A semantic network with four objects (John, Mary, 1, and 2) and four categories. Relations are denoted by labeled links.

Inheritance Reasoning

- Semantic network notation makes it convenient to perform inheritance reasoning
- Inheritance is the result of transitivity reasoning over paths in a network
- E.g. Mary inherits the property of having two legs:
  Q: How many legs Mary has?

$$Mary \rightarrow (MemberOf) \rightarrow FemalePersons \rightarrow$$

$$(SubsetOf) \rightarrow Persons \rightarrow \textbf{Legs} \rightarrow 2$$

- Multiple inheritance
  - An object can belong to more than one category, or a category can be a subset of more than one other category
  - Can find two or more conflicting values answering the query

Description Logics

Description logics (DLs): a family of logics with notations designed to describe the definitions and properties of categories

Subsumption: checking if one category is a subset of another based on their definitions

Classification: checking if an object belongs to a category

Description logics focus on tractability of inference and serve as theoretic foundation for ontology

Description Logics

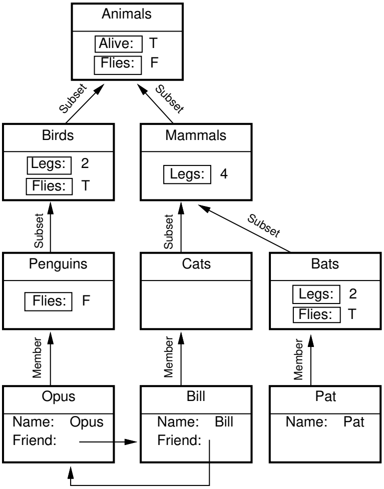
$$
\begin{array}{rcl}
\mathit{Concept} & \rightarrow & \textbf{Thing} \mid \mathit{ConceptName} \\
& \mid & \textbf{And}(\mathit{Concept}, \ldots) \\
& \mid & \textbf{All}(\mathit{RoleName}, \mathit{Concept}) \\
& \mid & \textbf{AtLeast}(\mathit{Integer}, \mathit{RoleName}) \\
& \mid & \textbf{AtMost}(\mathit{Integer}, \mathit{RoleName}) \\
& \mid & \textbf{Fills}(\mathit{RoleName}, \mathit{IndividualName}, \ldots) \\
& \mid & \textbf{SameAs}(\mathit{Path}, \mathit{Path}) \\
& \mid & \textbf{OneOf}(\mathit{IndividualName}, \ldots) \\
\mathit{Path} & \rightarrow & [\mathit{RoleName}, \ldots] \\
\mathit{ConceptName} & \rightarrow & \mathit{Adult} \mid \mathit{Female} \mid \mathit{Male} \mid \ldots \\
\mathit{RoleName} & \rightarrow & \mathit{Spouse} \mid \mathit{Daughter} \mid \mathit{Son} \mid \ldots
\end{array}
$$

Figure 10.6 The syntax of descriptions in a subset of the CLASSIC language.

Example: Man with at least three sons who are all unemployed and married to doctors, and at most two daughters who are all professors in physics or math departments:

$$
\begin{array}{l}
\mathit{And}(\mathit{Man}, \mathit{AtLeast}(3, \mathit{Son}), \mathit{AtMost}(2, \mathit{Daughter}), \\
\quad \mathit{All}(\mathit{Son}, \mathit{And}(\mathit{Unemployed}, \mathit{Married}, \mathit{All}(\mathit{Spouse}, \mathit{Doctor}))), \\
\quad \mathit{All}(\mathit{Daughter}, \mathit{And}(\mathit{Professor}, \mathit{Fills}(\mathit{Department}, \mathit{Physics}, \mathit{Math}))))
\end{array}
$$

Frame



(a) A frame-based knowledge base

Rel(Alive,Animals,T)
Rel(Flies,Animals,F)

Birds $\subset$ Animals
Mammals $\subset$ Animals

Rel(Flies,Birds,T)
Rel(Legs,Birds,2)
Rel(Legs,Mammals,4)

Penguins $\subset$ Birds
Cats $\subset$ Mammals
Bats $\subset$ Mammals

Rel(Flies,Penguins,F)
Rel(Legs,Bats,2)
Rel(Flies,Bats,T)

Opus $\in$ Penguins
Bill $\in$ Cats
Pat $\in$ Bats

Name(Opus,"Opus")
Name(Bill,"Bill")
Friend(Opus,Bill)
Friend(Bill,Opus)
Name(Pat,"Pat")

(b) Translation into first–order logic

Semantic Web

Semantic Web is the next generation of Web led by W3C (World Wide Web Consortium, `http://www.w3c.org`) that makes the Web pages understandable to machine

- RDF (Resource Description Framework) as underlying meta-data representational language is a language of categories (ontology)
- The "semantics" of Web (data and facts) is realized by ontology, OWL (Web Ontology Language) is an ontology representation language
- Various Knowledge Graphs are implementations of KB
- Description logics are theoretic foundation of ontology and the standards of ontology language
- The spirit of semantic Web came from AI and can be viewed as an application of AI (so-called Internet + AI)

Resource Description Framework

RDF: W3C specifications as a metadata data model
- based on description logics
- syntax in XML (eXtensible Markup Language)

Triples: $subject - predicate - object$, or
(h,r,t) $(head - relation - tail)$, entity-relationship (ER model)
- $subject, object$ (entities): the (web) resources
- $predicate$: relationship between the subject and the object

RDF graph: a collection of RDF triples represents a labeled, directed multi-graph

SPARQL: query language for RDF graphs
- an SQL-like language

Web Ontology Language

OWL: a family of ontology representation languages
- based on description logics
- RDF/OWL syntax in XML (eXtensible Markup Language)

OWL extends RDF Schema
- Class equivalent
  Property
  sameIndividualAs
  ...
- RDFS
  subClassOf
  resource
  ID
  ...

Knowledge Graph

Knowledge graph (KG, or linked data, multi-relational data): representing entities and relations

Triples (h,r,t): $head - relation - tail$

– based on RDF and description logics

E.g. KG services to enhance search engine's results with information gathered from a variety of sources

– Google, knowledge panel

Summary

- Large-scale knowledge representation requires a general-purpose ontology to organize and tie together various specific domains of knowledge
- An ontology is based on categories and event calculus
- Categories, subcategories, parts, structured objects, measurements, substances, events, time and space are covered
- Actions, events, and time can be represented with the event calculus; Such representations enable an agent to construct sequence of actions and make logical inferences about what will be true when these actions happen
- Special-purpose representation systems, such as semantic networks and description logics, have been devised to help in organizing a hierarchy of categoreis; Inheritance is an important form of inference, allowing the properties of objects to be deducted from their membership in categories