Ontology Tran Duc Minh

Outline

- Ontology
- Description Logic
- Sematics
- Knowledge Base
- Reasoning
- Semantic Web
- Semantic Web Rule Language
- Inductive Logic Programming

Ontology

- An ontology is a formal conceptualization of the world.
- An ontology specifies a set of constraints, which declare what should necessarily hold in any possible world.
- Any possible world should conform to the constraints expressed by the ontology.
- Given an ontology, a legal world description is a possible world satisfying the constraints.

Ontology languages (1)

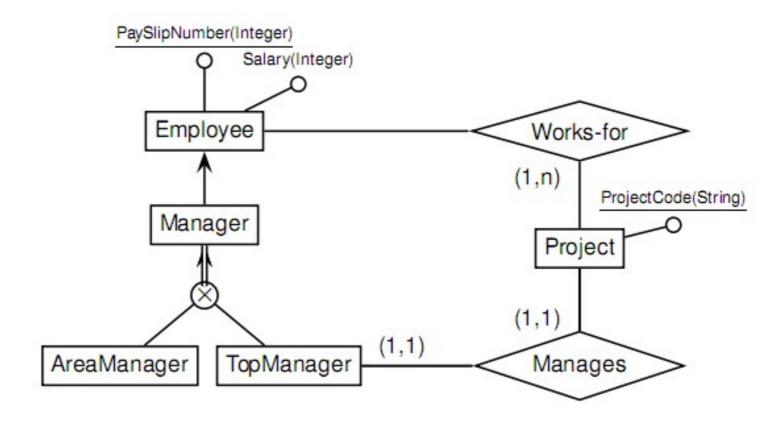
- An ontology language usually introduces concepts (aka classes, entities), properties of concepts (aka slots, attributes, roles), relationships between concepts (aka associations), and additional constraints.
- Ontology languages may be simple (e.g., having only concepts), frame-based (having only concepts and properties), or logicbased (e.g. Ontolingua and DAML+OIL).
 - Ontolingua is mechanism for writing ontologies in a canonical format, such that they can be easily translated into a variety of representation and reasoning systems.
 - DAML (DARPA Agent Markup Language)
 - OIL (the Ontology Inference Layer)

Ontology languages (2)

- Ontology languages are typically expressed by means of diagrams.
- The Entity-Relationship conceptual data model and UML Class Diagrams can be considered as ontology languages.

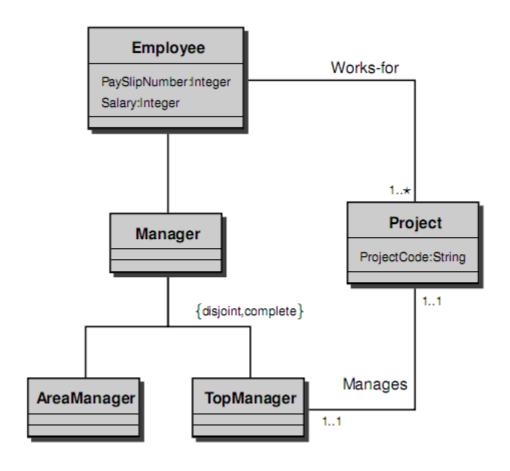
Ontology languages (3)

For example: Entity-Relationship Schema



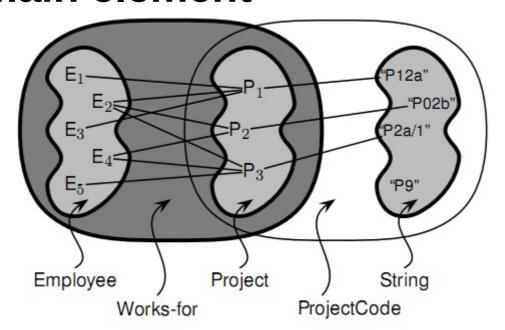
Ontology languages (4)

UML Class Diagram

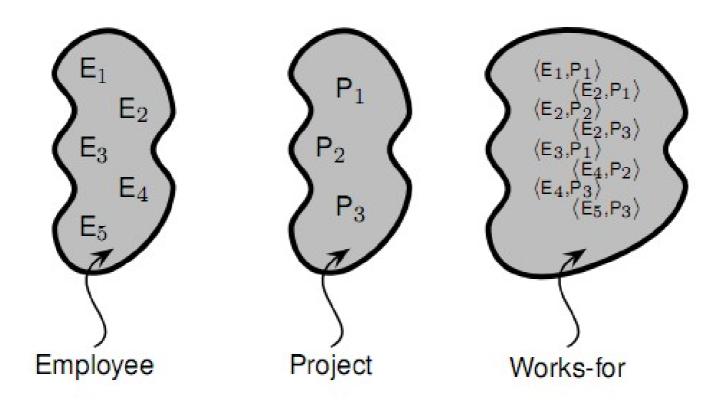


Meaning of basic constructs

- An entity/class is a set of instances
- An association (n-ary relationship) is a set of pairs (n-tuples) of instances
- An attribute is a set of pairs of an instance and a domain element



A world is described by sets of instances



The relational representation

Employee

employeeld	
E_1	
E_2	
E ₃	
E_4	
E ₅	

Project

projectld
P_1
P_2
P ₃

String

anystring	
"P12a"	
"P02b"	
"P2a/1"	
"P9"	
• • •	

Works-for

employeeld	projectld
E ₁	P_1
E_2	P_1
E ₂	P_2
E_2	P_3
E ₃	P_1
E_4	P_2
E_4	P_3
E ₅	P_3

ProjectCode

projectId	pcode
P_1	"P12a"
P_2	"P02b"
P_3	"P2a/1"

Meaning of the diagram

```
Works-for \subseteq Employee \times Project
Manages ⊂ TopManager × Project
Employee \subseteq \{e \mid \sharp (PaySlipNumber \cap (\{e\} \times Integer)) \ge 1\}
Employee \subseteq \{e \mid \sharp (Salary \cap (\{e\} \times Integer)) \geq 1\}
Project \subseteq \{p \mid \sharp(ProjectCode \cap (\{p\} \times String)) \ge 1\}
TopManager \subseteq \{m \mid 1 \geq \sharp (Manages \cap (\{m\} \times \Omega)) \geq 1\}
Project \subseteq \{p \mid 1 \geq \sharp (\mathsf{Manages} \cap (\Omega \times \{p\})) \geq 1\}
Project \subseteq \{p \mid \sharp (\mathsf{Works\text{-}for} \cap (\Omega \times \{p\})) \geq 1\}
Manager ⊆ Employee
AreaManager ⊆ Manager
TopManager ⊆ Manager
AreaManager \cap TopManager = \emptyset
Manager ⊆ AreaManager ∪ TopManager
```

Description Logic

- Description logics (DLs) is a family of knowledge representation languages that are widely used in ontological modelling.
- DLs provide one of the main underpinnings for the Web Ontology Language as standardized by the W3C.
- DLs are equipped with a formal semantics: a precise specification of the meaning DLs ontologies.
- The formal sematics allows human and computer systems
 - Exchanging DL ontologies without ambiguity as to their meaning.
 - Using logical deduction to infer additional information from the facts stated explicitly in an ontology
- The computation of inferences is called reasoning.

- The set N_I of individual names contains all names used to denote singuler entities in our domain of interest.
 - E.g. sun, moon,
- The set N_c of concept names contains names refer to types, categories, classes of entities, usually characterized by common properties.
 - E.g: Mammal, Country, ..
- The set N_R of role names contains names that denote binary relationships which may hold between individuals of a domain.
 - E.g: fatherOf, marriedWith, ...

(Complex) concept expressions are constructed as:

$$C, D ::= A \mid \top \mid \bot \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \forall R.C \mid \exists R.C$$

- A TBox (Terminological) is a set of statements of the form C

 □ D or C

 □ D, where C and D are concept expressions. They are called general inclusion axioms.
- An ABox (Assertion) consists of statements of the form C(a) or R(a,b), where C is a class expression, R is a role, and a, b are individuals.

- ABox expressions:
 - Individual assignments. E.g: Father(phong)
 - Property assignments. E.g: hasWife(phong,thuan)
- TBox expressions:
 - Subclass relationships. E.g: Father

 Man
 - Equivalence relationships. E.g: Mouse ≡ Mice
 - Conjunction. E.g: C □ D
 - Disjunction. E.g: C ⊔ D
 - Negation. E.g: ¬C
 - Top. E.g: T
 - Bottom. E.g: ⊥
 - Property restrictions. E.g: ∀R.C ∃R.C

(Complex) concept expressions are constructed as:

$$C, D ::= A \mid \top \mid \bot \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \forall R.C \mid \exists R.C$$

- A TBox (Terminological) is a set of statements of the form C

 □ D or C

 □ D, where C and D are concept expressions. They are called general inclusion axioms.
- An ABox (Assertion) consists of statements of the form C(a) or R(a,b), where C is a class expression, R is a role, and a, b are individuals.

```
Human ⊑ ∃hasParent.Human

Orphan ⊑ Human □ ∀hasParent.¬Alive

Orphan(harrypotter)

hasParent(harrypotter, jamespotter)
```

- The Description Logic SROIQ(D) includes:
 - SR = ALC + role chains
 - E.g: hasParent o hasBrother
 □ hasUncle
 FOL: ∀x ∀y (∃z ((hasParent(x, z) ∧ hasBrother(z, y) → hasUncle(x, y)))
 - includes S = ALC + transitivity
 - includes SH = S + role hierarchies
 - E.g: hasFather **□** hasParent

- The Description Logic SROIQ(D) includes:
 - O nominals (closed concepts)
 - E.g.: MyBirthdayGuests \equiv {lành, thơm, nụ}
 - I inverse roles
 - E.g: hasParent ≡ hasChild⁻
 Orphan ≡ ∀hasChild⁻.Dead
 - Q qualified cardinality restrictions
 - HappyFather ≡ ≥2 hasChildren.Female
 - Car = =4 hasTyre. T

- Property characteristics: Properties can be declared to be
 - Transitive
 - E.g. has Ancestor R(a,b) and R(b,c) => R(a,c)
 - Symmetric
 - E.g: hasSpouse

 $R(a,b) \Rightarrow R(b,a)$

- Asymmetric
 - E.g: hasChild

R(a,b) => not R(b,a)

- Reflexive
 - E.g: hasRelative

R(a,a) for all a

- Property characteristics: Properties can be declared to be
 - Irreflexive

E.g: parentOf not R(a,a) for all a

Functional

• E.g. has Husband R(a,b) and R(a,c) => b = c

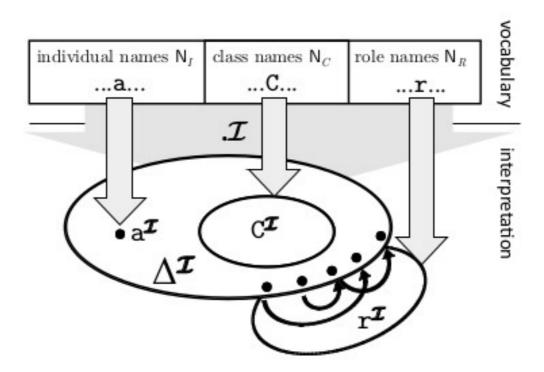
InverseFunctional

• E.g: has Husband R(a,b) and R(c,b) => a = c

- The Description Logic SROIQ(D) includes:
 - (D) datatypes
 - properties with datatype literals in second argument are called data properties or concrete roles
 - E.g 1: hasAge(john, "51"^^xsd:integer)
 - E.g 2: Teenager = Person □ ∃hasAge.(xsd:integer: ≥12 and ≤19)
 - note: this is not standard DL notation! It's really only used in OWL.

- The semantics of description logic is defined in a model-theoretic way.
- An interpretation (normally denoted with I) provides:
 - A nonempty set Δ^{I} , called the *domain* which can be understood as the entire of individuals that I represents.
 - A function I , called *interpretation function* which connects the vocabulary elements to Δ^{I} , by providing:

- for each individual name $\mathbf{a} \in \mathbf{N}_{l}$ a corresponding individual $\mathbf{a}^{l} \in \Delta^{l}$ from the domain.
- for each concept name $A \in N_c$ a corresponding set $A^I \in \Delta^I$ of the domain elements.
- for each role name $\mathbf{r} \in \mathbf{N}_{R}$ a corresponding set $\mathbf{r}^{l} \subseteq \Delta^{l} \times \Delta^{l}$ of ordered pairs of domain elements.



 Interpretation function extends to concept expressions in the obvious way

$$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$$

$$(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$$

$$(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$$

$$\{x\}^{\mathcal{I}} = \{x^{\mathcal{I}}\}$$

$$(\exists R.C)^{\mathcal{I}} = \{x \mid \exists y. \langle x, y \rangle \in R^{\mathcal{I}} \land y \in C^{\mathcal{I}}\}$$

$$(\forall R.C)^{\mathcal{I}} = \{x \mid \forall y. (x, y) \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}\}$$

$$(\leqslant nR)^{\mathcal{I}} = \{x \mid \#\{y \mid \langle x, y \rangle \in R^{\mathcal{I}}\} \leqslant n\}$$

$$(\geqslant nR)^{\mathcal{I}} = \{x \mid \#\{y \mid \langle x, y \rangle \in R^{\mathcal{I}}\} \geqslant n\}$$

• E.g 1:

```
\begin{array}{l} \text{Bird} \doteq \text{Animal} \; \sqcap \; \forall \text{SKIN.Feather} \\ \Delta^{\mathcal{I}} = \{ \text{tweety}, \text{goofy}, \text{fea1}, \text{fur1} \} \\ \text{Animal}^{\mathcal{I}} = \{ \text{tweety}, \text{goofy} \} \\ \text{Feather}^{\mathcal{I}} = \{ \text{fea1} \} \\ \text{SKIN}^{\mathcal{I}} = \{ \langle \text{tweety}, \text{fea1} \rangle, \langle \text{goofy}, \text{fur1} \rangle \} \\ \Longrightarrow \quad \text{Bird}^{\mathcal{I}} = \{ \text{tweety} \} \end{array}
```

• E.g 2:

Let $\mathcal{I}=(\Delta^{\mathcal{I}},\cdot^{\mathcal{I}})$ be defined by setting

$$\bullet \ \Delta^{\mathcal{I}} = \{a,b,c,d\};$$

•
$$A^{\mathcal{I}} = \{b, d\}, B^{\mathcal{I}} = \{c\};$$

$$ullet \ r^{\mathcal{I}} = \{(a,b), (a,c)\}, s^{\mathcal{I}} = \{(a,b), (a,d)\}.$$

Then

$$\bullet$$
 $(\forall r.A)^{\mathcal{I}} = \{b, c, d\}, (\forall s.A)^{\mathcal{I}} = \{a, b, c, d\};$

•
$$(\exists r.A \sqcap \forall r.A)^{\mathcal{I}} = \emptyset$$
, $(\exists s.A \sqcap \forall s.A)^{\mathcal{I}} = \{a\}$;

•
$$(\exists r.B \sqcap \exists r.A)^{\mathcal{I}} = \{a\}, (\exists r.(A \sqcap B))^{\mathcal{I}} = \emptyset;$$

$$ullet$$
 $(orall r.
eg A)^{\mathcal{I}} = \{b, c, d\}$, $(orall s.
eg A)^{\mathcal{I}} = \{b, c, d\}$.

Knowledge Base

 $\Sigma = (Tbox, Abox)$

- E.g: _{TBox:}

 - Mother \doteq Woman $\sqcap \exists ParentOf.Person$

 - MotherWithoutDaughter \doteq Mother $\sqcap \forall ParentOf. \neg Female$
 - GrandMother \doteq Woman $\sqcap \exists ParentOf.Parent$

ABox:

- GrandMother(Sally)
- $(Person \sqcap Male)(John)$

Reasoning

- Sentences: Axioms or assertions
- I is a model for a sentence S iff S is true in I
- I is a model for a DL knowledge base KB iff it is a model for every sentence in KB
- Models of KB are denoted by [KB]
- S is entailed by KB, written KB ⊨ S iff [KB] ⊆
 [S] (i.e. every model of K is a model of S)

Types of reasoning

KB a DL knowledge base

C and D are concepts

R is a role

a and b are individual names

- Instance checking: $KB \models C(a)$ or $KB \models R(a, b)$
- Subsumption checking: KB ⊨ C ⊑ D
- Equivalence checking: $KB \models C \equiv D$
- Consistency checking: KB ⊭ T ⊑ ⊥
- Concept satisfiability: KB ⊭ C ⊑ ⊥
- Disjoint concept: KB ⊨ C ⊓ D ⊑ ⊥

Semantic Web

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in co-operation."

Tim Berners-Lee - 2001

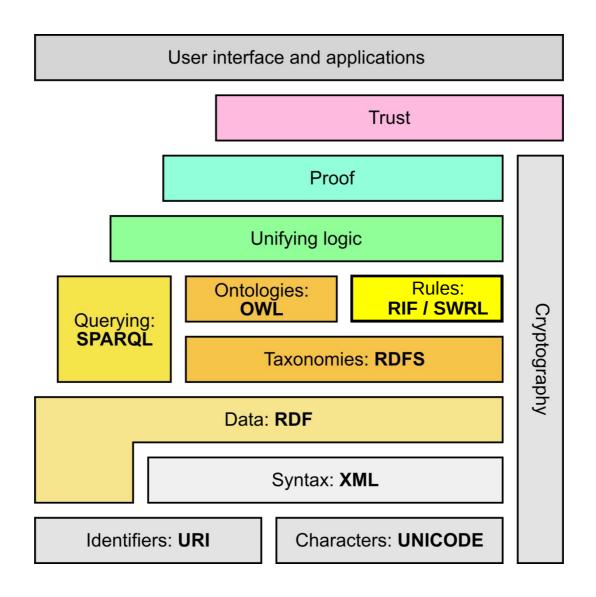
Semantic Web

- The content of the present Word Wide Web is nowadays only accessible and can be elaborated only by people.
- The Semantic Web is an enlargement of the WWW with semantic information that can be used by computers
- With the help of semantic information the content of pages could be processed automatically and computers could make inferences about a search.

Semantic Web

- The semantic web is not different from the www, is actually a developing part of it.
- The infrastructures and characteristics should be common
 - Use URI (Uniform resource Identifiers) addressing
 - Use protocols that a have a small and universally understood set of commands (like HTTP: Hypertext Transfer Protocol)

Semantic Web Vision



RESOURCE, URIs and NAMESPACES

- A resource is anything that has an identity
 - Digital (i.e an electronic document)
 - Physical (i.e. a book)
- A URI (Uniform Resource Identifier) is a character string that identifies a resource on the Web
 - URIs can follow different schemes
 - FTP (File Transfer Protocol)
 - HTTP (Hypertext Transfer Protocol
 - http://www.mysite.com/food.html

RESOURCE, URIs and NAMESPACES

- Namespaces are contexts, the domain of specific elements
- Namespaces are identified by a URI
- URIref: It is a URI with an optional fragment identifier attached to it, preceded by #

XML: Extensible Markup Language

- It is a general purpose markup Language for creating specific purpose mark-up languages
- With XML the single users can create their own tags (which is not possible with HTML)
- Differences between HTML and XML
 - HTML (Hypertext Markup Language)
 - Has a fixed set of tag
 - It is most frequently used to define the lay-out
 - Does not focus on the logical content or on the structur
 - XMI
 - It is possible to personally define the tags
 - Tags reflect a content
 - The layout is defined in a separate document (stylesheet)

- RDF is a general-purpose language for representing information in the web
 - Useful to represent metadata about Web resources
- RDF describes resources (Both abstract or concrete subjects) identifiable via an URI
- The syntax of RDF is based on XML
- RDF-documents are written as XML-documents with the tag rdf:RDF

- A RDF-statement is described by a triple (Subject, Predicate, Object)
 - Subject is a resource
 - Property is a property of a resource
 - **Object** is the value of a property of a resource

- E.g:
 - A Fact: Ora Lassila is the creator of the resource http://www.w3.org/Home/Lassila

```
<rdf:RDF>
  <rdf:Description about=
    "http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Lassila</s:Creator>
  </rdf:Description>
  </rdf:RDF>
```

How to represent above fact with DL?

 RDF gives a formalism for meta data annotation, and a way to write it down in XML, but it does not give any special meaning to vocabulary such as subClassOf or type

RDF Schema (RDFS)

- RDF Schema allows you to define vocabulary terms and the relations between those terms
 - it gives "extra meaning" to particular RDF predicates and resources
 - this "extra meaning", or semantics, specifies how a term should be interpreted

RDF Schema (RDFS)

- RDF Schema terms
- Class
- Property
- type
- subClassOf
- range
- domain
- These terms are the RDF Schema building blocks (constructors) used to create vocabularies:
- <Person,type,Class>
- <hasColleague,type,Property>
- <Professor,subClassOf,Person>
- <Carole,type,Professor>
- <hasColleague,range,Person>
- <hasColleague,domain,Person>

RDF Schema (RDFS)

- :mary rdf:type :Person
 - Person(mary)
- :Mother rdfs:subClassOf :Woman
- :john :hasWife :Mary
 - hasWife(john,mary)
- :hasWife rdfs:subPropertyOf :hasSpouse
 - hasWife ⊑ hasSpouse
- :hasWife rdfs:range :Woman
 - T

 ☐ ∀hasWife.Woman
- :hasWife rdfs:domain :Man
 - ∃hasWife. T ⊑ Man

- They are designed to define ontologies
- They are based on RDF and RDF-schema
- OWL (Web Ontology Language)

http://www.w3.org/TR/owl-features/

- It is an ontology description language
- It is a standard language for the modeling of ontologies
- Has additional vocabulary based on description logic

- OWL was based on SHOIN
 - only simple role hierarchy, and unqualified Nrs
- OWL 2 based on SHOIQ
 - ALC extended with transitive roles, a role box nominals, inverse roles and qualified number restrictions

Class/Concept Constructors

OWL Constructor	DL Syntax	Example
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human □ Male
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor ⊔ Lawyer
complementOf	$\neg C$	¬Male
oneOf	$\{x_1\}\sqcup\ldots\sqcup\{x_n\}$	{john} ⊔ {mary}
allValuesFrom	$\forall P.C$	∀hasChild.Doctor
someValuesFrom	$\exists P.C$	∃hasChild.Lawyer
maxCardinality	$\leqslant nP$	≤1hasChild
minCardinality	$\geqslant nP$	≥2hasChild

 Ontology Axioms: OWL ontology is a mixed set of TBox and ABox axioms

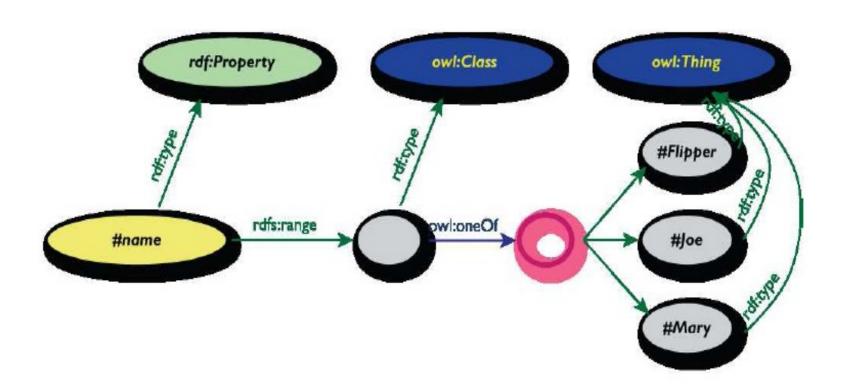
OWL Syntax	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human ⊑ Animal □ Biped
equivalentClass	$C_1 \equiv C_2$	Man ≡ Human □ Male
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter \sqsubseteq hasChild
equivalentProperty	$P_1 \equiv P_2$	cost ≡ price
transitiveProperty	$P^+ \sqsubseteq P$	ancestor ⁺ ⊑ ancestor

OWL Syntax	DL Syntax	Example
type	a:C	John : Happy-Father
property	$\langle a,b \rangle$: R	$\langle John, Mary \rangle$: has-child

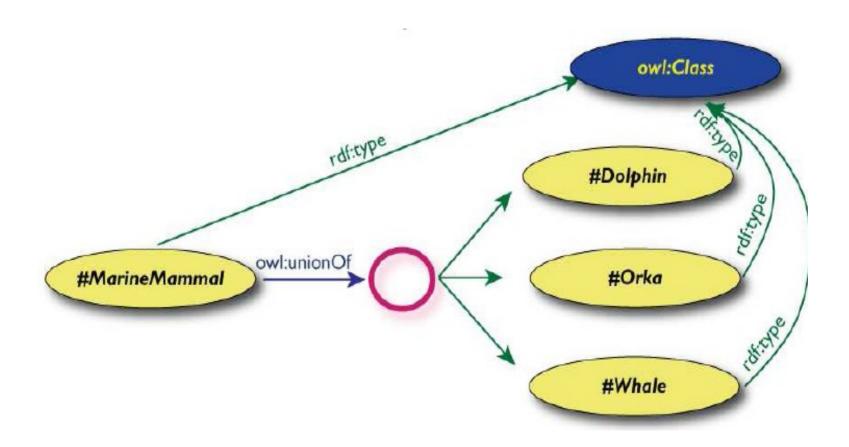
```
E.g., Person \sqcap \forall has Child. (Doctor \sqcup \exists has Child. Doctor):
<owl:Class>
  <owl:intersectionOf rdf:parseType=" collection">
    <owl:Class rdf:about="#Person"/>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasChild"/>
      <owl:allValuesFrom>
        <owl:unionOf rdf:parseType=" collection">
          <owl:Class rdf:about="#Doctor"/>
          <owl:Restriction>
             <owl:onProperty rdf:resource="#hasChild"/>
             <owl:someValuesFrom rdf:resource="#Doctor"/>
          </owl:Restriction>
        </owl:unionOf>
      </owl:allValuesFrom>
    </owl:Restriction>
  </owl:intersectionOf>
```

</owl:Class>

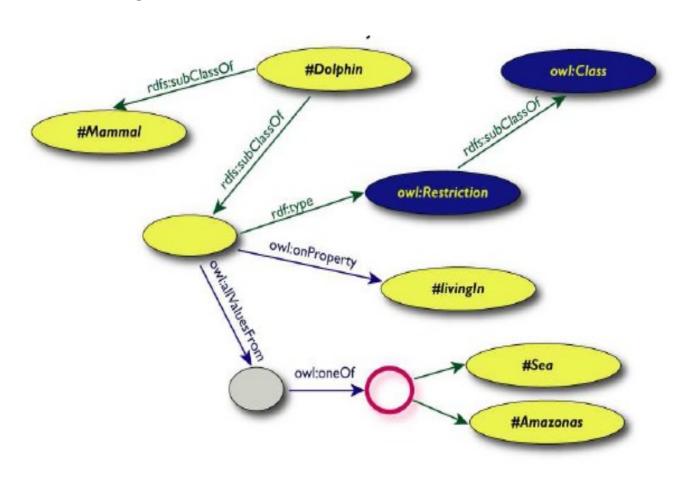
Definition of "name"



Definition of "marine mammal"



What is dolphin?



The SWRL rule syntax follows:

a
$$<=$$
 $b_1 \wedge b_2 \wedge \dots \wedge b_n$ where
a: head (an atom) b_s : body (all atoms)

 A SWRL knowledge base (K) is defined as follows:

```
K = (\sum, P) where

\sum = Knowledge base of SROIQ(D)

P = A finite set of rules
```

SWRL atoms are defined as follows:

Atom <= C(i) | D(v) | R(i, j) | U(i, v) |
builtIn(p,
$$v_1$$
, ..., v_n) | i = j | i <> j

C: Class D: Data type

R: Object Property **U:** Data type Property

i, j: Object variable names or Object individual names

 $\mathbf{v_1}$, ..., $\mathbf{v_n}$: Data type variable names or Data type value names

p: Built-in names

SWRL Semantics

Let
$$I = (\Delta^I, \Delta^D, I^I, D)$$
 where

I: interpretation

Δ^I: Object Interpretation domain

Δ^D: Datatype Interpretation domain

¹: Object Interpretation function

^{.D}: Datatype Interpretation function

$$\Delta^{I} \cap \Delta^{D} = \emptyset$$

such that
$$V_{IX} => P(\Delta^{I})$$
 $V_{DX} => P(\Delta^{D})$

 V_{IX} : object variables V_{DX} : datatype variables

P: the powerset operator

 The following table shows Binding B(I) for the SWRL atoms:

SWRL Atoms	Condition on Interpretation
C(i)	$i' \in C'$
R(i,j)	$(i^I,j^I)\in R^I$
U(i, v)	$(i^I, v^D) \in U^I$
D(v)	$v^D \in D^D$
$builtIn(p, v_1, \ldots, v_n)$	$(\mathbf{v}_1^D,\ldots,\mathbf{v}_n^D\in\mathbf{p}^D)$
i = j	i'=j'
$i \neq j$	$i' \neq j'$

- SWRL atoms in the antecedent are satisfied
 - if it is empty (trivially true)
 - or every atom of it is satisfied
- SWRL atom in the consequent is satisfied,
 - if it is not empty
 - and it is satisfied
- A rule is satisfied by an interpretation of I iff
 - every binding B(I) that satisfies the antecedent
 - B(I) satisfies the consequent

A rule asserting a fast computer:

FastComputer(?c) <= Computer(?c) ^ hasCPU(?c, ?cpu) ^ hasSpeed(?cpu, ?sp) ^ HighSpeed(?sp)

 The above rule using only DL can be expressed as follows:

Computer Π \exists hasCPU. \exists hasSpeed. \exists highSpeed \sqsubseteq FastComputer

- The translating of rules from SWRL to DL, depends on the number of variables based on shared variables between the consequent and antecedent.
- The number of variables shared between consequent and antecedent:
 - translating is possible, if 0 variable is shared, but at least one individual is shared
 - translating is possible, if 1 variable is shared
 - translating not possible, if 2 or more variables are shared

A Rule in SWRL but not in DL:

```
hasUncle(?x, ?y) <= hasParent(?x, ?z) ^ hasBrother(?z, ?y)
```

- The above rule cannot be translated into DL:
 - two different variables in the consequent