INF3580/4580 – Semantic Technologies – Spring 2017

Lecture 12: OWL: Loose Ends

Ernesto Jiménez-Ruiz

3rd April 2017

Department of Informatics

University of Oslo

Mandatory exercises

- Oblig 6 published after lecture.
- First attempt by April 25th.
- Second attempt by May 16th.

Outline

Make it simple!

- "Data level" with resources
- "Ontology level" with properties and "classes"
- Can have `rdf:type` relation between data objects and classes
- Allow a fixed vocabulary for relations between classes and properties
- Interpret:
  - Class as set of data objects
  - Property as relation between data objects

OWL 2 TBox and ABox

- The TBox
  - is for *terminological knowledge*
  - is independent of any actual instance data
  - is a set of axioms:
    - Class inclusion $\sqsubseteq$, equivalence $\equiv$
    - roles symmetric, asymmetric, reflexive, irreflexive, transitive,...
    - roles functional, inverse functional
    - inverse roles: $\mathit{hasParent} = \mathit{hasChild}^{-1}$
    - role inclusion $\mathit{hasBrother} \sqsubseteq \mathit{hasSibling}$
    - role chains $\mathit{hasParent} \circ \mathit{hasBrother} \sqsubseteq \mathit{hasUncle}$
  - Only certain combinations allowed

OWL 2 TBox and ABox

- The ABox
  - is for *assertional knowledge*
  - contains facts about concrete instances $a, b, c, \ldots$
  - A set of (negative) concept assertions $C(a)$, $\neg D(b) \ldots$
  - and (negative) role assertions $R(b, c)$, $\neg S(a, b)$
  - also `owl:sameAs`: $a = b$ and `owl:differentFrom`: $a \neq b$.

Assumptions

- Closed World Assumption
- Open World Assumption

- Unique Name Assumption
- Non-Unique Name Assumption

A Strange Catalogue

- We have seen many nice things that can be said in OWL
- Why the strange restrictions, e.g. on role axioms?
- Why not use 1st-order logic, could say much more?

- Because of the reasoning
  - Class satisfiability ($C \not\equiv \bot$)
  - Classification ($C \sqsubseteq D$)
  - Instance Check ($C(a)$)
  - ...
- All *decidable*
- Algorithm gives a correct answer after finite time

- Add a little more to OWL, and this is lost

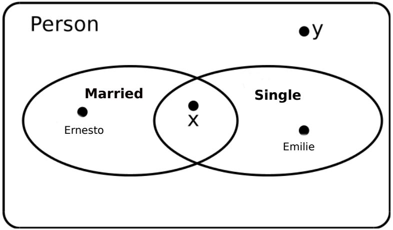
Outline

Single and Married

- Try to model the relationship between the concepts *Person*, *Married* and *Single*:
- First try:

$$\begin{aligned} \mathit{Single} &\sqsubseteq \mathit{Person} \\ \mathit{Married} &\sqsubseteq \mathit{Person} \end{aligned}$$

- General shape of a model:



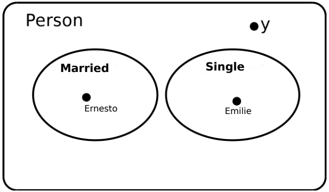
- x is both *Single* and *Married*, y is neither but a *Person*.

Disjointness Axioms

- Nothing should be both a *Single* and a *Married*
- Add a *disjointness* axiom for *Single* and *Married*
- Equivalent possibilities:

$$Single \sqcap Married \equiv \bot$$
$$Single \sqsubseteq \neg Married$$
$$Married \sqsubseteq \neg Single$$

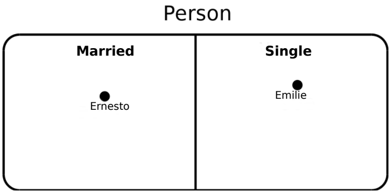
- General shape of a model:



- Specific support in OWL (`owl:disjointWith`) and Protégé

Covering Axioms

- Any *Person* should be either *Single* or *Married*.
- Add a *covering axiom* $Person \sqsubseteq Married \sqcup Single$
- General shape of a model (with disjointness):



- Specific support in Protégé (Edit Menu: "Add Covering Axiom")

Meat and Veggies

- Careful: not all subclasses are disjoint and covering
- Subclasses can be covering but not disjoint.
- E.g.

$$\begin{aligned} \mathit{MeatEatingMammal} &\sqsubseteq \mathit{Mammal} \\ \mathit{VeggieEatingMammal} &\sqsubseteq \mathit{Mammal} \end{aligned}$$

- All mammals eat either meat or vegetables. . .
- $\mathit{Mammal} \sqsubseteq \mathit{MeatEatingMammal} \sqcup \mathit{VeggieEatingMammal}$
- But there are mammals eating both
- No disjointness axiom for *MeatEatingMammal* and *VeggieEatingMammal*

Cats and Dogs

- Subclasses can be disjoint but not covering.
- E.g.

$$\begin{aligned} Cat &\sqsubseteq Mammal \\ Dog &\sqsubseteq Mammal \end{aligned}$$

- Nothing is both a cat and a dog: $Cat \sqsubseteq \neg Dog$
- But there are mammals which are neither
- No covering axiom with subclasses *Cat* and *Dog* for *Mammal*

Teachers and Students

- Subclasses can be neither disjoint nor covering.
- E.g.

$$\begin{aligned} \mathit{Teacher} &\sqsubseteq \mathit{Person} \\ \mathit{Researcher} &\sqsubseteq \mathit{Person} \end{aligned}$$

- There are people who are neither a researcher nor a teacher (yet)
- No covering axiom for these subclasses of *Person*
- There are people who are both a researcher and a teacher
- E.g. most PhD students
- No disjointness axiom for *Reasearcher* and *Teacher*

Outline

Keys

- A Norwegian is uniquely identified by his/her "personnummer"
  - Different Norwegians have different numbers
- Each customer in the DB is uniquely identified by the customer ID
  - No two customers with the same customer ID
  - Referred to as a *key* for a database table.
- A course is uniquely determined by code, semester, year.
  - E.g. $\langle$INF3580/4580, Spring, 2017$\rangle$
- R is a key for some set A if for all $x, y \in A$

$$x \; R \; k \quad \text{and} \quad y \; R \; k \quad \text{imply} \quad x = y$$

- So R is a key if it is "inverse functional"
  - There is a function giving exactly one object for every key value

Keys

- Keys in applications are usually (tuples of) literals
- Can we use "inverse functional datatype properties"?
- Reasoning about these is problematic
- Their exixtence would imply a literal as subject in a triple (not allowed in RDF)
- Therefore, datatype properties cannot be declared inverse functional in OWL 2

OWL 2 Keys

- OWL 2 includes special "hasKey" axioms
- Example: `Course hasKey {hasCode, hasSemester, hasYear}`
- Works for object properties and datatype properties.
- OWL Keys apply only to explicitly **named instances**
  - Makes reasoning tractable.
  - It may not be uspported by all OWL 2 reasoners

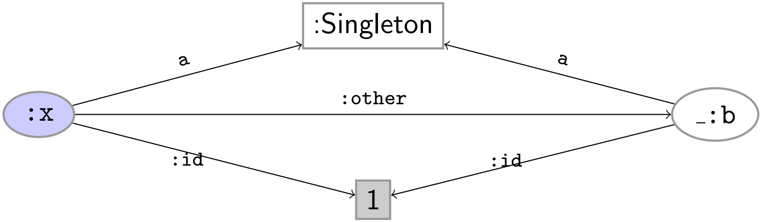
Reasoning with OWL Keys

- Given:
  - `:Norwegian hasKey {:personnr}`
  - `:drillo a :Norwegian`
  - `:drillo :personnr "12345698765"`
  - `:egil a :Norwegian`
  - `:egil   :personnr "12345698765"`
- Can infer:
  - `:drillo owl:sameAs :egil`

- Given:
  - `:Singleton hasKey {:id}`
  - `:Singleton` $\sqsubseteq$ `:id value 1`
  - `:x a :Singleton`
  - `:y a :Singleton`
- Can infer:
  - `:x owl:sameAs :y`

What's with the "named instances"?

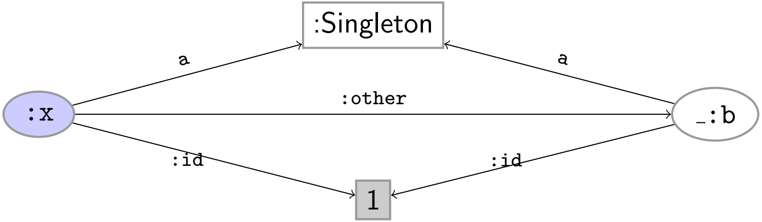
- Given:
  - :Singleton hasKey {:id}
  - :Singleton ⊑ :id value 1
  - :x a :Singleton
  - :Singleton ⊑ :other some :Singleton



- Since _:b is a blank node, and therefore not an explicitly named instance,
- the reasoner does not infer :x owl:sameAs _:b.

What's with the "named instances"?

- Given:
  - `:Singleton hasKey {:id}`
  - `:Singleton ⊑ :id value 1`
  - `:x a :Singleton`
  - `:Singleton ⊑ :other some (:Singleton and not {:x})`



- This is *not* inconsistent.
- Distinct keys only required for explicitly named individuals.

Outline

Punning

- Remember: In OWL strict separation of classes, properties and individuals. However, not entirely correct...
- OWL 2 introduces *punning*, allowing one URI to be used for, e.g., both a class and an individual,
- but not both a class and a datatype property, or for different property types.
- Example:
  ```
  :Joe       rdf:type    :Eagle .
  :Eagle     rdf:type    :Species .
  ```
  `:Eagle` is both a class and an individual.
- However, semantically, "punned" URI are treated as different terms. (under the hood)
  - Meaning, the class `:Eagle` is different from the individual `:Eagle`.
  - Axioms about the class is not transferred to the individual, or vice versa.

Outline

A tempting mistake

- Cardinality restrictions are not suitable to express
  - durations
  - intervals
  - or any kind of sequence
  - and they cannot be used for arithmetic
- Anti-pattern:
  - Scotch whisky is aged at least 3 years:
  - Use a datatype property *age* with range *int*.
  - $Scotch \sqsubseteq Whisky \sqcap \geq_3 age.int$

- Why?
  - This says that Scotch has at least 3 *different ages*
  - For instance -1, 0, 15

A possible solution

- Idea: don't use age.
- Use a property *casked*
  - domain *Whisky*
  - range *int*
  - relates the whisky to each year it is in the cask.

e.g. `:young :casked "2000"^^int, "2001"^^int, "2002"^^int`

- $Scotch \sqsubseteq Whisky \sqcap \geq_3 casked.int$
- Works, but...
- Can't express e.g. that the years are consecutive
  - Knowing a whisky is casked in 2000 and 2009 doesn't imply it is casked for 10 years.
- Reasoning about $\geq_n$ often works by generating n sample instances
  - $Town \equiv \geq_{10000} inhabitant.Person$
  - $Metropolis \equiv \geq_{1000000} inhabitant.Person$
  - Will kill almost any reasoner

Reminder: Datatype properties

- OWL distinguishes between
  - object properties: go from resources to resources
  - datatype properties: go from resources to literals
- OWL (2) prescribes a list of available built-in datatypes for literals
  - Numbers: real, rational, integer, positive integer, double, long,...
  - Strings
  - Booleans
  - Binary data
  - IRIs
  - Time Instants
  - XML Literals
- Varying tool support (e.g., depending on editor and reasoner)
- Possible to define custom datatypes (e.g. datatype “age” as xsd:integer[≥ 0, ≤ 130])

Data Ranges

- Like concept descriptions, only for data types
- Boolean combinations allowed (Manchester syntax)
  - `xsd:integer or xsd:string`
  - `xsd:integer and not xsd:byte`
- Each basic datatype can be restricted by a number of *facets*
  - `xsd:integer[`$\geq$ `9]` – integers ≥ 9.
  - `xsd:integer[`$\geq$ `9,` $\leq$ `11]` – integers between 9 and 11.
  - `xsd:string[length 5]` – strings of length 5.
  - `xsd:string[maxLength 5]` – strings of length ≤ 5.
  - `xsd:string[minLength 5]` – strings of length ≥ 5.
  - `xsd:string[pattern "[01]*"]` – strings consisting of 0 and 1.

Range Examples

- A whisky that is at least 12 years old:
  `Whisky and age some integer[>= 12]`
- A teenager:
  `Person and age some integer[>= 13, <= 19]`
- A metropolis:
  `Place and noInhabitants some integer[>= 1000000]`

- Note: often makes best sense with functional properties
  Why?

Pattern Examples

- An integer or a string of digits
  - `xsd:integer or xsd:string[pattern "[0-9]+"]`
- ISBN numbers: 13 digits in 5 --separted groups, first 978 or 979, last a single digit.
  - `Book` $\sqsubseteq$ `ISBN some string[length 17 ,`
    `pattern "97[89]-[0-9]+-[0-9]+-[0-9]+-[0-9]"]`
- Reasoning about patterns:
  - `R` a functional datatype property
  - $A \equiv$ `R some string[pattern "(ab)*"]`
  - $B \equiv$ `R some string[pattern "a(ba)*b"]`
  - Reasoner can find out that $B \sqsubseteq A$.

Outline

Expressivity

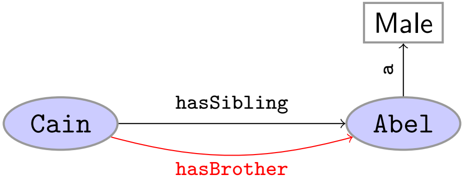
- Certain *relationships* between concepts and properties can't be expressed in OWL
- E.g.
  - Given that property *hasSibling* and class *Male* are defined...
  - ...cannot say that $hasBrother(x, y)$ iff $hasSibling(x, y)$ and $Male(y)$.
- Usually, adding such missing relationships would lead to undecidability
- *Not* easy to show that something is not expressible
  - We look at some examples, not proofs

Brothers

- Given terms

$$hasSibling \qquad Male$$

- . . . a brother is *defined* to be a sibling who is male



- Best try:

$$hasBrother \sqsubseteq hasSibling$$
$$\top \sqsubseteq \forall hasBrother.Male \qquad \text{or: } \mathrm{rg}(hasBrother, Male)$$
$$\exists hasSibling.Male \sqsubseteq \exists hasBrother.\top$$

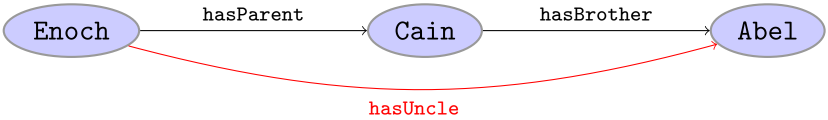
- Not enough to infer that *all* male siblings are brothers

Uncles

- Given terms

$$hasParent \qquad hasBrother$$

- ...an uncle is *defined* to be a brother of a parent.



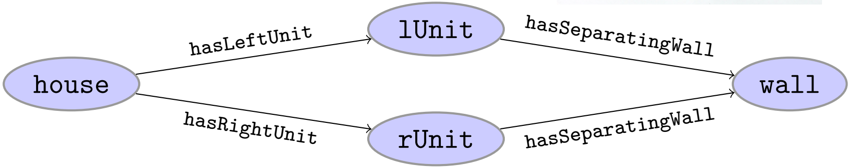
- Best try:

$$hasParent \circ hasBrother \sqsubseteq hasUncle$$
$$hasUncle \sqsubseteq hasParent \circ hasBrother$$

- properties cannot be declared sub-properties of property chains in OWL 2.
  - problematic for reasoning

Diamond Properties

- A semi-detached house has a left and a right unit
- Each unit has a separating wall
- The separating walls of the left and right units are the same
- "diamond property"



- Try...

$$SemiDetached \sqsubseteq \exists hasLeftUnit.Unit \sqcap \exists hasRightUnit.Unit$$
$$Unit \sqsubseteq \exists hasSeparatingWall.Wall$$

- But this does not guarantee to use the same wall

Connecting Datatype Properties

- Given terms

$$Person \qquad hasChild \qquad hasBirthday$$

- A twin parent is defined to be a person who has two children with the same birthday.
- Try...

$$\begin{aligned} TwinParent \equiv Person \quad & \sqcap \; \exists hasChild.\exists hasBirthday[\ldots] \\ & \sqcap \; \exists hasChild.\exists hasBirthday[\ldots] \end{aligned}$$

- No way to connect the two birthdays to say that they're the same.
  - (and no way to say that the children are *not* the same)
- Try...

$$TwinParent \equiv Person \sqcap \geq_2 hasChild.\exists hasBirthday[\ldots]$$

- Still no way of connecting the birthdays

Reasoning about Numbers

- Reasoning about natural numbers is undecidable in general.
- DL Reasoning is decidable
- Therefore, general reasoning about numbers can't be "encoded" in DL
- Cannot encode addition, multiplication, etc.
- Note: a lot can be done with other logics, but not with DLs
  - Outside the intended scope of Description Logics

Combining OWL 2 and Rules

Some limitation may be addressed

- SWRL: Semantic Web Rule Language
- Uses XML syntax based on RuleML
- OWL 2 + unrestricted SWRL leads to undecidability
- Restricted SWRL + OWL is decidable and very powerful
- A bit more in the next SPARQL lesson

Outline

OWL 2 profiles

- OWL 2 has various *profiles* that correspond to different DLs.
- OWL 2 DL is the "normal" OWL 2 (sublanguage): "maximum" expressiveness while keeping reasoning problems decidable—but still very expensive.
- (Other) profiles are tailored for specific ends, e.g.,
  - OWL 2 QL:
    - Specifically designed for efficient database integration.
  - OWL 2 EL:
    - A lightweight language with polynomial time reasoning.
  - OWL 2 RL:
    - Designed for compatibility with rule-based inference tools.
- OWL Full: Anything goes: classes, relations, individuals, ... like in RDFS, are not kept apart. Highly expressive, not decidable. But we want OWL's reasoning capabilities, so stay away if you can—and you almost always can.

OWL 2 Validator: `http://owl.cs.manchester.ac.uk/validator/`

OWL EL

Based on DL $\mathcal{EL}^{++}$.

$\mathcal{EL}^{++}$ concept descriptions, simplified

$C, D \rightarrow$	A	$\mid$	(atomic concept)
	$\top$	$\mid$	(universal concept)
	$\bot$	$\mid$	(bottom concept)
	$\{a\}$	$\mid$	(*singular* enumeration)
	$C \sqcap D$	$\mid$	(intersection)
	$\exists R.C$	$\mid$	(existential restriction)

Axioms

- $C \sqsubseteq D$ and $C \equiv D$ for concept descriptions D and C.
- $P \sqsubseteq Q$ and $P \equiv Q$ for roles P, Q. Also Domain and Range.
- $C(a)$ and $R(a, b)$ for concept C, role R and individuals a, b.

OWL EL contd.

Not supported, simplified:

- negation, (NB, disjointness of classes: $C \sqcap D \sqsubseteq \bot$ possible),
- disjunction,
- universal quantification,
- cardinalities,
- inverse roles,
- plus some role characteristics.
- reduced list of datatypes (e.g., not supported "boolean" nor "double")

Complete list: `http://www.w3.org/TR/owl2-profiles/#Feature_Overview`.

- Checking ontology consistency, class expression subsumption, and instance checking is in **P**.
- "Good for large ontologies."
- Used in many biomedical ontologies (e.g. SNOMED CT).

OWL QL

Based on DL-Lite$_R$.

DL-Lite$_R$ concept descriptions, simplified

$C \rightarrow$	A	$\mid$	(atomic concept)
	$\exists R.\top$	$\mid$	(existential restriction with $\top$ only)
$D \rightarrow$	A	$\mid$	(atomic concept)
	$\exists R.D$	$\mid$	(existential restriction)
	$\neg D$	$\mid$	(negation)
	$D \sqcap D'$	$\mid$	(intersection)

Axioms

- $C \sqsubseteq D$ for concept descriptions D and C (and $C \equiv C'$).
- $P \sqsubseteq Q$ and $P \equiv Q$ for roles P, Q. Also Domain and Range.
- $C(a)$ and $R(a, b)$ for concept C, role R and individuals a, b.

OWL QL contd.

Not supported, simplified:

- disjunction,
- universal quantification,
- cardinalities,
- functional roles, keys,
- $=$ (`SameIndividual`)
- enumerations (closed classes),
- subproperties of chains, transitivity
- reduced list of datatypes (e.g., not supported "boolean" nor "double")

Complete list: `http://www.w3.org/TR/owl2-profiles/#Feature_Overview_2`.

- Captures language for which queries can be translated to SQL.
- "Good for large datasets."
- We will see more in the Ontology Based Data Access (OBDA) lesson

OWL2: RL

OWL 2 RL is based on the description logic $\mathcal{RL}$ (also called DLP):

RL-concepts

$C \rightarrow$	A	$\mid$	(atomic concept)
	$C \sqcap C'$	$\mid$	(intersection)
	$C \sqcup C'$	$\mid$	(union)
	$\exists R.C$	$\mid$	(existential restriction)
$D \rightarrow$	A	$\mid$	(atomic concept)
	$D \sqcap D'$	$\mid$	(intersection)
	$\forall R.D$	$\mid$	(universal restriction)

Axioms

- $C \sqsubseteq D$, $C \equiv C'$, $\top \sqsubseteq \forall R.D$, $\top \sqsubseteq \forall R^{-}.D$ $R \sqsubseteq P$, $R \equiv P^{-}$ and $R \equiv P$ for roles R, P and concept descriptions C and D. Also Domain and Range.
- $C(a)$ and $R(a, b)$ for concept C, role R and individuals a, b.

OWL RL contd.

- Puts constraints in the way in which constructs are used (i.e., syntactic subset of OWL 2).
- So that OWL 2 RL axioms can be directly translated into datalog rules
- Enables desirable computational properties using rule-based reasoning engines.
- It also imposes a reduced list of allowed datatypes (e.g., not supported "real" nor "rational")

- We will see more in the next SPARQL lesson.

Complete list of characteristics: `http://www.w3.org/TR/owl2-profiles/#Feature_Overview_3`.

EXERCISE: Property axioms expressed as DL-axioms

$$
\begin{aligned}
\exists R.\top &\sqsubseteq C \\
\top &\sqsubseteq \forall R.C \\
R \circ R &\sqsubseteq R \\
\top &\sqsubseteq \ \leq 1\, R.\top \\
\top &\sqsubseteq \ \leq 1\, R^{-}.\top \\
R &\sqsubseteq R^{-} \\
R &\sqsubseteq \neg R^{-} \\
\top &\sqsubseteq \exists R.\mathit{Self} \\
\exists R.\mathit{Self} &\sqsubseteq \bot
\end{aligned}
$$

EXERCISE: Property axioms expressed as DL-axioms

$\exists R.\top \sqsubseteq C$	Domain	$(\exists hasPet.\top \sqsubseteq Person)$
$\top \sqsubseteq \forall R.C$	Range	$(\top \sqsubseteq \forall hasPet.(Animal \sqcap \neg Person))$
$R \circ R \sqsubseteq R$	Transitivity	$(\text{ancestorOf} \circ ancestorOf \sqsubseteq ancestorOf)$
$R_1 \equiv R_2^-$	Inverse	$(\text{partOf} \equiv hasPart^-)$
$\top \sqsubseteq\ \leq 1\, R.\top$	Functionality	$(\top \sqsubseteq\ \leq 1\, hasSpouse.\top)$
$\top \sqsubseteq\ \leq 1\, R^-.\top$	Inverse Functionality	$(\top \sqsubseteq\ \leq 1\, hasSpouse^-.\top)$
$R \sqsubseteq R^-$	Symmetry	$(\text{friendOf} \sqsubseteq friendOf^-)$
$R \sqsubseteq \neg R^-$	Asymmetry	$(\text{partOf} \sqsubseteq \neg partOf^-)$
$\top \sqsubseteq \exists R.Self$	Reflexive	$(\top \sqsubseteq \exists hasRelative.Self)$
$\exists R.Self \sqsubseteq \bot$	Irreflexive	$(\exists parentOf.Self \sqsubseteq \bot)$

Next

- Guest lecture:
  - April 24
  - **Veronika Hemsbakk** (Acando `https://www.acando.no/`)
  - Theoretic aspects of **SHACL** (`https://www.w3.org/TR/shacl/`) covering how to build up a shape, the different core constraints and validation result graphs.
  - Application (demo) within the **eInnsyn project** `https://einnsyn.difi.no/`
  - **Exam will include questions from guest lecture**
- May 8: More (practical) details about SPARQL and rules (Ernesto)
- May 15: OBDA, R2RML, query rewriting (Ernesto)
- May 22: Linked Open Data (Leif)