Practical Reasoning for the Semantic Web

Day 4

Stefan Schlobach & Heiner Stuckenschmidt

August 18, 2005

http://www.few.vu.nl/~schlobac/
http://www.few.vu.nl/~heiner/

Reasoning to support the modeling process

- Logical modeling
  - formal correctness
  - formal conciseness
  - formal completeness
- Explaining logical modeling errors
- Explaining subsumption
  - Explaining concept subsumption
  - Explaining terminological subsumption
- Modularization of Ontologies

Reasoning to support the modeling process

- Logical modeling
  - formal correctness
  - formal conciseness
  - formal completeness
- Explaining logical modeling errors
- Explaining subsumption
  - Explaining concept subsumption
  - Explaining terminological subsumption
- Modularization of Ontologies

It's not all that easy

$$
\begin{aligned}
Hepatitis &\sqsubseteq \exists tract.DigestiveS \sqcap \forall tract.DigestiveS \sqcap \exists located.Liver \sqcap LiverD \\
&\quad \sqcap \exists abnormality.(Infarct \sqcup Drug \sqcup Inflammation \sqcup Infection \sqcup \ldots) \\
InfHepatitis &\doteq Hepatitis \sqcap \exists abnormality.Infection \sqcap \forall abnormality.Infection \\
&\quad \sqcap \forall etiology.(EBVirus \sqcup HVirus \sqcup Amoeba \sqcup Spirochaetes) \\
VirHep &\doteq \forall etiology.(HVirus \sqcup EB\text{-}Virus \sqcup CytomegaloVirus) \sqcap InfHepatitis \\
LiverD &\sqsubseteq \forall tract.DigestiveS \sqcap \exists tract.DigestiveS \sqcap Diagnosis \sqcap \exists located.Liver \\
Diagnosis &\sqsubseteq ReasonForAdmission \\
Liver &\sqsubseteq \exists part.DigestiveS \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \forall region.Abdomen \\
Kidney &\sqsubseteq \exists part.GUSystem \sqcap \exists part.ESystem \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \\
&\quad \forall region.Abdomen \sqcap \forall side.(Left \sqcup Right)
\end{aligned}
$$

Is this a good terminology?

No, from a formal perspective, it is not !

Formally Good Modeling

Let's take logic a step back, and answer the following questions:

- Is my terminology formally correct?
  - Is it non-contradictory, and is the implicit information correct?
- Is my terminology formally concise?
  - Is it more complex than necessary?
- Is my terminology formally complete?
  - Is all the relevant information modeled?

We have to do two things:

1. Define criteria for formal correctness, conciseness and completeness.
2. Give methods of how to ensure the criteria

Criteria and Methods for formal correctness

- Criterion 1: Consistency
  - all concepts are satisfiable (but this is not enough)
  - local consistency
- Criterion 2: verifiability (making implicit information explicit)
  - taxonomy (order concept-names according to isa ordering)
  - non-atomic subsumption (subsumption of complex concepts)

Criteria and Methods for formal correctness

- Criterion 1: Consistency
  - all concepts are satisfiable (but this is not enough)
  - local consistency
- Criterion 2: verifiability (making implicit information explicit)
  - taxonomy (order concept-names according to isa ordering)
  - non-atomic subsumption (subsumption of complex concepts)

Correctness: local inconsistency

$$\begin{aligned}
\mathit{Hepatitis} &\sqsubseteq \exists \mathit{tract.DigestiveS} \sqcap \forall \mathit{tract.DigestiveS} \sqcap \exists \mathit{located.Liver} \sqcap \mathit{LiverD} \\
&\quad \sqcap \exists \mathit{abnormality.(Infarct} \sqcup \mathit{Drug} \sqcup \mathit{Inflammation} \sqcup \mathit{Infection} \sqcup \ldots) \\
\mathit{InfHepatitis} &\doteq \mathit{Hepatitis} \sqcap \exists \mathit{abnormality.Infection} \sqcap \forall \mathit{abnormality.Infection} \\
&\quad \sqcap \forall \mathit{etiology.(EBVirus} \sqcup \mathit{HVirus} \sqcup \mathit{Amoeba} \sqcup \mathit{Spirochaetes}) \\
\mathit{VirHep} &\doteq \forall \mathit{etiology.(HVirus} \sqcup \mathit{EB\text{-}Virus} \sqcup \mathit{CytomegaloVirus}) \sqcap \mathit{InfHepatitis} \\
\mathit{LiverD} &\sqsubseteq \forall \mathit{tract.DigestiveS} \sqcap \exists \mathit{tract.DigestiveS} \sqcap \mathit{Diagnosis} \sqcap \exists \mathit{located.Liver} \\
\mathit{Diagnosis} &\sqsubseteq \mathit{ReasonForAdmission} \\
\mathit{Liver} &\sqsubseteq \exists \mathit{part.DigestiveS} \sqcap \mathit{BodyPart} \sqcap \exists \mathit{region.Abdomen} \sqcap \forall \mathit{region.Abdomen} \\
\mathit{Kidney} &\sqsubseteq \exists \mathit{part.GUSystem} \sqcap \exists \mathit{part.ESystem} \sqcap \mathit{BodyPart} \sqcap \exists \mathit{region.Abdomen} \sqcap \\
&\quad \forall \mathit{region.Abdomen} \sqcap \forall \mathit{side.(Left} \sqcup \mathit{Right})
\end{aligned}$$

Formally: There is a part ($\sqcup$ CytomegaloVirus) of a formula that can be left out without changing the formal semantics?

Computational challenge: We cannot check all sub-formulae. How do we check for local inconsistency?

Criteria and Methods for formal conciseness

- Criterion 1: Structured Modeling
  - Modularity: Is the terminology modeled in a modular way?
    - clustering of hierarchically related concepts
    - separation of semantically independent axioms
  - Similarity: are similar concepts treated similarly?
- Criterion 2: Presentation of Information
  - Redundancy: Is information modeled in a redundant way?
    - check whether a terminology is equivalent to a terminology where a concept in the first is replaced by a superconcept according to the subsumption hierarchy.

Criteria and Methods for formal conciseness

- Criterion 1: Structured Modeling
  - Modularity: Is the terminology modeled in a modular way?
    - clustering of hierarchically related concepts
    - separation of semantically independent axioms
  - Similarity: are similar concepts treated similarly?
- Criterion 2: Presentation of Information
  - Redundancy: Is information modeled in a redundant way?
    - check whether a terminology is equivalent to a terminology where a concept in the first is replaced by a superconcept according to the subsumption hierarchy.

Conciseness - modules

$$\begin{aligned}
\mathit{Hepatitis} &\sqsubseteq \exists \mathit{tract.DigestiveS} \sqcap \forall \mathit{tract.DigestiveS} \sqcap \exists \mathit{located.Liver} \sqcap \mathit{LiverD} \\
&\quad \sqcap \exists \mathit{abnormality.(Infarct} \sqcup \mathit{Drug} \sqcup \mathit{Inflammation} \sqcup \mathit{Infection} \sqcup \ldots) \\
\mathit{InfHepatitis} &\doteq \mathit{Hepatitis} \sqcap \exists \mathit{abnormality.Infection} \sqcap \forall \mathit{abnormality.Infection} \\
&\quad \sqcap \forall \mathit{etiology.(EBVirus} \sqcup \mathit{HVirus} \sqcup \mathit{Amoeba} \sqcup \mathit{Spirochaetes}) \\
\mathit{VirHep} &\doteq \forall \mathit{etiology.(HVirus} \sqcup \mathit{EB\text{-}Virus} \sqcup \mathit{CytomegaloVirus}) \sqcap \mathit{InfHepatitis} \\
\mathit{LiverD} &\sqsubseteq \forall \mathit{tract.DigestiveS} \sqcap \exists \mathit{tract.DigestiveS} \sqcap \mathit{Diagnosis} \sqcap \exists \mathit{located.Liver} \\
\mathit{Diagnosis} &\sqsubseteq \mathit{ReasonForAdmission}
\end{aligned}$$

$$\begin{aligned}
\mathit{Liver} &\sqsubseteq \exists \mathit{part.DigestiveS} \sqcap \mathit{BodyPart} \sqcap \exists \mathit{region.Abdomen} \sqcap \forall \mathit{region.Abdomen} \\
\mathit{Kidney} &\sqsubseteq \exists \mathit{part.GUSystem} \sqcap \exists \mathit{part.ESystem} \sqcap \mathit{BodyPart} \sqcap \exists \mathit{region.Abdomen} \sqcap \\
&\quad \forall \mathit{region.Abdomen} \sqcap \forall \mathit{side.(Left} \sqcup \mathit{Right})
\end{aligned}$$

Based on semantic independence

Criteria and methods for formal completeness

- Criterion 1: Concept Exhaustiveness
  - similarity-based
    - highlight structural differences between similar concepts in the terminology
- Criterion 2: Terminological Completeness
  - similarity based
    - extension based on similarity between concepts
  - concept grouping
    - look for co-occurrences of concepts in atomic disjunctions
    - this suggests to introduce a super-concept.

Criteria and methods for formal completeness

- Criterion 1: Concept Exhaustiveness
  - similarity-based
    - highlight structural differences between similar concepts in the terminology
- Criterion 2: Terminological Completeness
  - similarity based
    - extension based on similarity between concepts
  - concept grouping
    - look for co-occurrences of concepts in atomic disjunctions
    - this suggests to introduce a super-concept.

An Example for concept exhaustiveness

$$
\begin{aligned}
Hepatitis \sqsubseteq\ & \exists tract.DigestiveS \sqcap \forall tract.DigestiveS \sqcap \exists located.Liver \sqcap LiverD \\
& \sqcap \exists abnormality.(Infarct \sqcup Drug \sqcup Inflammation \sqcup Infection \sqcup \ldots) \\
InfHepatitis \doteq\ & Hepatitis \sqcap \exists abnormality.Infection \sqcap \forall abnormality.Infection \\
& \sqcap \forall etiology.(EBVirus \sqcup HVirus \sqcup Amoeba \sqcup Spirochaetes) \\
VirHep \doteq\ & \forall etiology.(HVirus \sqcup EB\text{-}Virus \sqcup CytomegaloVirus) \sqcap InfHepatitis \\
LiverD \sqsubseteq\ & \forall tract.DigestiveS \sqcap \exists tract.DigestiveS \sqcap Diagnosis \sqcap \exists located.Liver \\
Diagnosis \sqsubseteq\ & ReasonForAdmission \\
Liver \sqsubseteq\ & \exists part.DigestiveS \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \forall region.Abdomen \\
Kidney \sqsubseteq\ & \exists part.GUSystem \sqcap \exists part.ESystem \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \\
& \forall region.Abdomen \sqcap \forall side.(Left \sqcup Right) \\
Appendix \sqsubseteq\ & \exists part.DigestiveS \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \forall region.Abdomen
\end{aligned}
$$

⟶ shouldn't we be able to distinguish an appendix from a liver?

Criteria and Methods for formal completeness

- Criterion 1: Concept Exhaustiveness
  - similarity-based
    - highlight structural differences between similar concepts (as discussed above) in the terminology
- Criterion 2: Terminological Completeness
  - similarity based
    - extension based on similarity between concepts
  - concept grouping
    - look for co-occurrences of concepts in atomic disjunctions
    - this suggests to introduce a super-concept.

An Example for terminological completeness

$$
\begin{aligned}
Hepatitis \sqsubseteq\ & \exists tract.DigestiveS \sqcap \forall tract.DigestiveS \sqcap \exists located.Liver \sqcap LiverD \\
& \sqcap \exists abnormality.(Infarct \sqcup Drug \sqcup Inflammation \sqcup Infection \sqcup \ldots) \\
InfHepatitis \doteq\ & Hepatitis \sqcap \exists abnormality.Infection \sqcap \forall abnormality.Infection \\
& \sqcap \forall etiology.(EBVirus \sqcup HVirus \sqcup Amoeba \sqcup Spirochaetes) \\
VirHep \doteq\ & \forall etiology.(HVirus \sqcup EB\text{-}Virus \sqcup CytomegaloVirus) \sqcap InfHepatitis \\
LiverD \sqsubseteq\ & \forall tract.DigestiveS \sqcap \exists tract.DigestiveS \sqcap Diagnosis \sqcap \exists located.Liver \\
Diagnosis \sqsubseteq\ & ReasonForAdmission \\
Liver \sqsubseteq\ & \exists part.DigestiveS \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \forall region.Abdomen \\
Kidney \sqsubseteq\ & \exists part.GUSystem \sqcap \exists part.ESystem \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \\
& \forall region.Abdomen \sqcap \forall side.(Left \sqcup Right)
\end{aligned}
$$

⟶ shouldn't we also have a definition for a disease of the kidney?

Reasoning to support the modeling process

- Logical modeling
  - formal correctness
  - formal conciseness
  - formal completeness
- Explaining logical modeling errors
  - Debugging
  - Diagnosis
- Explaining subsumption
  - Explaining concept subsumption
  - Explaining terminological subsumption
- Modularization of Ontologies

Initial motivation: The DICE terminology

- DICE = Frame based terminology for Intensive Care Units
- Developed at Amsterdam Medical Center (AMC) for
  - unambiguous and
  - unified classification at patient admission.
  - initiated and supported by NICE foundation
- To evaluate the modeling, DICE is
  - translated into Description Logics and
  - checked with RACER, e.g., for consistency or equivalence

Frames and Description Logic

- Frame

Class	*SuperClass*	*Slot*	*Slot-Value*
Brain	*BodyPart*	*region*	*Head and Neck*
	CNS	*system part*	*NervousSystem*
CNS	*NervousSystem*		
BodyPart	*AnatLocation*		

- Description Logic Terminological Axioms

$$\begin{aligned}
&Brain \,\dot{\sqsubseteq}\, CNS \sqcap \exists systempart.NervousSystem \sqcap \\
&\quad BodyPart \sqcap \exists region.HeadAndNeck \sqcap \forall region.HeadAndNeck \\
&CNS \,\dot{\sqsubseteq}\, NervousSystem \\
&BodyPart \,\dot{\sqsubseteq}\, AnatLocation
\end{aligned}$$

How can logic help?

$$\begin{aligned}
\text{Brain} &\sqsubseteq \exists\text{systemPart.NervousSystem} \sqcap \\
&\quad \text{CentralNervousS} \sqcap \text{BodyPart} \sqcap \\
&\quad \exists\text{region.HeadAndNeck} \sqcap \forall\text{region.HeadAndNeck} \\
\text{AnatLocation} &\sqsubseteq \text{ROOT} \sqcap \exists\text{region.TOP} \sqcap \forall\text{region.TOP} \\
\text{BodyPart} &\sqsubseteq \text{AnatLocation} \\
\text{CentralNervousS} &\sqsubseteq \text{NervousSystem} \\
\text{NervousSystem} &\sqsubseteq \text{AnatLocation} \\
&(\text{DISJOINT BodyPart Skeletton BloodSystem NervousSystem}) \\
\text{Cerebrum} &\sqsubseteq \exists\text{systempart.NervousSystem} \sqcap \\
&\quad \exists\text{haspart.Brain} \sqcap \text{BodyPart} \sqcap \ldots
\end{aligned}$$

Is my brain satisfiable?

Is my brain satisfiable?

$$\begin{aligned}
\text{Brain} &\sqsubseteq \exists\text{systemPart.NervousSystem} \sqcap \\
&\quad \text{CentralNervousS} \sqcap \text{BodyPart} \sqcap \\
&\quad \exists\text{region.HeadAndNeck} \sqcap \forall\text{region.HeadAndNeck} \\
\text{AnatLocation} &\sqsubseteq \text{ROOT} \sqcap \exists\text{region.TOP} \sqcap \forall\text{region.TOP} \\
\text{BodyPart} &\sqsubseteq \text{AnatLocation} \\
\text{CentralNervousS} &\sqsubseteq \text{NervousSystem} \\
\text{NervousSystem} &\sqsubseteq \text{AnatLocation} \\
&(\text{DISJOINT BodyPart Skeletton BloodSystem NervousSystem}) \\
\text{Cerebrum} &\sqsubseteq \exists\text{systempart.NervousSystem} \sqcap \\
&\quad \exists\text{haspart.Brain} \sqcap \text{BodyPart} \sqcap \ldots
\end{aligned}$$

Is my brain satisfiable?

$$\begin{aligned}
\text{Brain} &\sqsubseteq \exists\text{systemPart.NervousSystem} \sqcap \\
&\quad \text{CentralNervousS} \sqcap \text{BodyPart} \sqcap \\
&\quad \exists\text{region.HeadAndNeck} \sqcap \forall\text{region.HeadAndNeck} \\
\text{AnatLocation} &\sqsubseteq \text{ROOT} \sqcap \exists\text{region.TOP} \sqcap \forall\text{region.TOP} \\
\text{BodyPart} &\sqsubseteq \text{AnatLocation} \\
\text{CentralNervousS} &\sqsubseteq \text{NervousSystem} \\
\text{NervousSystem} &\sqsubseteq \text{AnatLocation} \\
&(\text{DISJOINT BodyPart Skeletton BloodSystem NervousSystem}) \\
\text{Cerebrum} &\sqsubseteq \exists\text{systempart.NervousSystem} \sqcap \\
&\quad \exists\text{haspart.Brain} \sqcap \text{BodyPart} \sqcap \ldots
\end{aligned}$$

Is my brain satisfiable?

$$\begin{aligned}
\text{Brain} &\sqsubseteq \exists\text{systemPart.NervousSystem} \sqcap \\
&\quad \text{CentralNervousS} \sqcap \text{BodyPart} \sqcap \\
&\quad \exists\text{region.HeadAndNeck} \sqcap \forall\text{region.HeadAndNeck} \\
\text{AnatLocation} &\sqsubseteq \text{ROOT} \sqcap \exists\text{region.TOP} \sqcap \forall\text{region.TOP} \\
\text{BodyPart} &\sqsubseteq \text{AnatLocation} \\
\text{CentralNervousS} &\sqsubseteq \text{NervousSystem} \\
\text{NervousSystem} &\sqsubseteq \text{AnatLocation} \\
&(\text{DISJOINT BodyPart Skeletton BloodSystem NervousSystem}) \\
\text{Cerebrum} &\sqsubseteq \exists\text{systempart.NervousSystem} \sqcap \\
&\quad \exists\text{haspart.Brain} \sqcap \text{BodyPart} \sqcap \ldots
\end{aligned}$$

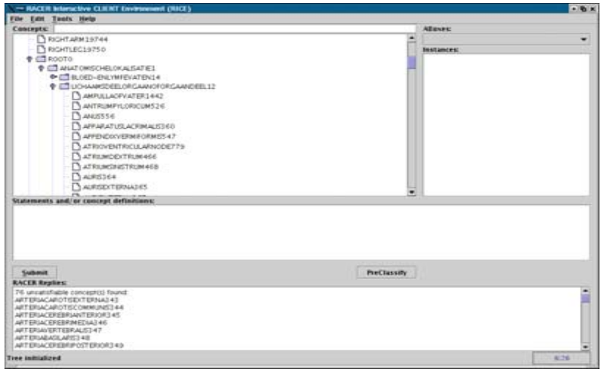
Logic proves that concept *Brain* is unsatisfiable.

Is my brain satisfiable?

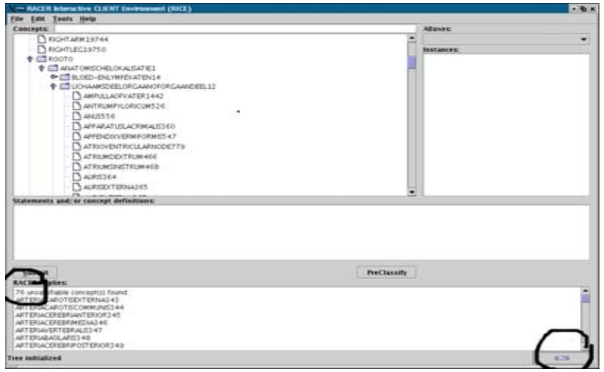
$$\begin{aligned}
\text{Brain} &\sqsubseteq \exists\text{systemPart.NervousSystem} \sqcap \\
&\quad \text{CentralNervousS} \sqcap \text{BodyPart} \sqcap \\
&\quad \exists\text{region.HeadAndNeck} \sqcap \forall\text{region.HeadAndNeck} \\
\text{AnatLocation} &\sqsubseteq \text{ROOT} \sqcap \exists\text{region.TOP} \sqcap \forall\text{region.TOP} \\
\text{BodyPart} &\sqsubseteq \text{AnatLocation} \\
\text{CentralNervousS} &\sqsubseteq \text{NervousSystem} \\
\text{NervousSystem} &\sqsubseteq \text{AnatLocation} \\
&(\text{DISJOINT BodyPart Skeletton BloodSystem NervousSystem}) \\
\text{Cerebrum} &\sqsubseteq \exists\text{systempart.NervousSystem} \sqcap \\
&\quad \exists\text{haspart.Brain} \sqcap \text{BodyPart} \sqcap \ldots
\end{aligned}$$

Logic proves that concept *Cerebrum* is unsatisfiable. Easy?

Modelling with RICE



Modelling with RICE



The Debugging Problem

- Checking the DICE anatomy with RACER:
  - many, unordered unsatisfiable concepts, long runtime for reasoning
  - Therefore: Trial and Error doesn't work
  - debugging is tedious, time-consuming, and error prone
- Two (classical diagnostic) solutions: detect
  - minimal conflicts - MIPS (debugging)
  - minimal fixes (diagnoses)

Reminder: Terminological Reasoning

$$\mathcal{T} = \{ Painting \sqsubseteq Artwork \sqcap \neg Sculpture, \quad Painter \sqsubseteq \exists creates.Paintings, \quad Sculpturer \sqsubseteq \exists creates.Artwork \sqcap \forall creates.Sculpture \}$$

- *Concept satisfiability*, $\Sigma \models C \neq \bot$.
  - Check whether there is an interpretation $\mathcal{I}$ such that $\mathcal{I} \models \Sigma$ and $C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$.
  - *Concept unsatisfiability:* $\Sigma \models Painter \sqcap Sculpturer = \bot$.
- *Incoherence*:
  - Check whether there is unsatisfiable concepts

Why is my brain inconsistent?

Brain ⊑ ∃systemPart.NervousSystem ⊓ CentralNervousS ⊓ BodyPart ⊓ ∃region.HeadAndNeck ⊓ ∀region.HeadAndNeck
AnatLocation ⊑ ROOT ⊓ ∃region.TOP ⊓ ∀region.TOP
BodyPart ⊑ AnatLocation
CentralNervousS ⊑ NervousSystem
NervousSystem ⊑ AnatLocation
(DISJOINT BodyPart Skeletton BloodSystem NervousSystem)
Cerebrum ⊑ ∃systempart.NervousSystem ⊓ ∃haspart.Brain ⊓ BodyPart ⊓ . . .

Debugging: Minimal conflict sets

Brain ⊑ ∃systemPart.NervousSystem ⊓ CentralNervousS ⊓ BodyPart ⊓ ∃region.HeadAndNeck ⊓ ∀region.HeadAndNeck
AnatLocation ⊑ ROOT ⊓ ∃region.TOP ⊓ ∀region.TOP
BodyPart ⊑ AnatLocation
CentralNervousS ⊑ NervousSystem
NervousSystem ⊑ AnatLocation
(DISJOINT BodyPart Skeletton BloodSystem NervousSystem)
Cerebrum ⊑ ∃systempart.NervousSystem ⊓ ∃haspart.Brain ⊓ BodyPart ⊓ . . .

Minimal Incoherence Preserving Sub-terminologies are not unique!

Diagnoses: an orthogonal notion

Brain	$\sqsubseteq$	∃systemPart.NervousSystem ⊓ ∃is_part_of CentralNervousS ⊓ BodyPart ⊓ ∃region.HeadAndNeck ⊓ ∀region.HeadAndNeck
AnatLocation	$\sqsubseteq$	ROOT ⊓ ∃region.TOP ⊓ ∀region.TOP
BodyPart	$\sqsubseteq$	AnatLocation
CentralNervousS	$\sqsubseteq$	NervousSystem
NervousSystem	$\sqsubseteq$	AnatLocation
(DISJOINT BodyPart Skeletton BloodSystem NervousSystem)		
Cerebrum	$\sqsubseteq$	∃systempart.NervousSystem ⊓ ∃haspart.Brain ⊓ BodyPart ⊓ . . .

Again, diagnoses are not unique!

Another diagnosis

Brain	$\sqsubseteq$	∃systemPart.NervousSystem ⊓ CentralNervousS ⊓ BodyPart ⊓ ∃region.HeadAndNeck ⊓ ∀region.HeadAndNeck
AnatLocation	$\sqsubseteq$	ROOT ⊓ ∃region.TOP ⊓ ∀region.TOP
BodyPart	$\sqsubseteq$	AnatLocation
CentralNervousS	$\sqsubseteq$	NervousSystem
NervousSystem	$\sqsubseteq$	AnatLocation
(DISJOINT BodyPart Skeletton BloodSystem NervousSystem)		
Cerebrum	$\sqsubseteq$	∃systempart.NervousSystem ⊓ ∃haspart.Brain ⊓ BodyPart ⊓ . . .

Minimal Conflict Sets and Diagnosis

Let $\mathcal{T}$ be an incoherent terminology. A TBox $\mathcal{T}' \subseteq \mathcal{T}$ is a

- minimal conflict set for incoherence of $\mathcal{T}$ iff
  - $\mathcal{T}'$ is incoherent, and
  - every sub-TBox $\mathcal{T}'' \subset \mathcal{T}'$ is coherent.
- a diagnosis for the incoherence problem of $\mathcal{T}$ iff,
  - $\mathcal{T} \setminus \mathcal{T}'$ is coherent, and
  - $\mathcal{T} \setminus \mathcal{T}''$ is incoherent for every sub-TBox $\mathcal{T}'' \subset \mathcal{T}'$

Definitions are independent of language

Algorithm for minimal conflict sets

A naive bottom-up algorithm.

- Build smallest incoherent sub-terminology of a terminology $\mathcal{T}$, by:
  1. $i = 1$.
  2. Check all sub-terminologies $\mathcal{T}' \subseteq \mathcal{T}$ of size i
  3. If $\mathcal{T}'$ is coherent, go back to (2).
- If terminated $\mathcal{T}'$ is a minimal conflict set.
- Advantage: optimised tools can be used to check for coherence. Debugging is possible in any language that the prover can understand.
- Disadvantage: can be very slow.

An example

(1)	Brain	$\sqsubseteq$	Bodypart ⊓ CentralNervousS
(2)	AnatLocation	$\sqsubseteq$	ROOT ⊓ ∃region.TOP ⊓ ∀region.TOP
(4)	CentralNervousS	$\sqsubseteq$	NervousSystem
(5)	(DISJOINT BodyPart Skeletton BloodSystem NervousSystem)		
(6)	Cerebrum	$\sqsubseteq$	∃systempart.NervousSystem ⊓ . . .

Check terminologies:
$\{1\}, \{2\}, \dots, \{6\} \sim 6$ checks
$\{1,2\}, \{1,3\}, \dots, \{1,6\}, \{2,3\}, \dots, \{2,6\}, \dots \sim 15$ checks
$\{1,2,3\}, \dots, \{1,5,6\}, \{2,3,4\}, \dots \sim 30$ checks
$\{1,2,3,4\}, \dots, \{1,2,3,6\}, \{1,2,4,5\} \sim$
87 checks in total
around 2.000.000.000 checks for DICE (to find MIPS of length 3)

Vocabulary induced ordering

(1)	Brain	$\sqsubseteq$	Bodypart ⊓ CentralNervousS
(2)	AnatLocation	$\sqsubseteq$	ROOT ⊓ ∃region.TOP ⊓ ∀region.TOP
(4)	CentralNervousS	$\sqsubseteq$	NervousSystem
(5)	(DISJOINT BodyPart Skeletton BloodSystem NervousSystem)		
(6)	Cerebrum	$\sqsubseteq$	∃systempart.NervousSystem ⊓ . . .

Check terminologies:
$\{1\}, \{2\}, \dots, \{6\} \sim 6$ checks
$\{1,4\}, \{1,5\}, \{4,5\}, \{4,6\}, \{5,6\} \sim 5$ checks
$\{1,4,5\}, \{1,5,6\}, \{4,5,6\}, \dots$

Much more efficient, but incomplete

Algorithm for Axiom Pinpointing (for unfoldable TBoxes)

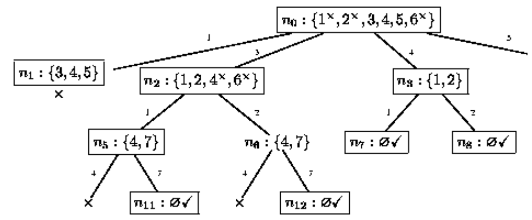
- For each unsatisfiable concept A
  - Fully expand a labelled tableau (starting $(a : A)^{\varnothing}$)
  - Labels are sets of axioms
  - Labels added while lazy-unfolding
  - Collect labels of contradicting atoms
  - This gives a propositional minimisation formula
  - Boolean Minimisation gives Minimal Unsatisfiability Preserving Sub-TBox (MUPS) for the TBox and the concept A
- Combine these MUPS to the set of MIPS.
- Status: Implemented and evaluated on DICE

Three ways of calculating diagnoses

- Small conflict sets
  - derived from the clashes in a tableau proof.
  - **Property:** generic prover needs to be adapted.
- Maximal conflict sets:
  - use generic DL reasoner to return a conflict sets
  - for an incoherent TBox $\mathcal{T}$ is to return $\mathcal{T}$ itself
  - **Property:** any language that the prover understands.
- Minimal conflict sets
  - using the algorithms of IJCAI 2003 to calculate minimal conflict sets.
  - **Property:** Specialized algorithms for unfoldable $\mathcal{ALC}$ TBoxes.

Hitting tree algorithm for diagnosis

Reiter proposes an algorithm to calculate diagnosis from conflict sets using hitting trees.



Still, the problem is NP-complete

An alternative Strategy for Pinpointing

Look at the following MIPS for the DICE Anatomy

{BRAINS292 CENTRAALZENUWSTELSEL151 ZENUWSTELSEL141}
{BRAINS292 CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12}
{CENTRAALZENUWSTELSEL151 MENINGES302 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12 MENINGES302}
{CENTRAALZENUWSTELSEL151 VENTRICULUSTERTIUS337 LICHAAMSDEELORGAANOFORGAANDEEL12}
{CENTRAALZENUWSTELSEL151 VENTRICULUSTERTIUS337 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12 VENTRICULUSQUARTUS339}
{CENTRAALZENUWSTELSEL151 VENTRICULUSQUARTUS339 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 MEDULLAESPINALISCERVICALIS399 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12 MEDULLAESPINALISCERVICALIS399}
{CENTRAALZENUWSTELSEL151 MEDULLAESPINALISTHORACOLUMBALIS400 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 MEDULLAESPINALISTHORACOLUMBALIS400 LICHAAMSDEELORGAANOFORGAANDEEL12}
{LICHAAMSDEELORGAANOFORGAANDEEL12 MUSCULIPAPILLARES463 SPIEREN17}
{LICHAAMSDEELORGAANOFORGAANDEEL12 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}
{BLOED-ENLYMFEVATEN14 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}

A Strategy for Pinpointing

Look at the following MIPS for the DICE Anatomy

{BRAINS292 CENTRAALZENUWSTELSEL151 ZENUWSTELSEL141}
{BRAINS292 CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12}
{CENTRAALZENUWSTELSEL151 MENINGES302 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12 MENINGES302}
{CENTRAALZENUWSTELSEL151 VENTRICULUSTERTIUS337 LICHAAMSDEELORGAANOFORGAANDEEL12}
{CENTRAALZENUWSTELSEL151 VENTRICULUSTERTIUS337 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12 VENTRICULUSQUARTUS339}
{CENTRAALZENUWSTELSEL151 VENTRICULUSQUARTUS339 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 MEDULLAESPINALISCERVICALIS399 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 LICHAAMSDEELORGAANOFORGAANDEEL12 MEDULLAESPINALISCERVICALIS399}
{CENTRAALZENUWSTELSEL151 MEDULLAESPINALISTHORACOLUMBALIS400 ZENUWSTELSEL141}
{CENTRAALZENUWSTELSEL151 MEDULLAESPINALISTHORACOLUMBALIS400 LICHAAMSDEELORGAANOFORGAANDEEL12}
{LICHAAMSDEELORGAANOFORGAANDEEL12 MUSCULIPAPILLARES463 SPIEREN17}
{LICHAAMSDEELORGAANOFORGAANDEEL12 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}
{BLOED-ENLYMFEVATEN14 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}

A Strategy for Pinpointing

What remains after deleting

{LICHAAMSDEELORGAANOFORGAANDEEL12 MUSCULIPAPILLARES463 SPIEREN17}
{LICHAAMSDEELORGAANOFORGAANDEEL12 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}
{BLOED-ENLYMFEVATEN14 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}

Again, what remains after deleting:

{BLOED-ENLYMFEVATEN14 THYMUS514 LYMFEVATENENLYMFEKLIEREN7}

So we have a *Pinpoint*:

{CENTRAALZENUWSTELSEL151,LICHAAMSDEELORGAANOFORGAANDEEL12,THYMUS514}

Which means: Taking away these three axioms turns terminology coherent!

Summary

- Explanation is necessary
- Explaining incoherence by minimal conflict sets
- Diagnoses are more useful for reasoning with inconsistency, because they give rise to maximally coherent sub-terminologies.
- both can be calculated
  - bottom-up methods for arbitrary languages (incomplete)
  - top-down methods for $\mathcal{ALC}$ (complete)
  - but both with high complexity
- Approximations are necessary, such as the pinpoints, which are not necessarily minimal diagnoses.

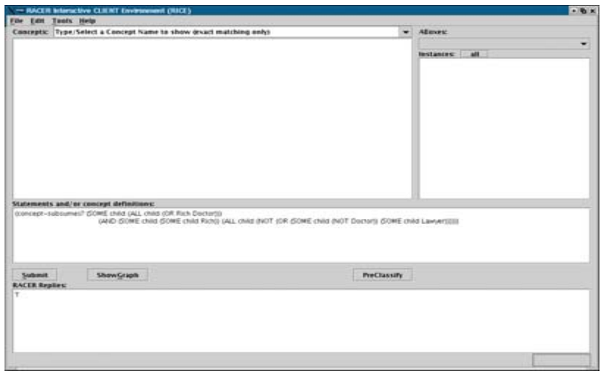
Reasoning to support the modeling process

- Logical modeling
  - formal correctness
  - formal conciseness
  - formal completeness
- Explaining logical errors
- Explaining subsumption
  - Explaining concept subsumption
  - Explaining terminological subsumption
- Modularization of Ontologies

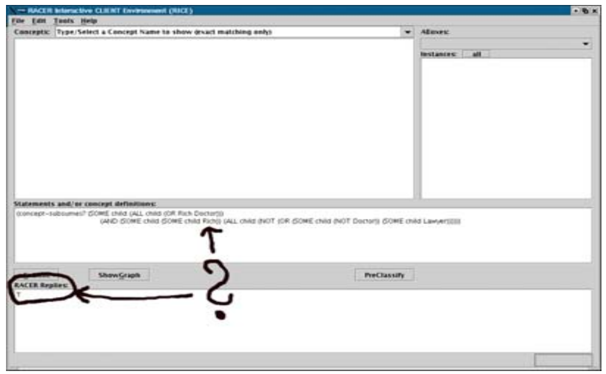
Explaining Subsumption as Proof Fragment

- McGuinness PhD 96, Borgida et.al 2000, ...
- Take: *Suppose somebody has rich grandchildren, and each of his children has neither a child which is not a doctor nor a child which is a Lawyer. Then, this person must have a child, every child of which is either rich or a doctor.*
- Consider the following subsumption $C_{ex} \sqsubseteq D_{ex}$.
  - $\exists child.\exists child.Rich \sqcap \forall child.\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer)) \sqsubseteq \exists child.\forall child\ (Rich \sqcup Doctor).$

Modelling and Querying with RICE (RACER)



Modelling and Querying with RICE (RACER)



Explaining Subsumption as Proof Fragment (McGuinness PhD 96, Borgida et.al 2000)

- $\exists child.\exists child.Rich \sqcap$
  $\forall child.\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer))$
  $\sqsubseteq \exists child.\forall child\ (Rich \sqcup Doctor).$
- $\exists child.Rich \sqcap$
  $\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer))$
  $\sqsubseteq \forall child\ (Rich \sqcup Doctor).$

Explaining Subsumption as Proof Fragment (McGuinness PhD 96, Borgida et.al 2000)

- $\exists child.\exists child.Rich \sqcap$
  $\forall child.\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer))$
  $\sqsubseteq \exists child.\forall child\ (Rich \sqcup Doctor).$
- $\exists child.Rich \sqcap$
  $\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer))$
  $\sqsubseteq \forall child\ (Rich \sqcup Doctor).$
- $\exists child.Rich \sqcap$
  $\neg(\exists child.\neg Doctor) \sqcap \neg(\exists child.Lawyer)$
  $\sqsubseteq \forall child\ (Rich \sqcup Doctor).$

Explaining Subsumption as Proof Fragment (McGuinness PhD 96, Borgida et.al 2000)

- $\exists child.\exists child.Rich \sqcap$
  $\forall child.\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer))$
  $\sqsubseteq \exists child.\forall child\ (Rich \sqcup Doctor).$
- $\exists child.Rich \sqcap$
  $\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer))$
  $\sqsubseteq \forall child\ (Rich \sqcup Doctor).$
- $\exists child.Rich \sqcap$
  $\neg(\exists child.\neg Doctor) \sqcap \neg(\exists child.Lawyer)$
  $\sqsubseteq \forall child\ (Rich \sqcup Doctor).$
- $\neg\neg Doctor \sqcap \neg Lawyer \sqsubseteq (Rich \sqcup Doctor).$

Explanation by Illustration

- Suppose somebody has rich grandchildren, and each of his children has neither a child which is not a doctor nor a child which is a Lawyer. Then, this person must have a child, every child of which is either rich or a doctor.

Explanation by Illustration

- Suppose somebody has rich grandchildren, and each of his children has neither a child which is not a doctor nor a child which is a Lawyer. Then, this person must have a child, every child of which is either rich or a doctor.

Explanation by Illustration

- Suppose somebody has rich grandchildren, and each of his children has neither a child which is not a doctor nor a child which is a Lawyer. Then, this person must have a child, every child of which is either rich or a doctor.

Explanation by Illustration

- Suppose somebody has rich grandchildren, and each of his children has neither a child which is not a doctor nor a child which is a Lawyer. Then, this person must have a child, every child of which is either rich or a doctor.
- This person must have a child, every child of which is a doctor.

Explanation by Illustration

- Suppose somebody has rich grandchildren, and each of his children has neither a child which is not a doctor nor a child which is a Lawyer. Then, this person must have a child, every child of which is either rich or a doctor.
- This person must have a child, every child of which is a doctor.
- $\exists child.\exists child.Rich \sqcap \forall child.\neg((\exists child.\neg Doctor) \sqcup (\exists child.Lawyer)) \sqsubseteq \exists child.\forall child.(Rich \sqcup Doctor)$.
- Illustration: $\exists child.\forall child.Doctor$
- Only works because $LHS \sqsubseteq Ill$ and $Ill \sqsubseteq RHS$

Explaining $C \sqsubseteq D$ through Illustrations

- Properties of an Illustration I
  - Simplify the subsumption: $C \sqsubseteq I$, $I \sqsubseteq D$.
  - Collect the common vocabulary from C and D.
- Illustrations are Interpolants
  - Definition of common vocabulary gives control
  - Ordering interpolants gives an explanation strategy.

The common language: control

- Common language w.r.t.
  - Polarity
  - Quantifier structure
- Formally:
  - $\mathcal{L}(A) := (A,+)^\epsilon \cup \mathcal{L}(\top)$ if A is atomic
  - $\mathcal{L}(C \sqcap D) := \mathcal{L}(C) \cup \mathcal{L}(D)$
  - $\mathcal{L}(C \sqcup D) := \mathcal{L}(C) \cup \mathcal{L}(D)$
  - $\mathcal{L}(\exists r.C) = \mathcal{L}(\forall r.C) := \mathcal{L}(C)^r \cup \mathcal{L}(\top)^\epsilon$
  - $\mathcal{L}(\neg C) := \mathcal{L}_{\tilde{\mathcal{A}}\mathcal{L}\mathcal{C}}(C)$
  - $\mathcal{L}(\top) = \mathcal{L}(\bot) := \{(\top,\_)^\epsilon, (\bot,\_)^\epsilon\}$
- Example:
  - $\forall r.(D \sqcap \neg\exists s.C)$.
  - $\{(C,-)^{rs}, (D,+)^r, (\top,\_)^\epsilon, (\top,\_)^r, (\top,\_)^{rs}, (\bot,\_)^\epsilon, (\bot,\_)^r,$

The common language: control

- Common language w.r.t.
  - Polarity
  - Quantifier structure
- Formally:
  - $\mathcal{L}(A) := (A,+)^\epsilon \cup \mathcal{L}(\top)$ if A is atomic
  - $\mathcal{L}(C \sqcap D) := \mathcal{L}(C) \cup \mathcal{L}(D)$
  - $\mathcal{L}(C \sqcup D) := \mathcal{L}(C) \cup \mathcal{L}(D)$
  - $\mathcal{L}(\exists r.C) = \mathcal{L}(\forall r.C) := \mathcal{L}(C)^r \cup \mathcal{L}(\top)^\epsilon$
  - $\mathcal{L}(\neg C) := \mathcal{L}_{\tilde{\mathcal{A}}\mathcal{L}\mathcal{C}}(C)$
  - $\mathcal{L}(\top) = \mathcal{L}(\bot) := \{(\top,\_)^\epsilon, (\bot,\_)^\epsilon\}$
- Example:
  - $\forall r.(D \sqcap \neg\exists s.C)$.
  - $\{(C,-)^{rs}, (D,+)^r, (\top,\_)^\epsilon, (\top,\_)^r, (\top,\_)^{rs}, (\bot,\_)^\epsilon, (\bot,\_)^r,$

The common language: control

- Common language w.r.t.
  - Polarity
  - Quantifier structure
- Formally:
  - $\mathcal{L}(A) := (A,+)^\epsilon \cup \mathcal{L}(\top)$ if A is atomic
  - $\mathcal{L}(C \sqcap D) := \mathcal{L}(C) \cup \mathcal{L}(D)$
  - $\mathcal{L}(C \sqcup D) := \mathcal{L}(C) \cup \mathcal{L}(D)$
  - $\mathcal{L}(\exists r.C) = \mathcal{L}(\forall r.C) := \mathcal{L}(C)^r \cup \mathcal{L}(\top)^\epsilon$
  - $\mathcal{L}(\neg C) := \mathcal{L}_{\tilde{\mathcal{A}}\mathcal{L}\mathcal{C}}(C)$
  - $\mathcal{L}(\top) = \mathcal{L}(\bot) := \{(\top,\_)^\epsilon, (\bot,\_)^\epsilon\}$
- Example:
  - $\forall r.(D \sqcap \neg\exists s.C)$.
  - $\{(C,-)^{rs}, (D,+)^r, (\top,\_)^\epsilon, (\top,\_)^r, (\top,\_)^{rs}, (\bot,\_)^\epsilon, (\bot,\_)^r, ($

Criteria for Optimality

- Syntactic
  - Minimal size
  - Minimal number of role-names
  - Minimal number of concept
- Semantic
  - Most specific
  - Most general
- Here I will focus on
  - concept-name-optimal
  - some thoughts about size and generality

Optimal Interpolants

Definition: Let P and N be concepts, and let $\mathcal{N}(C)$ denote the set of concept-names occurring in an arbitrary concept C. A concept I is an *optimal interpolant* for P and N if

- $\models P \sqsubseteq I$ and $\models I \sqsubseteq N$,
- $\mathcal{L}(I) \subseteq \mathcal{L}(P) \cap \mathcal{L}(N)$, and if
- there is no interpolant I' for P and N with $\mathcal{N}(I') \subset \mathcal{N}(I)$.

Algorithms for Description Logic $\mathcal{ALC}$

- To calculate optimal interpolant for C and D:
- Fully expand labeled tableau
  - labels are either $(\cdot)^p$ or $(\cdot)^n$
  - begin with one branch $\{(i : C)^p, (i : \neg D)^n\}$
  - fully saturate all branches
- Calculate Reducts
  - calculate reduct function ϕ
  - calculate prime implicants of ϕ.
- Calculate Optimal Interpolants
  - read interpolants from tableau

A simple Propositional Example

$A \sqcap B \sqcap C \sqsubseteq (A \sqcup B \sqcup \neg C) \sqcap (A \sqcup \neg B \sqcup C)$

- Tableau proof by refutation:

A		
B		A
C	both branches	B
$\neg A$	are closed	$\neg A$
$\neg B$		B
C		$\neg C$

- Reduct function: $red(\mathcal{T}) = (A \vee B) \wedge (A \vee C)$.
- Reducts: $\{A\}$, $\{B, C\}$
- Optimal Interpolants: A or $B \sqcap C$.
- A is an Illustration for the subsumption relation, and so is $B \sqcap C$

Explaining terminological subsumption

- The above methods work for concept subsumption
- How do we explain terminological subsumption?
  - Reduction to a debugging problem.
  - Remember that $\Sigma \models C \sqsubseteq D$ iff $\Sigma \models C \sqcap \neg D = \bot$ (day 2)
  - every minimal conflict set explains the subsumption, because it is the smallest set of axioms responsible for it to hold.

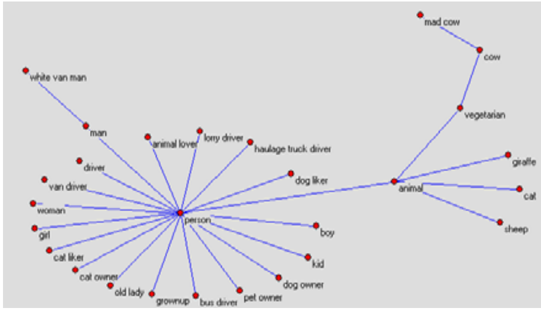
Reasoning to support the modeling process

- Logical modeling
  - formal correctness
  - formal conciseness
  - formal completeness
- Explaining logical errors
  - Debugging
  - Diagnosis
- Explaining subsumption
  - Explaining concept subsumption
  - Explaining terminological subsumption
- Modularization of Ontologies

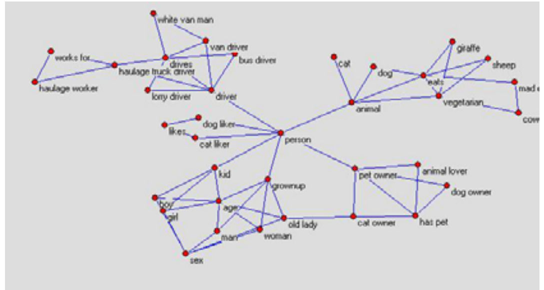
Modularizing Ontologies

- Often ontologies are too big to be used in SW applications
- Ideally one would like to import only parts, so-called modules.
- Here's a sketch of a method to modularise a given ontology:
  - partition the vocabulary
  - calculate the parts of the ontology which are relevant to keep the semantics of the concepts you want to use in your application

Partitioning Vocabularies



Partitioning Vocabularies



Semantic preserving transformations

Let $\mathcal{S}$ be a set of queries of the form $C \sqsubseteq D$ (subsumption) or $a : C$ (instance). An ontology $\mathcal{O}_\mathcal{S}$ is called a Semantic Preserving Transformation of an ontology $\mathcal{O}$ w.r.t. $\mathcal{S}$ if

$$\mathcal{O} \models \phi \text{ iff } \mathcal{O}_\mathcal{S} \models \phi \text{ for all } \phi \in \mathcal{S}$$

A strong module is a subset of the ontology, which is semantics preserving w.r.t. to its own vocabulary.

For a vocabulary V and a DL $\mathcal{L}$, let $\mathcal{L}_V$ be the set of concepts constructed from atoms in V only. An ontology $\mathcal{O}_{\mathcal{L}_V}$ is then a weak module iff it is semantic preserving w.r.t. the queries build from $\mathcal{L}_V$.

Back to our previous example

$Hepatitis \sqsubseteq \exists tract.DigestiveS \sqcap \forall tract.DigestiveS \sqcap \exists located.Liver \sqcap LiverD$
$\qquad \sqcap \exists abnormality.(Infarct \sqcup Drug \sqcup Inflammation \sqcup Infection \sqcup \ldots)$
$InfHepatitis \doteq Hepatitis \sqcap \exists abnormality.Infection \sqcap \forall abnormality.Infection$
$\qquad \sqcap \forall etiology.(EBVirus \sqcup HVirus \sqcup Amoeba \sqcup Spirochaetes)$
$VirHep \doteq \forall etiology.(HVirus \sqcup EB\text{-}Virus \sqcup CytomegaloVirus) \sqcap InfHepatitis$
$LiverD \sqsubseteq \forall tract.DigestiveS \sqcap \exists tract.DigestiveS \sqcap Diagnosis \sqcap \exists located.Liver$
$Diagnosis \sqsubseteq ReasonForAdmission$

$Liver \sqsubseteq \exists part.DigestiveS \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap \forall region.Abdomen$
$Kidney \sqsubseteq \exists part.GUSystem \sqcap \exists part.ESystem \sqcap BodyPart \sqcap \exists region.Abdomen \sqcap$
$\qquad \forall region.Abdomen \sqcap \forall side.(Left \sqcup Right)$

The last two axioms are probably a strong module, $Liver \sqsubseteq \top, Kidney \sqsubseteq \bot$ is a weak module w.r.t $V = \{Liver, Kidney\}$.

Caclulating Modules

Calculating modules is an open problem. Here are some thoughts:

- Strong modules:
  - Study syntactic dependencies
- Weak modules for an ontology $\mathcal{O}$:
  - Taking $\mathcal{O}$ itself is a trivial solution.
  - If the new ontology should be build from V only, this is in general not possible
  - Bottom-up methods
  - Tableau transformations
  - Quantifier elimination

Wrap-up

- Today
  - We have seen some examples for logical modeling support.
  - Often new techniques and thoughts.
  - Research in progress.
  - There's help needed, and lots of room for creativity.
- Tomorrow: Heiner