



W3C/IHTSDO 2009

Applications of SNOMED and DL in
Kaiser's EHR
Peter Hendler MD
Kaiser Permanente

Two Subjects

- How Does Kaiser Currently Use SNOMED
- How Can SNOMED Be Moved to OWL(SHIQ) Expressivity

The Three Layered Terminology

- There are three layers in Kaisers E.H.R. terminology system
- The providers (physicians) see the interface terminology
- It is specific to Kaiser and created by requests from our users
- This is mapped in the background to our Reference Terminology which is SNOMED
- The interface term is mapped to billing terminologies such as ICD and CPT

Interface terminology

KAISER PERMANENTE
HEALTHCONNECT.

- Display name heuristics developed for different areas of each masterfile
- Display names are created to be user friendly and are not necessarily the same as those created as FSNs to be submitted to SNOMED

EAP Laboratory Heuristic
KP HealthConnect
Convergent Medical Terminology
03/30/2007

For comments, please contact Michael T Montgomery (michael.Lmontgomery.kp.org)

1. Objective	2
2. Overview	2
3. Background	2
4. Context of Use in KP HealthConnect	2
5. CMT Guiding Principles	3
6. Strategy	5
7. Naming Conventions	5
8. Mapping Scenarios	5
9. Examples	5
10. Outstanding Issues	27
11. Open Questions	27
12. Reference Contacts	27

KAISER PERMANENTE
HEALTHCONNECT.

Example of interface terminology to administrative and reference terminology mapping

ICD-9 493, Asthma

SNOMED 57607007, Occupational asthma (disorder)

BRUCE GOLDBERG | © Kaiser Permanente | For Internal use only | page

SNOMED is EL+

- Multiple "Has Some" roles
- Multiple IsA (more than one parent allowed)
- Conjunction
- Can do Structural Subsumption.

EL+ can not do Negation or Disjunction

The screenshot displays the SNOMED CT browser interface for the concept 'TB - Pulmonary tuberculosis' (SNOMED ID: 154283005). The interface is divided into several panes:

- Top Panel:** Shows the concept name 'TB - Pulmonary tuberculosis' and its type 'clinical finding'. The parent concept is 'pulmonary tuberculosis' (SNOMED ID: 1784750013).
- Left Panel (Words):** Lists various terms related to pulmonary tuberculosis, with 'Pulmonary tuberculosis' highlighted in red.
- Left Panel (Subtype hierarchy):** Shows a list of subtypes under the 'Subtype hierarchy' view. The selected concept '154283005 pulmonary tuberculosis' is highlighted in red. Other subtypes include 'inactive tuberculosis of lung', 'infiltrative lung tuberculosis', 'isolated tracheal or bronchial tuberculosis', etc.
- Right Panel (Definition):** Provides a detailed definition of the concept. It includes:
 - Concept Status:** Current
 - Descriptions:** A list of descriptive terms such as 'pulmonary tuberculosis (disorder)', 'pulmonary tuberculosis', 'FTB - Pulmonary tuberculosis', and 'TB - Pulmonary tuberculosis'.
 - Definition:** Fully defined by ...
 - Is a:** A list of parent concepts including 'pneumonitis', 'inflammatory disorder of lower respiratory tract', 'disorder of lung', 'inflammation of specific body organs', 'tuberculosis', 'pulmonary disease due to Mycobacteria', 'infectious disease of lung', 'bacterial lower respiratory infection', and 'mycobacteriosis'.
 - causative agent:** Mycobacterium tuberculosis complex.
 - Group:** associated morphology.
 - finding site:** lung structure.
 - Qualifiers:** severity (severities), episodicity (episodicities), clinical course (courses).
 - Codes:** (Empty list)

Pulmonary Tuberculosis SNOMED Multiple Hierarchies

- It IsA pneumonitis which in turn IsA
 - Inflammatory disorder of lower respiratory tract
- It IsA Tuberculosis
- It IsA Pulmonary disease due to Mycobacteria

Pulmonary Tuberculosis SNOMED Relationships

- Has causative agent Mycobacterium tuberculosis complex
- Has associated morphology Granulomatous inflammation
- Has finding site Lung structure
- Overall a much richer structured model than single hierarchy with no relationships

Pulmonary TB

- A pure Subsumption approach
- It IsA Disorder
- It Has Finding Site Lung Structure
- It Has Causative Agent Mycobacterium

- Using Kaiser Query Tool to find Patients who have any form of Pulmonary TB

Query Model

Save CQML VCG/HCM Metadata Run Clear

Criteria Link | Export | Help

Query Details

Query Name

Query Desc

NI EDG Clinical (Diagnosis)

Load CQML Browse... Re-Load

Property Based Query

Value	300000	300010
cd-9 Code		
cd-9 Code		
cd-9 Code		
cd-9 Code		

Load more rows +

Hierarchical Based Query

SCTID

SCTID

SCTID

Load more rows +

Subsumption Based Query

Load concept with +

SA 64572001

Role

Causative agent (attribute)	243368001 +
Finding site (attribute)	39607008
-- Select Role Type --	

1 Total Results = 15

Exclude	CSM_ID	item_2	item_1	item_11	item_40	item_2000
<input type="checkbox"/>	105302	TB LUNG, BRONCHIECTASIS	26224	12126224	011.50A	011.50
<input type="checkbox"/>	105303	TB LUNG, FIBROSIS	26225	12126225	011.40A	011.40
<input type="checkbox"/>	105825	TB LUNG, CAVITARY	26747	12126747	011.20A	011.20
<input type="checkbox"/>	124660	TB LUNG, CULTURE DIAGNOSIS	27295	12127295	011.94A	011.94
<input type="checkbox"/>	127809	MYCOBACTERIAL PNEUMONIA, NON TB	28774	12128774	500725	500725
<input type="checkbox"/>	132112	TUBERCULOSIS (TB), PULMONARY, ACTIVE.	31139	12131139	011.90D	011.90
<input type="checkbox"/>	171105	PULMONARY MYCOBACTERIAL AVIUM	37507	12137507	031.0D	031.0
<input type="checkbox"/>	184096	TB LUNG, CONFIRMED HISTOLOGICALLY	39390	12139390	011.95A	011.95
<input type="checkbox"/>	189147	TB OF LUNG, INFILTRATIVE	39861	12139861	011.00A	011.00
<input type="checkbox"/>	189149	TB PNEUMONIA	39863	12139863	011.60A	011.60

Create Patient Cohorts

- Given this list of Kaiser Master File Diagnosis terms, we can go into our E.H.R. and bring back the MRNs of all patients that have one of these conditions that is subsumed by pulmonary TB.
- This can be used for Quality of Care or Reporting or even Decision Support
- But we can't do subsumption on terms that include negation or disjunction

Structural Subsumption in EL+

- Compare terms in Normalized form
- Terms can get roles in two ways
- It can have roles asserted and it can inherit roles
- In normal form a term is flattened to two lists
- Atomic Primitive Parents (with no roles)
- Flattened list of all roles inherited and asserted

Normalized TB Pneumonia

- 154283005pulmonary tuberculosis
[Defined]
- is a=disease
- ,causative agent=Mycobacterium tuberculosis complex
- {associated morphology=granulomatous inflammation
- ,finding site=lung structure}

Structural Subsumption

- If you have two normalized terms you can see if one is subsumed by the other
- If they have exactly the same form, they are equal.
- If one of them has everything the other does, plus one or more additional features, then it is more specialized and it is subsumed by the other
- If they share features but each has additional features, they may be siblings

But Kaiser Needs Negation and Disjunction

- Non infectious Pneumonia
- Non genital herpes
- Infectious or Malignant disorder of lung
- Burn injury of face neck or scalp
- Poisoning by antineoplastic AND/OR immunosuppressive drug

More Samples of KP terms rejected because of negation or disjunction

Non renal secondary hyperparathyroidism

Localized infection of skin AND/OR subcutaneous tissue

Warts non-genital

Sprain or strain of wrist

Closed fracture of skull or face with intracranial hemorrhage without coma



OWL Web Ontology Language

Because Owl in Winnie the Pooh spelled his name WOL.
If OWL can do that, then
Web Ontology Language can be spelled OWL

Most often is SHIQ

But adds two things Kaiser wants.
Negation and Disjunction

Up until now, OWL/SHIQ expressivity could only be
used with small ontologies of 20k or so terms in tools like
Protege

Complexity of reasoning in Description Logics
Note: the information here is (always) incomplete and **updated** often

Base description logic: *Attributive Language with Complements*
 $ALC ::= \perp \mid A \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \exists R.C \mid \forall R.C$

Concept constructors:

- \mathcal{F} - functionality²: $(\leq 1 R)$
- \mathcal{N} - (unqualified) number restrictions: $(\geq n R)$, $(\leq n R)$
- \mathcal{Q} - qualified number restrictions: $(\geq n R.C)$, $(\leq n R.C)$
- \mathcal{O} - nominals: $\{a\}$ or $\{a_1, \dots, a_n\}$ ("one-of" constructor)
- μ - least fixpoint operator: $\mu X.C$
- $R \sqsubseteq S$ - role-value-maps
- $f = g$ - agreement of functional role chains ("same-as")

TBox is internalized in extensions of *ALC/O*, see [76, Lemma 4.12], [54, p.3]

- Empty TBox
- Acyclic TBox ($A \sqsubseteq C$, A is a concept name; no cycles)
- General TBox ($C \sqsubseteq D$ for arbitrary concepts C and D)

Role constructors:

- I - role inverses: R^{-}
- \cap - role intersection²: $R \sqcap S$
- \cup - role union: $R \sqcup S$
- \neg - role complement: full
- \circ - role chain (composition): $R \circ S$
- $*$ - reflexive-transitive closure⁴: R^*
- id - concept identity: $id(C)$
- Forbid complex roles⁵ in number restrictions⁶

Role axioms (RBox):

- \mathcal{S} - Role transitivity: $\text{Trans}(R)$
- \mathcal{H} - Role hierarchy: $R \sqsubseteq S$
- \mathcal{R} - Complex role inclusions: $R \circ S \sqsubseteq R$, $R \circ S \sqsubseteq S$
- \mathcal{I} - some additional features

You have selected the Description Logic: **SHIQ($\cap, \cup, -(full)$)**
Complex roles in number restrictions are: forbidden

Reasoning problem	Complexity ⁸	Comments and references

Tableaux Reasoners

Once you add negation and/or disjunction to an ontology, you can no longer do structural subsumption.

You must go to Tableaux reasoners that use a different form of symbolic logic.

In a nutshell they test for the negative hypothesis.

If A is subsumed by B, then the only thing that can not be true is that you can't be an A and NOT a B.

If Chevy is subsumed by Vehicle, then the only thing that can't be true is that you are a Chevy and NOT a Vehicle

Tableaux reasoners

The reasoner has to prove that given all of the assertions (triplets) in the ontology, that this one assertion is incompatible.

This happens when all branches of the tableaux are closed with a contradiction such as A and NOT A in the same branch

Pellet and FaCT++ are free open source Tableaux classifiers and they come packaged with the free tool Protege

They apparently need to load the entire ontology into memory

<p>AND</p> <p>1) A and B x</p> <p>-----</p> <p>2) A</p> <p>3) B</p>	<p>NOT AND</p> <p>1) !(A and B) x</p> <p>-----</p> <p>2) !A 3) !B</p>
<p>OR</p> <p>1) A or B x</p> <p>-----</p> <p>2) A 3) B</p>	<p>NOT OR</p> <p>1) !(A or B) x</p> <p>-----</p> <p>2) !A</p> <p>3) !B</p>
<p>IMPLIES</p> <p>1) A -> B x</p> <p>-----</p> <p>2) !A 3) A</p> <p>4) B</p>	<p>NOT IMPLIES</p> <p>1) !(A -> B) x</p> <p>-----</p> <p>2) A</p> <p>3) !B</p>
<p>IFF</p> <p>1) A <=> B x</p> <p>-----</p> <p>2) A 3) !A</p> <p>4) B 5) !B</p>	<p>NOT IFF</p> <p>1) !(A <=> B) x</p> <p>-----</p> <p>2) !A 3) A</p> <p>4) B 5) !B</p>
<p>Complex statements can be broken up into s</p> <p>1) (A or (B or C)) <=> (D and (E -> F))</p>	

Protege

Using Protege you can easily produce a small version of a SNOMED like ontology and add negated and disjunctive terms.

But in a 32 bit machine you can only address 4 gigs of RAM and it turns out that is not enough to load SNOMED.

Well, on my XP laptop that's true, but MAC and Linux can do it on 32 bit machines with enough memory.

A 64 bit machine can address enough memory

A 64 bit machine can have much more memory but..
They cost a lot.

Amazon EC2 (Elastic Compute Cloud) is able to virtualize various machines which can be run for 40 to 80 cents an hour.

These virtual machines are intended to be desktopless web servers, but you can create an image of 64 bit Ubuntu Linux which Has a nice desktop

In order to use the Desktop remotely you must install VNC or even better NoMachine NX server

Methods Continued

Once you have a 64 bit machine with a desktop and 14 Gigs of memory you can install 64 bit JAVA and you can install Protege

The Protege start up configuration must be changed to specify more memory for the Java Heap.

It turns out that you need a **minimum JAVA heap size of 2000Megs for 64 bit OS**

SNOMED is distributed in 3 very large text files

Kent Spackman has written a PERL script that will convert this format to OWL RDF/XML which is the input format for Protege

The First Experiment

The first test was to see if the entire (the large inferred version) of SNOMED OWL/SHIQ could be loaded and navigated in Protege on a machine with 14 Gigs of RAM.

Happily it could, but it was indeed quite sluggish.

The next test was whether it could be classified.

The screenshot shows the Protege software interface with a terminal window open. The terminal displays system statistics and a table of running processes. A 'Reasoner progress' dialog box is overlaid on the Protege window, showing a progress bar at 56% and a 'Cancel' button. The Protege window shows the 'Inferred class hierarchy' for the ontology 'http://www.snomed.org/'.

NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
0	14.8g	4.1g	11m	S	100	27.0	13:05.25	java
0	18848	1248	928	R	0	0.0	0:00.53	top
0	3996	852	592	S	0	0.0	0:00.04	init
0	0	0	0	S	0	0.0	0:00.00	migration/0
19	0	0	0	S	0	0.0	0:00.00	ksoftirqd/0
0	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	0
19	0	0	0	S	0	0.0	0:00.00	0
-5	0	0	0	S	0	0.0	0:00.00	watchdog/1
-5	0	0	0	S	0	0.0	0:00.00	events/1
-5	0	0	0	S	0	0.0	0:00.00	migration/2
19	0	0	0	S	0	0.0	0:00.00	ksoftirqd/2

Resources Required

- Taking 37 min
- Using 100% CPU (it latched on to only one of the multiple CPUs)
- Using 5.5 Gigs of memory (It could get by with less)

The screenshot shows a terminal window and a Protege application window. The terminal window displays system statistics and a table of running processes. The Protege application window shows the 'Reasoner progress' dialog box, indicating that the ontology is being classified at 95%.

Terminal Output:

```

2 min, 3 users, load average: 1.00, 1.00, 0.96
1 running, 119 sleeping, 1 stopped, 1 zombie
.8%sy, 0.0%ni, 73.4%id, 0.1%wa, 0.0%hi, 0.0%si, 0.0%st
al, 5506064k used, 10230296k free, 16024k buffers
al, 0k used, 0k free, 495668k cached

```

NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
0	15.0g	4.2g	11m	S	100	28.2	37:29.12	java
0	86508	38m	6552	S	0	0.2	0:28.49	nxagent
0	18848	1248	928	R	0	0.0	0:01.81	top
0	3996	852	592	S	0	0.0	0:00.04	init
0	0	0	0	S	0	0.0	0:00.00	migration/0
19	0	0	0	S	0	0.0	0:00.00	migration/0
0	0	0	0	S	0	0.0	0:00.00	migration/0
-5	0	0	0	S	0	0.0	0:00.00	migration/0
-5	0	0	0	S	0	0.0	0:00.00	migration/0
-5	0	0	0	S	0	0.0	0:00.00	migration/0
-5	0	0	0	S	0	0.0	0:00.00	migration/0
-5	0	0	0	S	0	0.0	0:00.00	migration/0
-5	0	0	0	S	0	0.0	0:00.00	migration/0
19	0	0	0	S	0	0.0	0:00.00	ksortlrq0/1
-5	0	0	0	S	0	0.0	0:00.00	watchdog/1
-5	0	0	0	S	0	0.0	0:00.00	events/1
-5	0	0	0	S	0	0.0	0:00.00	migration/2

Reasoner progress dialog box:

```

Reasoner progress
Classifying ontology 95%
[Progress bar]
Cancel

```

Fully Classified

- Just under 40 minutes on High Mem machine with 14 gig RAM and 4 CPUs
- This is the inferred Class Hierarchy view
- Next slide shows fully classified inferred view of SNOMED CT
- Later research showed the classification is done on only one CPU and can be done in 2 gigs

The screenshot shows a terminal window on the left and a Protege application window on the right. The terminal window displays system statistics and a table of running processes. The Protege application window shows the inferred class hierarchy for SNOMED CT, with a tree view of classes and their relationships.

Terminal Output:

```

7 min, 3 users, load average: 0.22, 0.80, 0.91
1 running, 119 sleeping, 1 stopped, 1 zombie
.1%sy, 0.0%ni, 99.9%id, 0.0%wa, 0.0%hi, 0.0%si, 0.0%st
al, 5544440k used, 10191920k free, 16512k buffers
al, 0k used, 0k free, 495776k cached

```

NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
0	434m	38m	15m	S	0	0.3	0:01.70	nautilus
0	3996	852	592	S	0	0.0	0:00.04	init
0	0	0	0	S	0	0.0	0:00.00	migration/0
19	0	0	0	S	0	0.0	0:00.00	ksoftirqd/0
0	0	0	0	S	0	0.0	0:00.00	watchdog/0
-5	0	0	0	S	0	0.0	0:00.00	events/0
-5	0	0	0	S	0	0.0	0:00.00	khelper
-5	0	0	0	S	0	0.0	0:00.00	kthread
-5	0	0	0	S	0	0.0	0:00.00	xenwatch
-5	0	0	0	S	0	0.0	0:00.00	xenbus
-5	0	0	0	S	0	0.0	0:00.00	migration/1
19	0	0	0	S	0	0.0	0:00.00	ksoftirqd/1
-5	0	0	0	S	0	0.0	0:00.00	watchdog/1
-5	0	0	0	S	0	0.0	0:00.00	events/1
-5	0	0	0	S	0	0.0	0:00.00	migration/2
19	0	0	0	S	0	0.0	0:00.00	ksoftirqd/2
-5	0	0	0	S	0	0.0	0:00.00	watchdog/2

Protege Application Output:

```

changed
snomed.org/ from file:/protege/Protege_4.0_beta/snomed.owl
logy to http://www.snomed.org/
ndices...
ms
changed
owner for the SR0IQ(D) Description Logic
y V. Tsarkov, 2002-2008. Version 1.2.0 (25 September 2008)
n 1809161ms

```

Protege Application Class Hierarchy:

- Thing
 - Nothing
 - SCTID_138875005
 - SCTID_105590001
 - SCTID_106237007
 - SCTID_123037004
 - SCTID_123038009
 - SCTID_243796009
 - SCTID_254291000
 - SCTID_260787004
 - SCTID_272379006
 - SCTID_308916002
 - SCTID_362981000
 - SCTID_363787002
 - SCTID_370115009
 - SCTID_373873005
 - SCTID_404684003
 - SCTID_410607006
 - SCTID_419891008
 - SCTID_48176007
 - SCTID_71388002
 - SCTID_78621006

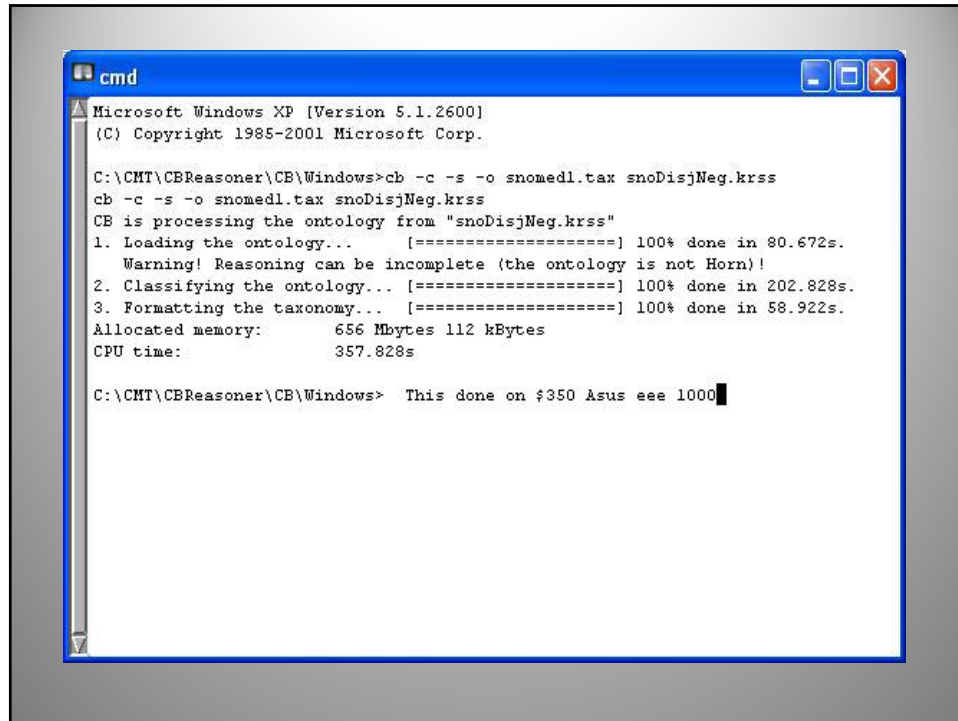
A Peek At a New Reasoner

- Using an entirely different flavor of reasoner, the resources needed could be reduced by orders of magnitude.
- In fact, a \$350 Asus EEE netbook was able to classify the test ontology in a few minutes.
- Is it complete? I can't test that myself, but the spot checks I did on the results were correct

The experimental CB reasoner

Ian Horrocks of Oxford is one of the experts in reasoner design for expressive ontologies like OWL

He and Yevgeny Kazakov (the author of CB) have created a forward chaining reasoner that classified all of asserted SNOMED in about 30 seconds. When the 50000 extra terms were added it took one minute.



```
cmd
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\CMT\CBReasoner\CB\Windows>cb -c -s -o snomedl.tax snoDisjNeg.krss
cb -c -s -o snomedl.tax snoDisjNeg.krss
CB is processing the ontology from "snoDisjNeg.krss"
1. Loading the ontology... [=====] 100% done in 80.672s.
   Warning! Reasoning can be incomplete (the ontology is not Horn)!
2. Classifying the ontology... [=====] 100% done in 202.828s.
3. Formatting the taxonomy... [=====] 100% done in 58.922s.
Allocated memory:      656 Mbytes 112 kBytes
CPU time:              357.828s

C:\CMT\CBReasoner\CB\Windows> This done on $350 Asus eee 1000
```

SNOMED OWL NEGATION

The Open World will get you if you don't
watch out

Modeled

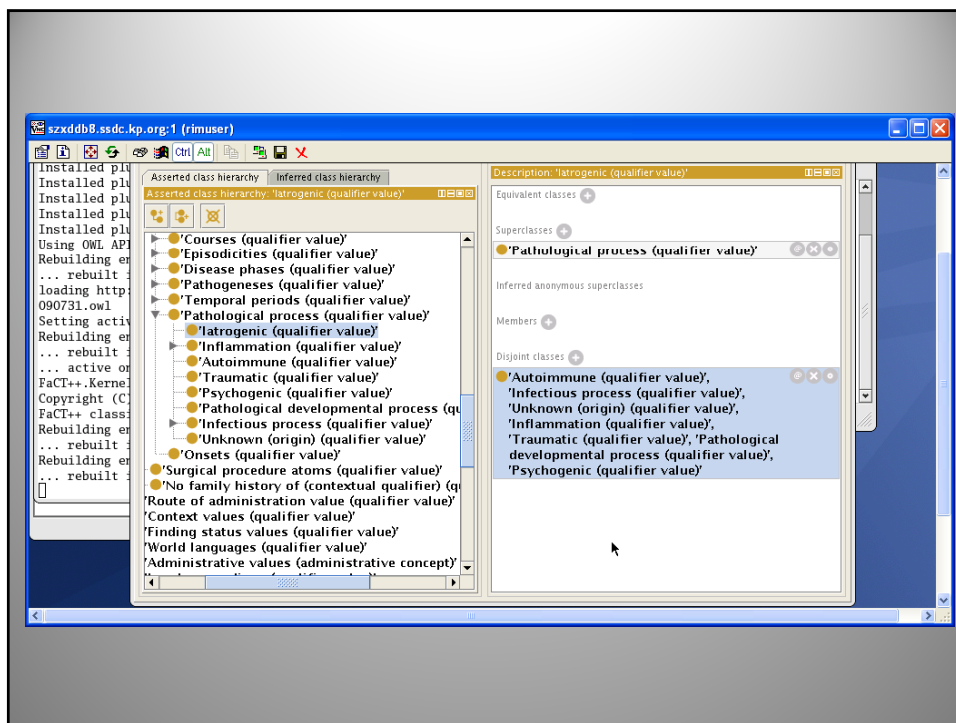
- All axioms are for fully defined terms
- Non infectious disease of lung
- `Isa Disease and roleGroup(findingSite some Lung and (not(infectiveDisease)))`

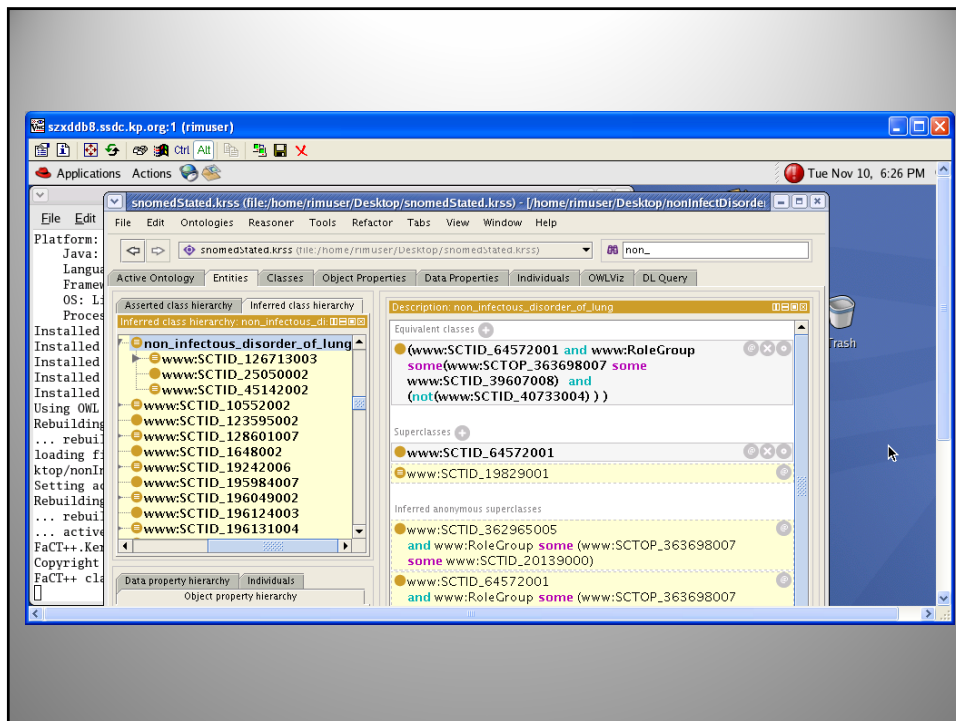
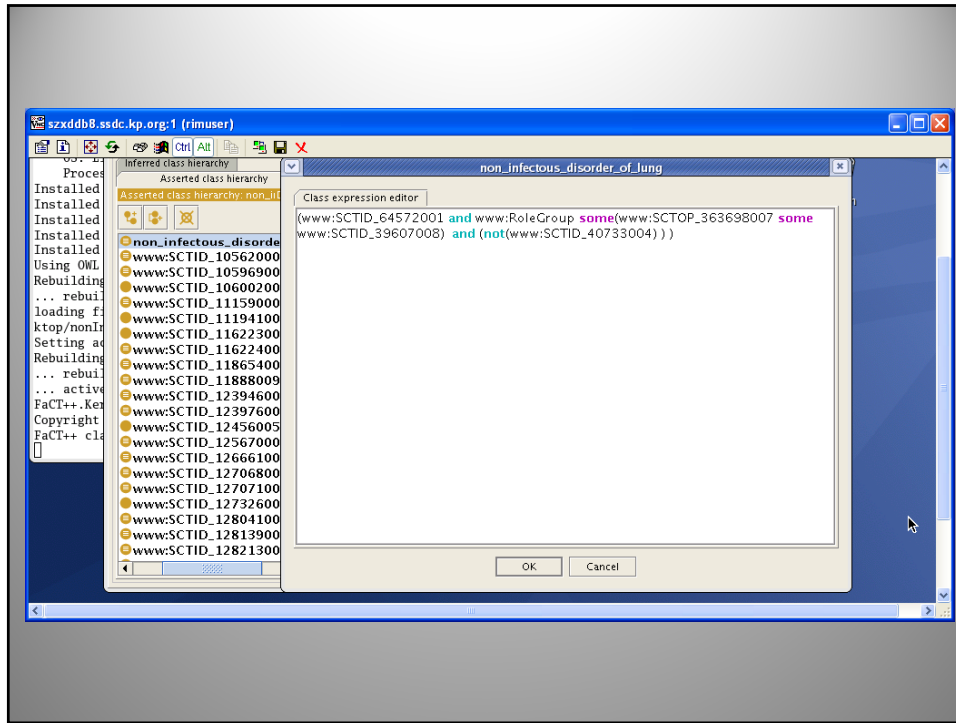
WARINING!!

Open World
Enter At Your Own
Risk

You must make disjoint

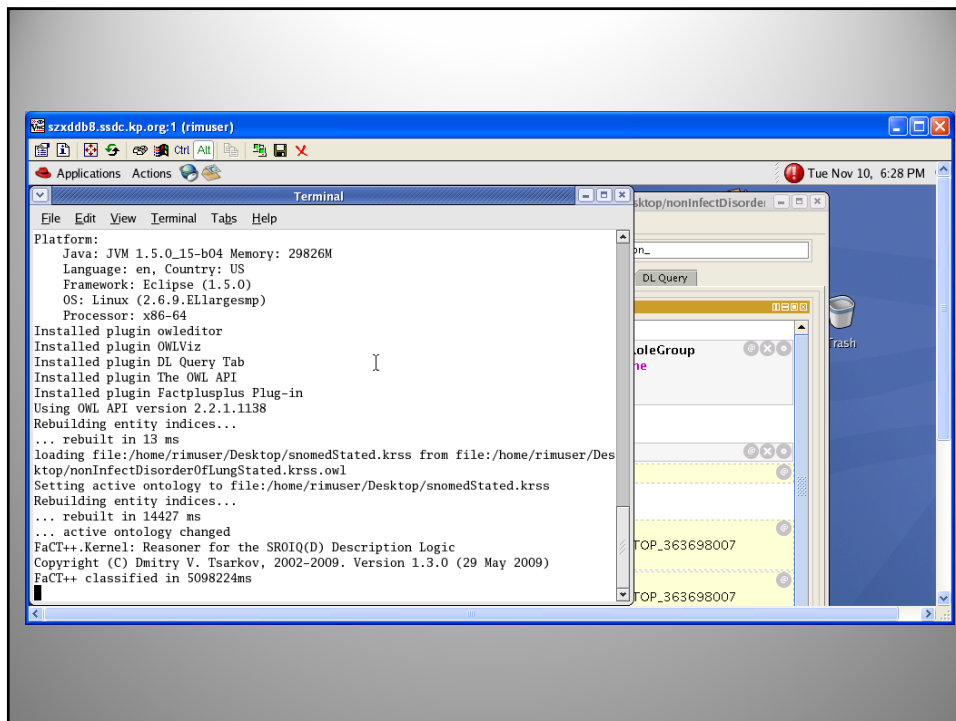
- In the open world, you can not assume that sibling classes are different or disjoint
- You can not assume that an Infectious disease is not a Neoplastic disease for example unless you explicitly assert that's the case
- Negation will not work unless disjoint classes are asserted





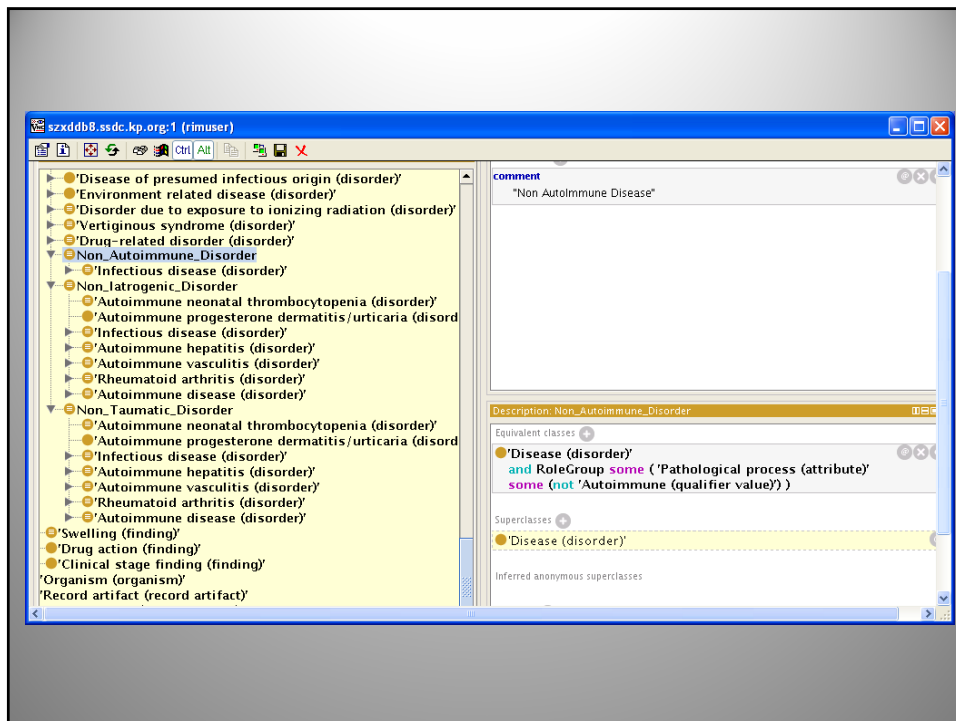
Classified

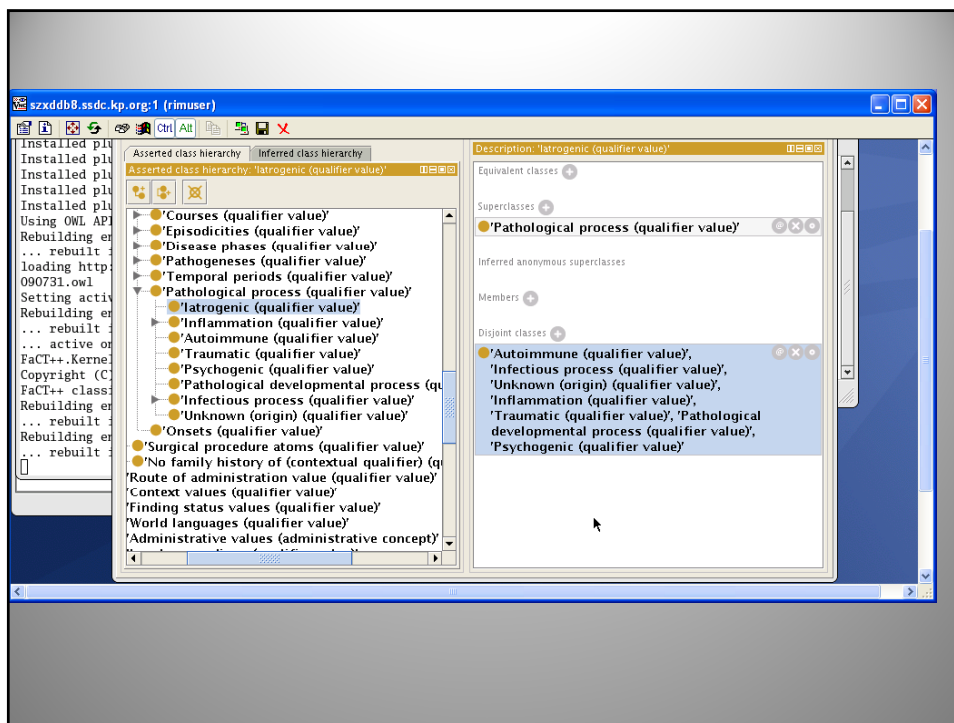
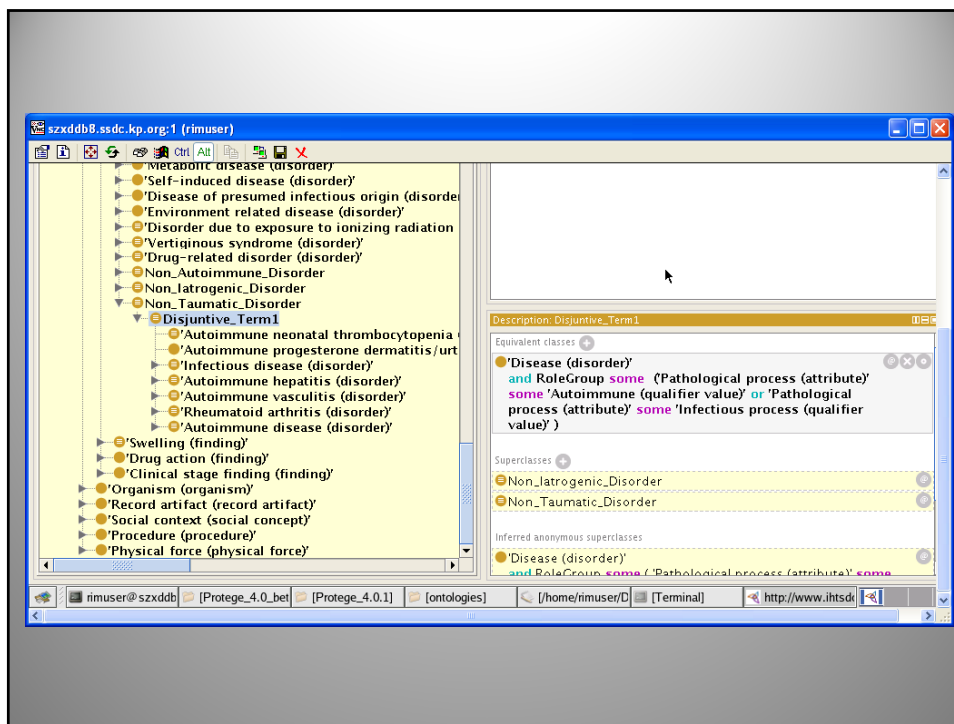
- 1.4 hours
- All children are correct
- Neoplasm of lung with many children
- Alpha heavy chain disease (respiratory form)
- Congenital pulmonary lymphangiectasis



Make it nice for humans

- It's easier to do the searching and setting up using the SCTID class names, but once you have results you can change Protégé to render the results by annotations which are human readable





The End

- Questions?