New Modular Reasoning Capabilities for SNOMED CT Classification Beyond the EL profile

Modular Reasoning – MORe (or Less)?

Robert Hausam MD

Hausam Consulting LLC

SNOMED CT Expo

31 October, 2019

Kuala Lumpur, Malaysia

The Issue

- Description logic underlying SNOMED CT is a subset of the OWL 2 EL profile
- The new DL capabilities that are available with the OWL refset still are within the EL profile
- This is good from the perspective of maintaining reasonable classification performance, but it limits the expressivity available for modeling concepts

Notably Excluded DL features

- Universal quantification to a class expression (**ObjectAllValuesFrom**)
- Disjunction (**ObjectUnionOf**, **DisjointUnion**, and **DataUnionOf**)
- Class negation (**ObjectComplementOf**)

SNOMED CT Logic Profile Specification

- Primary motivation for excluding these features and remaining within the OWL 2 EL profile is to ensure classifiers can reason over SNOMED CT content and expressions in a reasonable timeframe.
- While some of the features are desirable for content modelling, moving to OWL 2 DL incurs too high a cost to implementation of SNOMED CT to be supportable at this time.

SNOMED CT Logic Profile Specification

- If these features become reasonably practical to add in future (e.g. due to advancements in reasoner technologies) then they may be added.
- Until that time, modelling patterns must be devised to address the content and use cases which would benefit from these features to work around their absence.

Challenge

- Can we find and/or develop new reasoners or reasoning techniques that can make it feasible to use the “missing” OWL DL description logic capabilities in content modeling and still have tractable classification performance?
  - Can we remove the necessity to “work around” this absence?

Modular Reasoning

- Combines two (or more) DL reasoners in a way that takes advantage of their respective strengths
  - OWL EL reasoner
    - Fast, efficient performance
    - Examples: ELK, Snorocket
  - OWL DL reasoner
    - Significantly slower, but allows the additional expressive logic capabilities available in OWL DL (to varying degrees)
    - Examples: HermiT, FaCT++

MORe reasoner

- Developed by Ana Armas Romero, et al. at University of Oxford
- “Propose a novel technique where an efficient L-reasoner and a fully fledged OWL 2 reasoner are combined in a modular way to classify an OWL 2 ontology”
- Romero AA, Grau BC, Horrocks I. MORe: Modular Combination of OWL Reasoners for Ontology Classification. In Proceedings of the 11th International Semantic Web Conference (ISWC 2012). Springer. 2012.

Initial Work

- MORe 0.1.5 (or 0.1.6)
  - Latest distributed version
  - Based on earlier versions of ELK, HermiT and OWL API
  - Has been used previously with Snow Owl
  - Doesn’t work as a plugin with current versions of Protégé (and Snow Owl?)
  - Additional limitations
- This is not sufficient for work with the current versions of the tools, and the reasoner versions used don’t have all of the latest features

Updated MORe Reasoner

- Multiple updates to make it functional as a plugin for current versions of Protégé (5.1+) and other tools (Snow Owl)
- OWL API 4.2.6
- HermiT 1.3.8.413
- ELK 0.4.3
- FaCT++ 1.6.5 (uses native libraries)

Updated Ontology with Logical Negation

- Define “Non-bacterial infectious pneumonia”
- Add disjoint axioms between the “Bacteria” and “Virus” (and other) organism concept subhierarchies
- Add a universal restriction closure axiom to the definitions of the expected subtypes of “Non-bacterial infectious pneumonia”, including “Viral pneumonia”
  - Otherwise the reasoner doesn’t know that “Viral pneumonia” does not have additional unknown causative agents which may be a type of bacteria or other “non-viral” organism

Definition of “Non-bacterial infectious pneumonia”

SNOMED International

Description: 'Non-bacterial infectious pneumonia'

Equivalent To

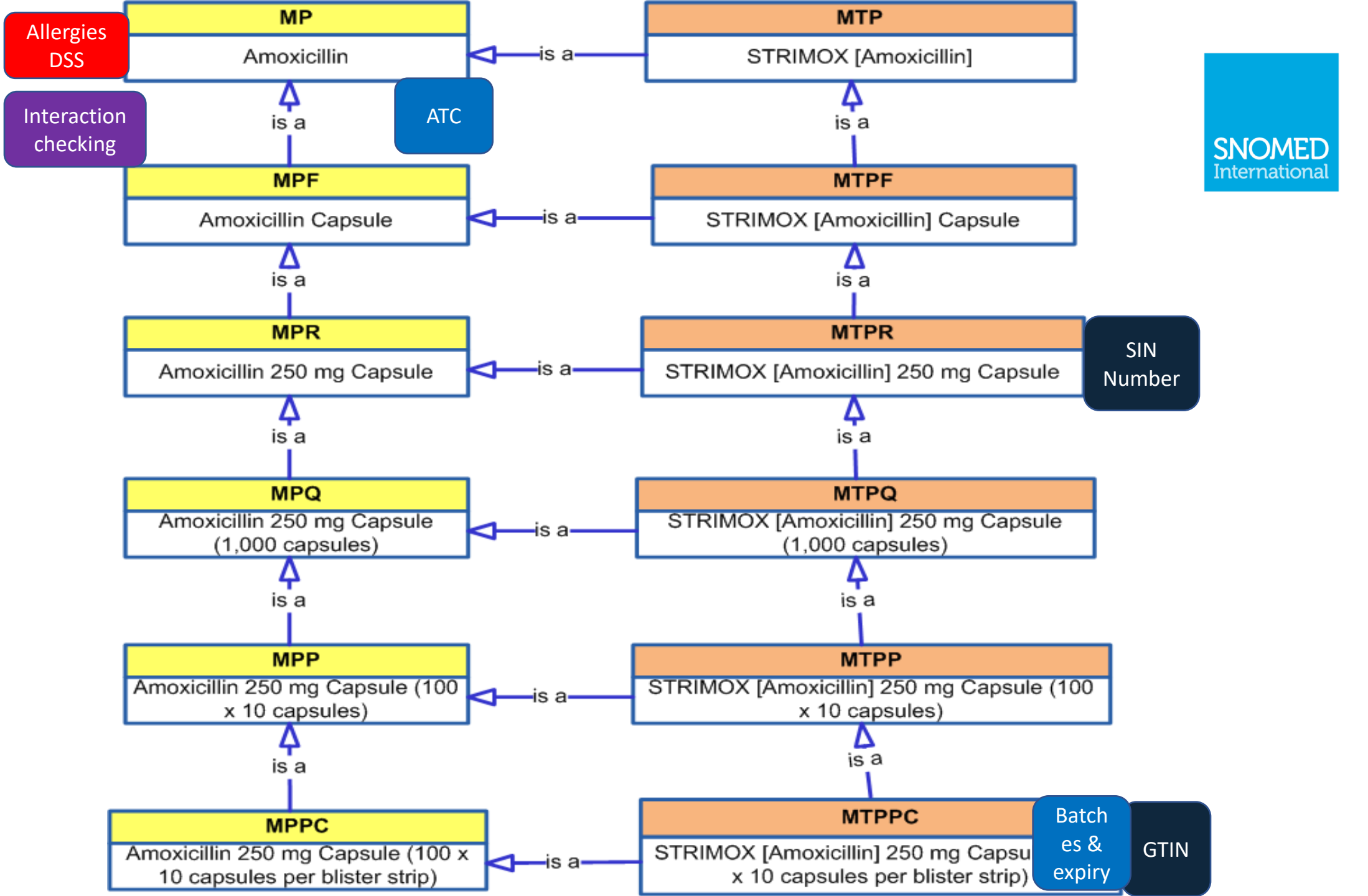
'Disease (disorder)'
and ('Role group (attribute)' some
((not ('Causative agent (attribute)' some
'Superkingdom Bacteria (organism)'))
and ('Associated morphology (attribute)'
some 'Inflammation and consolidation
(morphologic abnormality)')
and ('Causative agent (attribute)' some
'Microorganism (organism)')
and ('Finding site (attribute)' some 'Lung
structure (body structure)')
and ('Pathological process (attribute)'
some 'Infectious process (qualifier value)')))

Initial Results

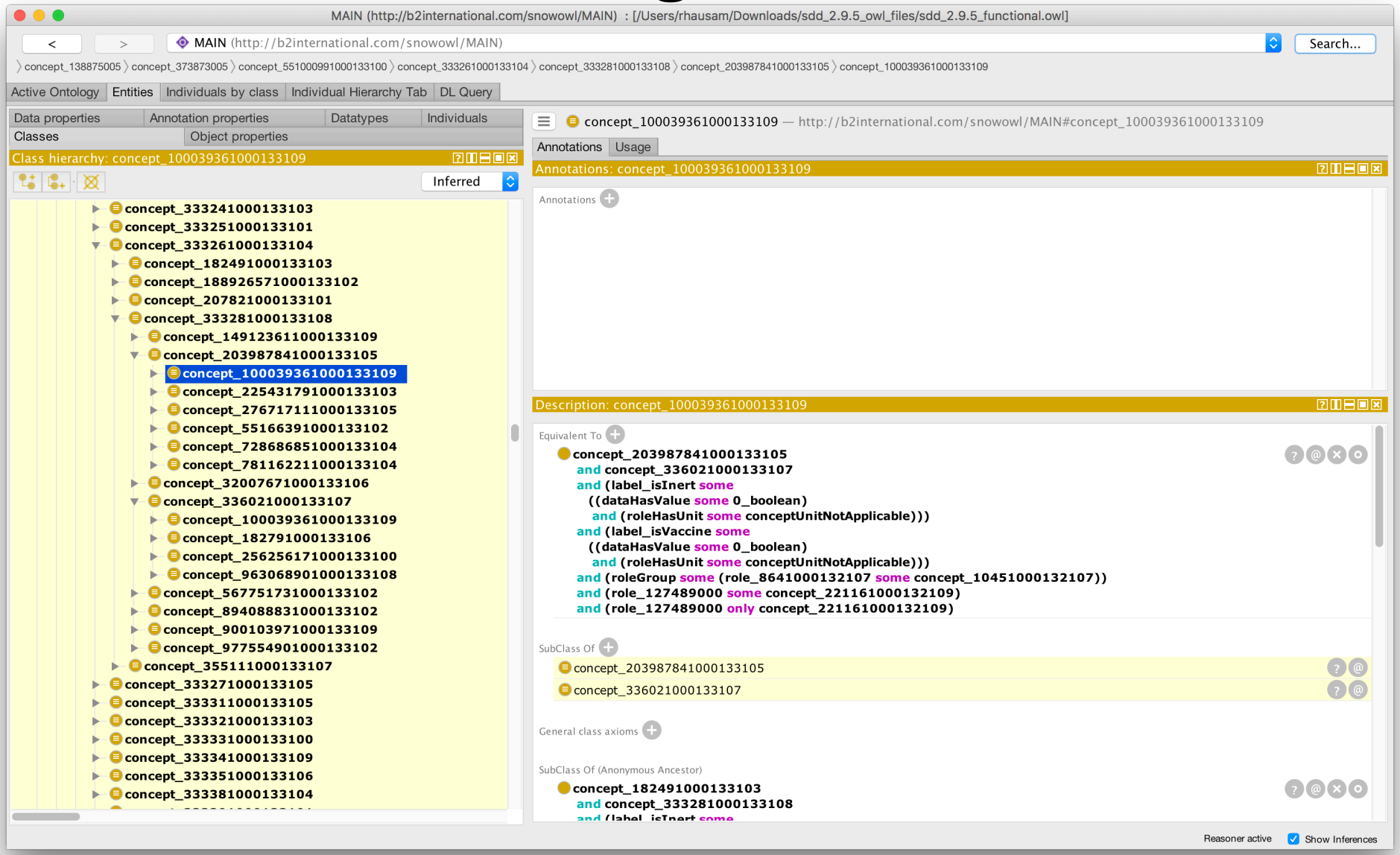
- Classification of the entire SNOMED CT ontology with the single additional negated concept of “Non-bacterial infectious pneumonia” and the supporting additional logic
  - MacBook Pro 2016 16 GB RAM
- Performed using the expressive OWL 2 reasoner FaCT++ in 11.9 minutes
- Performed using the updated MORe reasoner combining the ELK and FaCT++ OWL reasoners in 5.2 minutes
- A promising 56% reduction in classification time!

Further Testing

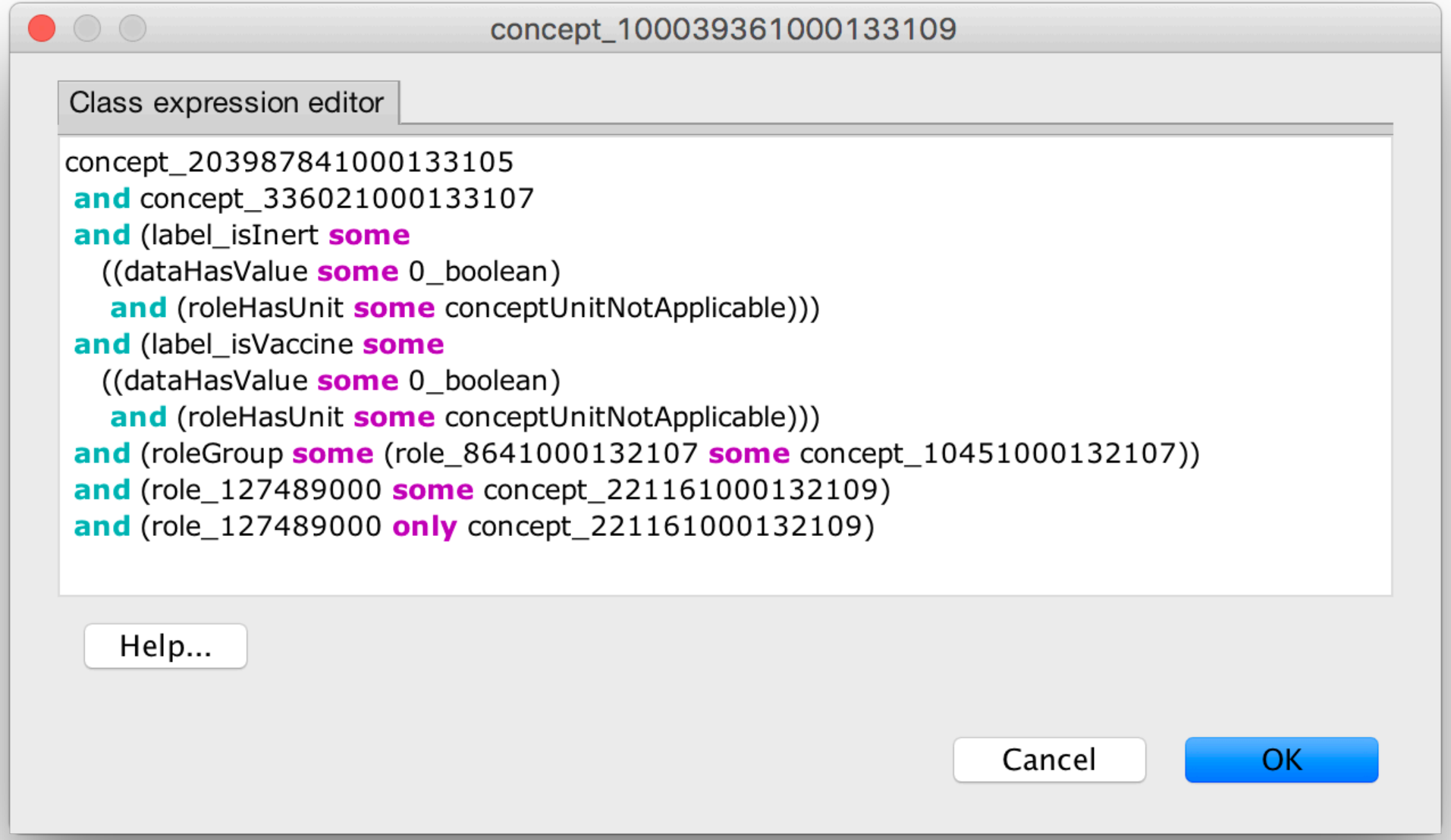
- Singapore Drug Dictionary (SDD)
  - Large numbers of universal restriction axioms
  - Known very long classification time with current reasoners
  - Initial external testing indicated a reduction in classification time from 14 to 7 hours using the MORe reasoner (50% reduction)


Allergies DSS
Interaction checking
ATC
SIN Number
Batches & expiry
GTIN
MP
Amoxicillin
MTP
STRIMOX [Amoxicillin]
MPF
Amoxicillin Capsule
MTPF
STRIMOX [Amoxicillin] Capsule
MPR
Amoxicillin 250 mg Capsule
MTPR
STRIMOX [Amoxicillin] 250 mg Capsule
MPQ
Amoxicillin 250 mg Capsule (1,000 capsules)
MTPQ
STRIMOX [Amoxicillin] 250 mg Capsule (1,000 capsules)
MPP
Amoxicillin 250 mg Capsule (100 x 10 capsules)
MTPP
STRIMOX [Amoxicillin] 250 mg Capsule (100 x 10 capsules)
MPPC
Amoxicillin 250 mg Capsule (100 x 10 capsules per blister strip)
MTPPC
STRIMOX [Amoxicillin] 250 mg Capsu
x 10 capsules per blister strip)
is a

SDD View in Protégé



SDD Class Expression Example



SDD Classification Results with Updated MORe Reasoner

- First upgrade the hardware (Amazon EC2 r5.4xlarge instance)
- FaCT++ reasoner alone: 2.626 hours
- Updated MORe Protégé plugin with ELK and FaCT++: 2.413 hours
- Only a disappointing 8.1% reduction! ☹
- But the upgraded hardware was able to significantly reduce classification time (2.6 hours is much better than 14!)

Next Step – MORe 2.0

SNOMED International

- This seems promising, as it:
  - Allows use of an improved module extraction algorithm, based on Datalog (logical query language) reasoning
  - Allows additional combinations of reasoners, including the fast Datalog / OWL 2 RL reasoner RDFox
- But there were issues
  - The available MORe 2.0 project is incompletely developed
    - Many reasoner methods were not supported, including getEquivalentClasses(), getSubClasses() and isConsistent()
    - getAllSuperClasses() and getAllUnsatisfiableClasses() were supported, but they aren't standard OWL API reasoner methods
  - Protégé plugin was partially developed, but not yet functional
    - getEquivalentClasses(), getSubClasses() and isConsistent() methods are required

MORe 2.0 Further Work

SNOMED International

- Completed initial implementation of getEquivalentClasses(), getSubClasses() and isConsistent()
- Updated OWL API and reasoner versions
- Completed functioning implementation of the Protégé plugin
  - For Protégé 5.5.0 (could also support earlier versions)

But Some Issues Arose ☹

- Bugs related to the module extraction logic ultimately prevented realizing the MORe 2.0 reasoner capabilities and being able to successfully classify useful ontologies
- So far have been unable to sufficiently resolve these issues
- At present, further work on the MORe 2.0 branch is postponed
  - Temporarily or permanently?

Back to the Future

- Go back to the earlier MORe branch (“0.1.7”)
  - See where to go from there
- Should still serve as a reasonable base for further development
- First reassess the current state
  - Updated ontology – new SNOMED CT release
  - Updated hardware – Amazon EC2 r5.4xlarge instance
    - Intel Xeon Platinum 8000 series (Skylake-SP) processors with a sustained all core Turbo CPU clock speed of up to 3.1 GHz
    - 16 vCPU
    - 128 GB RAM
  - Existing hardware
    - MacBook Pro late 2016 with 16 GB RAM

Current Results

- Use MORe FaCT++
- SNOMED CT July 2019 OWL file, generated by snomed-owl-toolkit
- Enhance with “Non-bacterial infective pneumonia”
  - Not quite as unrealistic as it seems, as relative numbers of “non EL” axioms should always be a small percentage, which is where modular reasoning is likely to perform best
  - Definition is updated from prior version, as some SNOMED CT concepts have been inactivated and replaced

Updated Ontology Stats

SNOMED International

- Axioms
  - Total ontology = 351,054
  - OWL 2 = 4
- Modules
  - EL signature = 271,676
  - OWL 2 = 117,957
  - Combined total of 389,633 (> 351,054)
  - OWL 2 module is 30.3% of total (from only 4 axioms!)

Prior Ontology Stats (2018)

- Axioms
  - Total ontology = 325,156
  - OWL 2 = 2
- Modules
  - EL signature = 250,192
  - OWL 2 = 107,021
  - Combined total of 357,213 (> 325,156)
  - OWL 2 module is 30.0% of total (from only 2 axioms!)

Classification Times - How Well Did It Do?

SNOMED International

2018 ontology – MacBook

- FaCT++ (alone)
  - 687,914 ms = 11.47 min
- MORe with FaCT++
  - 376,670 ms = 6.28 min (↓ 45.2%)

New ontology

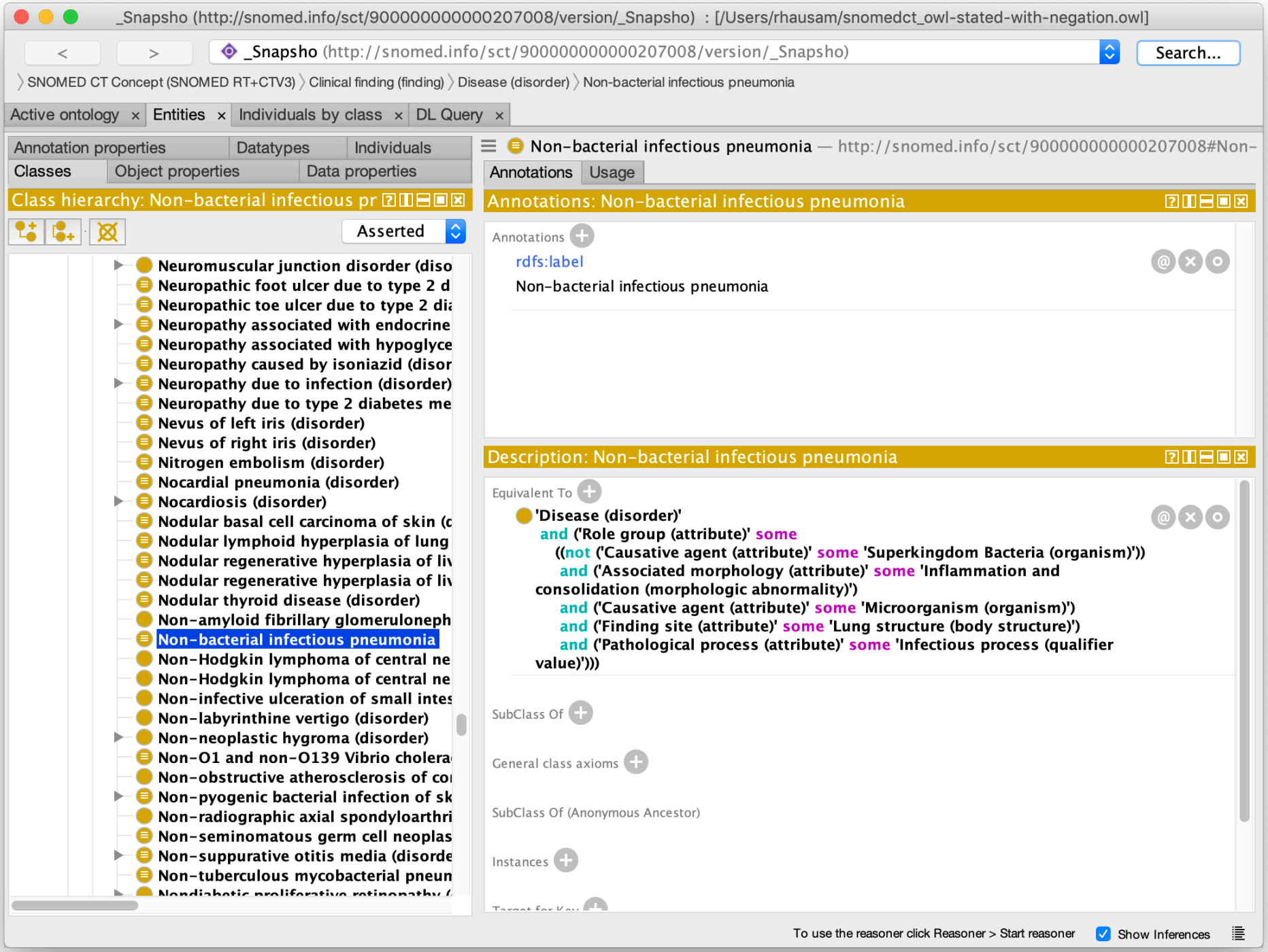
- FaCT++
  - *Way too long!*
- MORe with FaCT++
  - 1,699,151 ms = 28.32 min

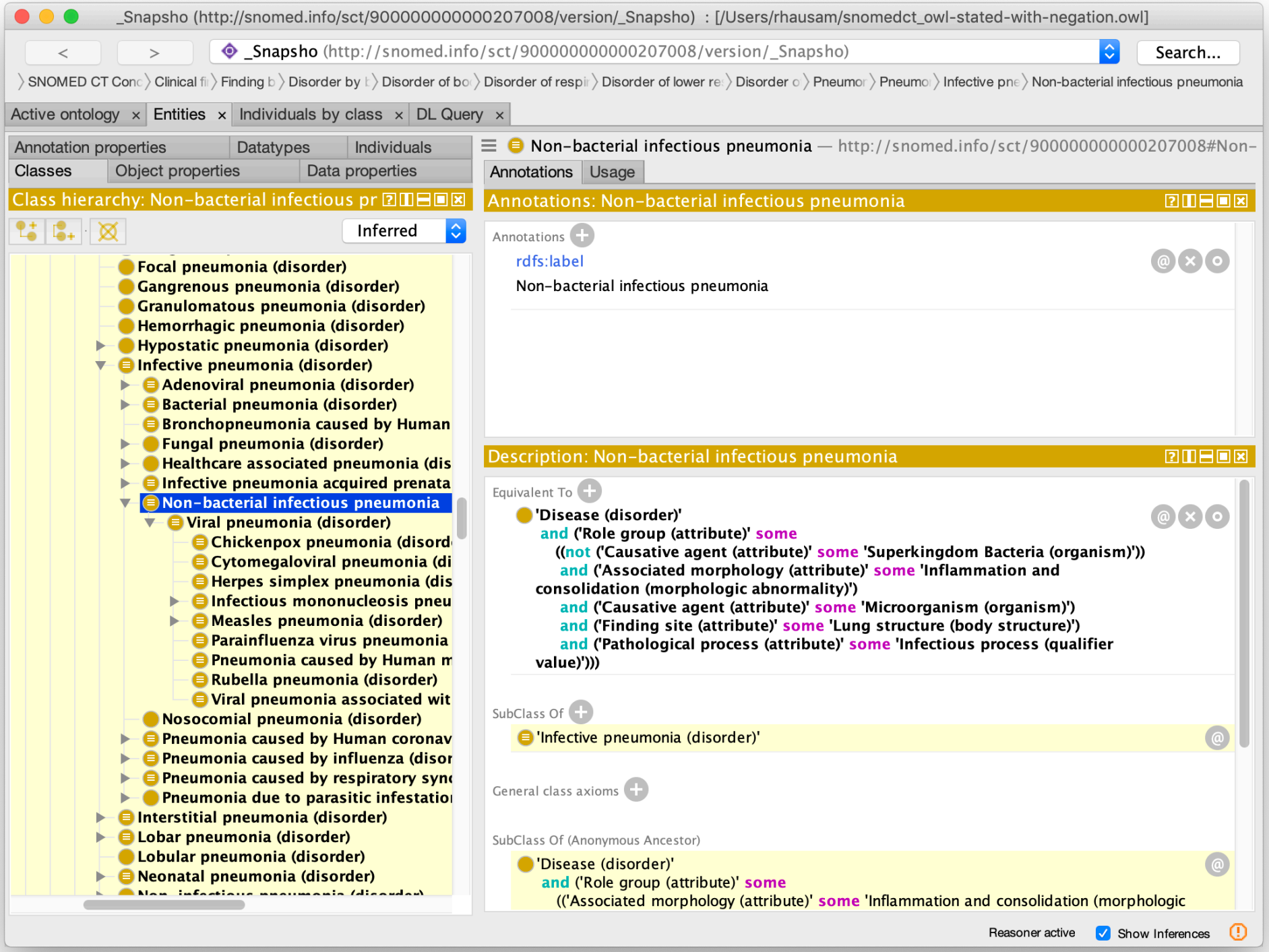
2018 ontology – Amazon EC2

- FaCT++ (alone)
  - 391,974 ms = 6.53 min
- MORe with FaCT++
  - 244,850 ms = min (↓ 37.5%)

New ontology

- FaCT++
  - *Way too long!*
- MORe with FaCT++
  - 911,362 ms = 15.19 min


_Snapsho (http://snomed.info/sct/900000000000207008/version/_Snapsho) : [/Users/rhausam/snomedct_owl-stated-with-negation.owl]
_Snapsho (http://snomed.info/sct/900000000000207008/version/_Snapsho)
Search...
SNOMED CT Concept (SNOMED RT+CTV3) 〉 Clinical finding (finding) 〉 Disease (disorder) 〉 Non-bacterial infectious pneumonia
Active ontology
Entities
Individuals by class
DL Query
Annotation properties
Datatypes
Individuals
Classes
Object properties
Data properties
Class hierarchy: Non-bacterial infectious pr
Asserted
Neuromuscular junction disorder (diso
Neuropathic foot ulcer due to type 2 d
Neuropathic toe ulcer due to type 2 dia
Neuropathy associated with endocrine
Neuropathy associated with hypoglyce
Neuropathy caused by isoniazid (disor
Neuropathy due to infection (disorder)
Neuropathy due to type 2 diabetes me
Nevus of left iris (disorder)
Nevus of right iris (disorder)
Nitrogen embolism (disorder)
Nocardial pneumonia (disorder)
Nocardiosis (disorder)
Nodular basal cell carcinoma of skin (c
Nodular lymphoid hyperplasia of lung
Nodular regenerative hyperplasia of liv
Nodular regenerative hyperplasia of liv
Nodular thyroid disease (disorder)
Non-amyloid fibrillary glomeruloneph
Non-bacterial infectious pneumonia
Non-Hodgkin lymphoma of central ne
Non-Hodgkin lymphoma of central ne
Non-infective ulceration of small intes
Non-labyrinthine vertigo (disorder)
Non-neoplastic hygroma (disorder)
Non-O1 and non-O139 Vibrio cholera
Non-obstructive atherosclerosis of co
Non-pyogenic bacterial infection of sk
Non-radiographic axial spondyloarthri
Non-seminomatous germ cell neoplas
Non-suppurative otitis media (disorde
Non-tuberculous mycobacterial pneun
Non-bacterial infectious pneumonia — http://snomed.info/sct/900000000000207008#Non-
Annotations
Usage
Annotations: Non-bacterial infectious pneumonia
Annotations
rdfs:label
Non-bacterial infectious pneumonia
Description: Non-bacterial infectious pneumonia
Equivalent To
'Disease (disorder)'
and ('Role group (attribute)' some
((not ('Causative agent (attribute)' some 'Superkingdom Bacteria (organism)'))
and ('Associated morphology (attribute)' some 'Inflammation and
consolidation (morphologic abnormality)')
and ('Causative agent (attribute)' some 'Microorganism (organism)')
and ('Finding site (attribute)' some 'Lung structure (body structure)')
and ('Pathological process (attribute)' some 'Infectious process (qualifier
value)')))
SubClass Of
General class axioms
SubClass Of (Anonymous Ancestor)
Instances
To use the reasoner click Reasoner > Start reasoner
Show Inferences


_Snapsho (http://snomed.info/sct/900000000000207008/version/_Snapsho) : [/Users/rhausam/snomedct_owl-stated-with-negation.owl]
_Snapsho (http://snomed.info/sct/900000000000207008/version/_Snapsho)
Search...
SNOMED CT Conc
Clinical fi
Finding b
Disorder by l
Disorder of bo
Disorder of respir
Disorder of lower re
Disorder o
Pneumor
Pneumor
Infective pne
Non-bacterial infectious pneumonia
Active ontology
Entities
Individuals by class
DL Query
Annotation properties
Datatypes
Individuals
Classes
Object properties
Data properties
Class hierarchy: Non-bacterial infectious pr
Inferred
Focal pneumonia (disorder)
Gangrenous pneumonia (disorder)
Granulomatous pneumonia (disorder)
Hemorrhagic pneumonia (disorder)
Hypostatic pneumonia (disorder)
Infective pneumonia (disorder)
Adenoviral pneumonia (disorder)
Bacterial pneumonia (disorder)
Bronchopneumonia caused by Human
Fungal pneumonia (disorder)
Healthcare associated pneumonia (dis
Infective pneumonia acquired prenata
Non-bacterial infectious pneumonia
Viral pneumonia (disorder)
Chickenpox pneumonia (disord
Cytomegaloviral pneumonia (di
Herpes simplex pneumonia (dis
Infectious mononucleosis pneu
Measles pneumonia (disorder)
Parainfluenza virus pneumonia
Pneumonia caused by Human m
Rubella pneumonia (disorder)
Viral pneumonia associated wit
Nosocomial pneumonia (disorder)
Pneumonia caused by Human coronav
Pneumonia caused by influenza (disor
Pneumonia caused by respiratory sync
Pneumonia due to parasitic infestatio
Interstitial pneumonia (disorder)
Lobar pneumonia (disorder)
Lobular pneumonia (disorder)
Neonatal pneumonia (disorder)
Non-bacterial infectious pneumonia — http://snomed.info/sct/900000000000207008#Non-
Annotations
Usage
Annotations: Non-bacterial infectious pneumonia
Annotations
rdfs:label
Non-bacterial infectious pneumonia
Description: Non-bacterial infectious pneumonia
Equivalent To
'Disease (disorder)'
and ('Role group (attribute)' some
((not ('Causative agent (attribute)' some 'Superkingdom Bacteria (organism)'))
and ('Associated morphology (attribute)' some 'Inflammation and
consolidation (morphologic abnormality)')
and ('Causative agent (attribute)' some 'Microorganism (organism)')
and ('Finding site (attribute)' some 'Lung structure (body structure)')
and ('Pathological process (attribute)' some 'Infectious process (qualifier
value)')))
SubClass Of
'Infective pneumonia (disorder)'
General class axioms
SubClass Of (Anonymous Ancestor)
'Disease (disorder)'
and ('Role group (attribute)' some
(('Associated morphology (attribute)' some 'Inflammation and consolidation (morphologic
Reasoner active
Show Inferences

Other Modular Reasoning Possibilities?

ComR (a combined OWL reasoner for ontology classification)

- From Tianjin University in China
  - https://www.researchgate.net/publication/323201927 (Feb 2018)
- Recently discovered this - haven't obtained access yet
- Features
  - Evokes the OWL 2 reasoner and EL reasoner in parallel
  - Can create a smaller OWL 2 (non-EL) module (compared to MORe)
- Results
  - Report that reasoning time is reduced by 83.7% compared against MORe

Modular Reasoning and SNOMED CT Where To Go From Here?

SNOMED International

- Further develop “experimental” options (not yet complete)
  - RL rewrite
  - RL rewrite and materialization by stages
- Develop and/or apply new module extraction algorithms
  - Goal is to minimize number of axioms in OWL 2 module while preserving all inferences
    - May be able to reconsider using HermiT if the OWL 2 module is small enough
- Add additional reasoners and combinations
  - RDFox?
- Evaluate ComR (if possible)

Questions?

- Email me at rob@hausamconsulting.com