



Medical University of Graz

Institute for Medical Informatics,  
Statistics and Documentation,

# Modularity Issues in Biomedical Ontologies

Stefan Schulz



# Structure of Presentation

- Overview of large ontology projects in the biomedical domain
- Modularization issues
- Study on purpose-driven modularization of SNOMED CT

# Structure of Presentation

- Overview of large ontology projects in the biomedical domain
- Modularization issues
- Study on purpose-driven modularization of SNOMED CT

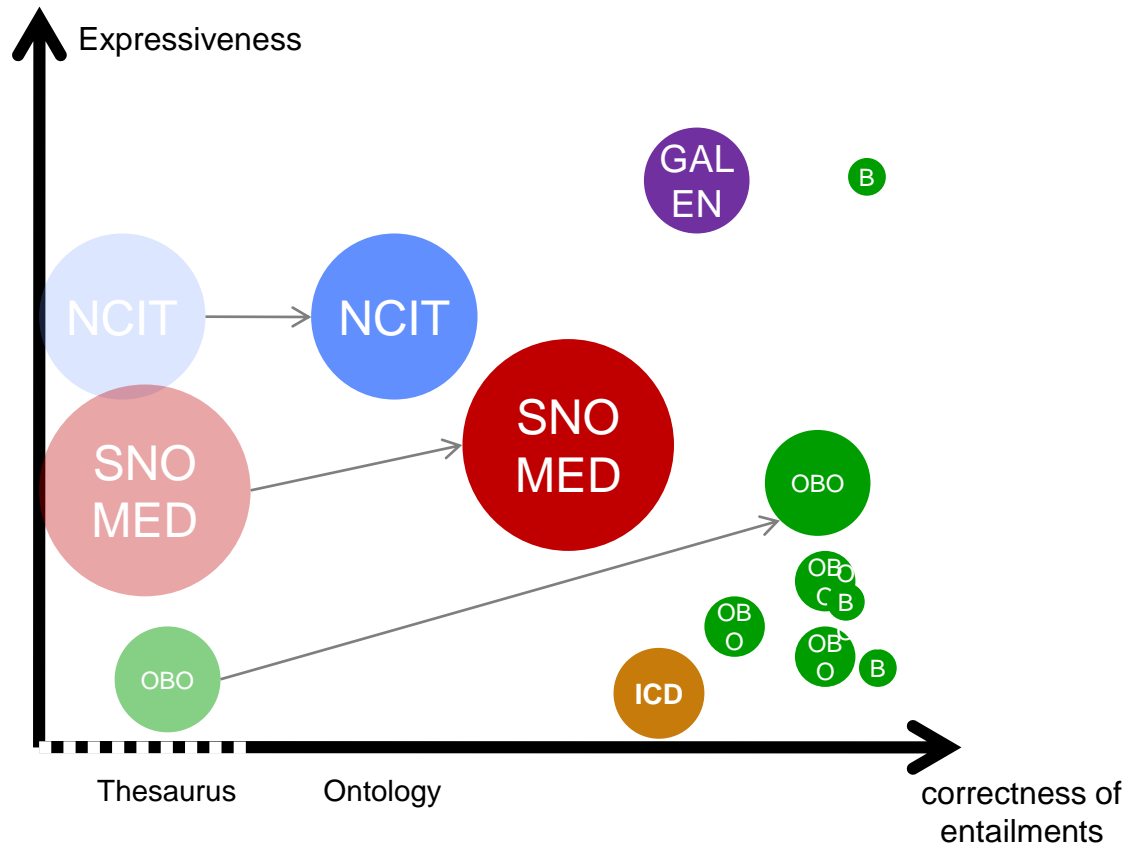
# Scoping: Large biomedical ontologies

- Size:  $> 10^4$  representational units
- Domain: life science research and health care
- Use of logic
  - Typical ontologies: OBO Foundry Ontologies (GO, ChEBI, FMA, OBI, ...), SNOMED CT, NCIT, GALEN,
  - No ontologies: UMLS meta, MeSH

# Large biomedical ontologies are slowly maturing

- Legacy: originally not conceived as ontologies but as systems for classification and semantic annotation
- "Ontologization" of thesaurus-like structures
- Few use cases for logic-based reasoning
- General tendency towards OWL
- Persisting problems
  - Understanding foundations of logic
  - Clarifying ontological commitment

# The gradual mutation from thesauri to ontologies



# Thesauri and Ontologies represent different things

## Informal Thesauri

---

- Examples: MeSH, UMLS Metathesaurus, WordNet
- Describe **terms** of a domain
- **Concepts**: represent the meaning of (quasi-) synonymous terms
- Concepts related by (informal) semantic relations
- Linkage of concepts:

*C1 Rel C2*



## Formal ontologies

---

- Examples: openGALEN, OBO, SNOMED
- Describe **entities** of a domain
- **Classes**: collection of entities according to their properties
- Axioms state what is universally true for all members of a class
- Logical expressions:

*C1 comp rel quant C2*

# Misconceptions of easy and straightforward thesaurus ontologization

- Upgrading a thesaurus to a formal ontology
- Rationales: use of standards (e.g. OWL-DL), enhanced reasoning, clarification of meaning, internal quality assurance...
- Underestimated:
  - Correctness of quantifications
  - Ontological commitment
  - Thesaurus triples cannot be unambiguously translated into ontology axioms

*C1 Rel C2*



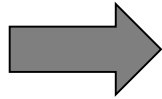
*C1 comp rel quant C2*



# The translation of triples into DL statements is ambiguous

Translation of triples

*C1 Rel C2*



*C1 subClassOf rel some C2*

or

*C1 subClassOf rel only C2*

or

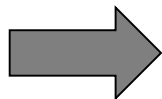
*C2 subclassOf inv(rel) some C2*

*or...*

Translation of groups  
of triples

*C1 Rel C2*

*C1 Rel C3*



*C1 subClassOf (rel some C2) and (rel some C3)*

or

*C1 equivalentTo (rel some C2) and (rel some C3)*

or

*C1 equivalentTo (rel some ((C2 or C3)))*

*or ...*

# Ontologies are not exactly made for represent contingent knowledge

- *“Aspirin Treats Headache”*  
*“Headache Treated-by Aspirin”*  
(seemingly intuitively understandable)
- Translation problem:
  - Not every aspirin tablet treats some headache
  - Not every headache is treated by some aspirin
- Description logics do not allow probabilistic, default, or normative assertions
- Axioms can only state what is true for all members of a class
- Introducing dispositions into ontology possible but not very intuitive  
("every aspirin tablet has an inherent disposition which is only realized by treating headache)

Schulz S, Stenzhorn H, Boeker M, Smith B: Strengths and limitations of formal ontologies in the biomedical domain. RECIIS - Electronic Journal in Communication, Information and Innovation in Health, 2009; 3 (1): 31-45:<http://dx.doi.org/10.3395/reciis.v3i1.241en>

Schulz S, Jansen L: Molecular interactions: On the ambiguity of ordinary statements in biomedical literature. Applied Ontology, 2009; 4 (1): 21-34: <http://dx.doi.org/10.3233/AO-2009-0061>

# Structure of Presentation

- Overview of large ontology projects in the biomedical domain
  - OBO Foundry
  - NCI Thesaurus
  - SNOMED CT
- Study on purpose-driven modularization of SNOMED CT

# The OBO Foundry – Modularity by design

- Collaborative, user-driven, bottom up initiative
- Driven by the success of the Gene Ontology  
(... also a thesaurus-like artifact incrementally mutating into an ontology)
- Guided by ontology development principles
- Rooted in upper ontologies (BFO + RO)
- Goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain
- Division of labor amongst domain experts
- Originally using semi-formal OBO syntax, now increasingly OWL-DL

# OBO Foundry: Orthogonality by Upper-level, and Granularity divisions

| RELATION<br>TO TIME<br><br>GRANULARITY | CONTINUANT                            |  |                                    |                                 | OCCURRENT                              |
|--|---------------------------------------|--|------------------------------------|---------------------------------|--|
|  | INDEPENDENT                           |  | DEPENDENT                          |                                 |  |
| ORGAN AND ORGANISM                     | Organism<br>(NCBI<br>Taxonomy)        | Anatomical<br>Entity<br>(FMA,<br>CARO)     | Organ<br>Function<br>(FMP, CPRO)   | Phenotypic<br>Quality<br>(PaTO) | <b>Biological<br/>Process<br/>(GO)</b> |
| CELL AND CELLULAR COMPONENT            | Cell<br>(CL)                          | <b>Cellular<br/>Component</b><br>(FMA, GO) | Cellular<br>Function<br>(GO)       |                                 |  |
| MOLECULE                               | Molecule<br>(ChEBI, SO,<br>RnaO, PrO) |  | <b>Molecular Function<br/>(GO)</b> |                                 | Molecular Process<br>(GO)              |

# Linking OBO Foundry ontologies

- OBO Foundry principles
  - Single *is\_a* parenthood, further *is\_a* relations *inferred*
  - Genus-species definitions
- Reality: most Foundry ontologies have no equivalence axioms at all
- Example:
  - *Calcitonin secreting cell* (Cell Ontology) can be defined as a *Secretory cell* which secretes *Calcitonin* (ChEBI)
  - *Heart development* (Gene Ontology) can be defined as a *Develomental process* which has *Heart* (FMA) as participant
- Conclusion: Sufficient representation of definitional knowledge requires cross-linking OBO modules.

# OBO Foundry cross products

|            | ChEBI | PRO | CC | cell | anat | envo | PATO | MF   | BP   |
|------------|-------|-----|----|------|------|------|------|------|------|
| ChEBI      |       |     |    |      |      |      |      |      |      |
| PRO        |       |     |    |      |      |      |      |      |      |
| GO-CC      |       |     |    | 16   |      |      |      |      | 174  |
| Cell       | 14    |     | 8  |      |      |      | 70   |      | 13   |
| Gross anat |       |     |    |      |      |      |      |      |      |
| Envo       | 19    |     |    |      |      |      | 28   |      |      |
| phen       | MP    | 273 | 10 | ?    | 654  | 3300 |      | 5815 | 1040 |
|            | WP    | 115 |    | ?    | 45   | 432  |      | 1450 | 686  |
|            | TO    | 127 |    |      | 2    | 355  | 19   |      | 1    |
| disease    |       |     |    |      |      |      |      |      |      |
| GO-MF      | 2407  |     |    |      |      |      |      |      |      |
| GO-BP      |       |     |    | 511  | 1085 |      | 55   |      |      |

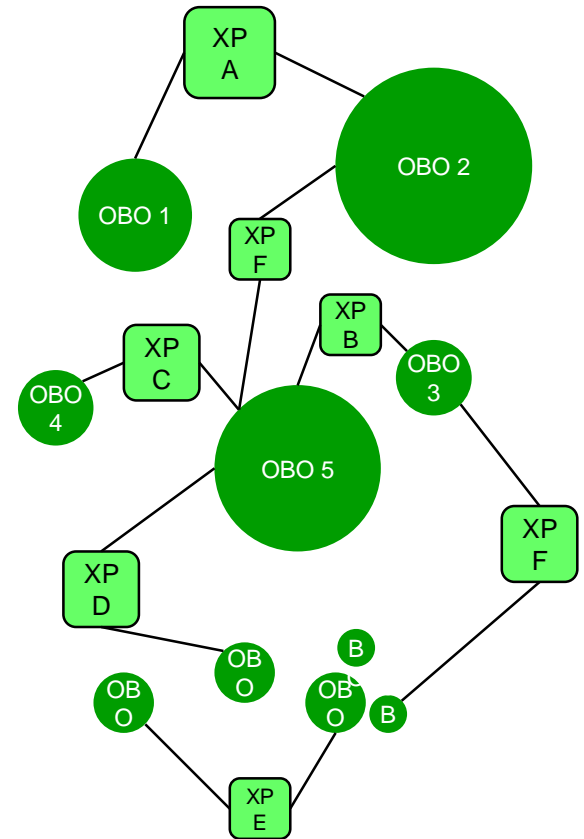
# OBO Foundry architecture

● Typical axioms in Foundry Modules:

- A subclassOf B
- A subclassOf r some C

□ Typical axioms in Cross Products

- A equivalentTo r some D
- A equivalentTo r some D and s some E

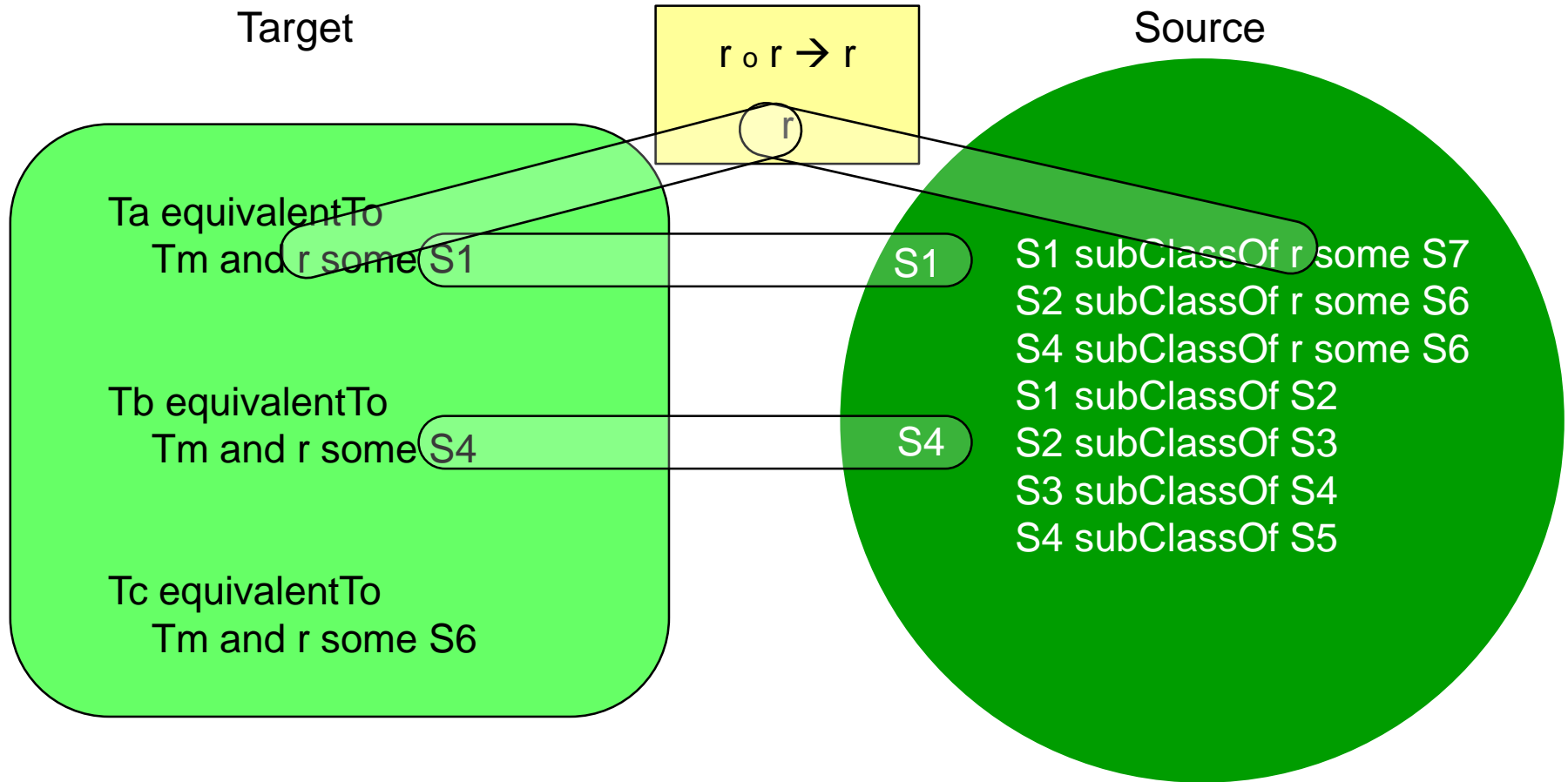




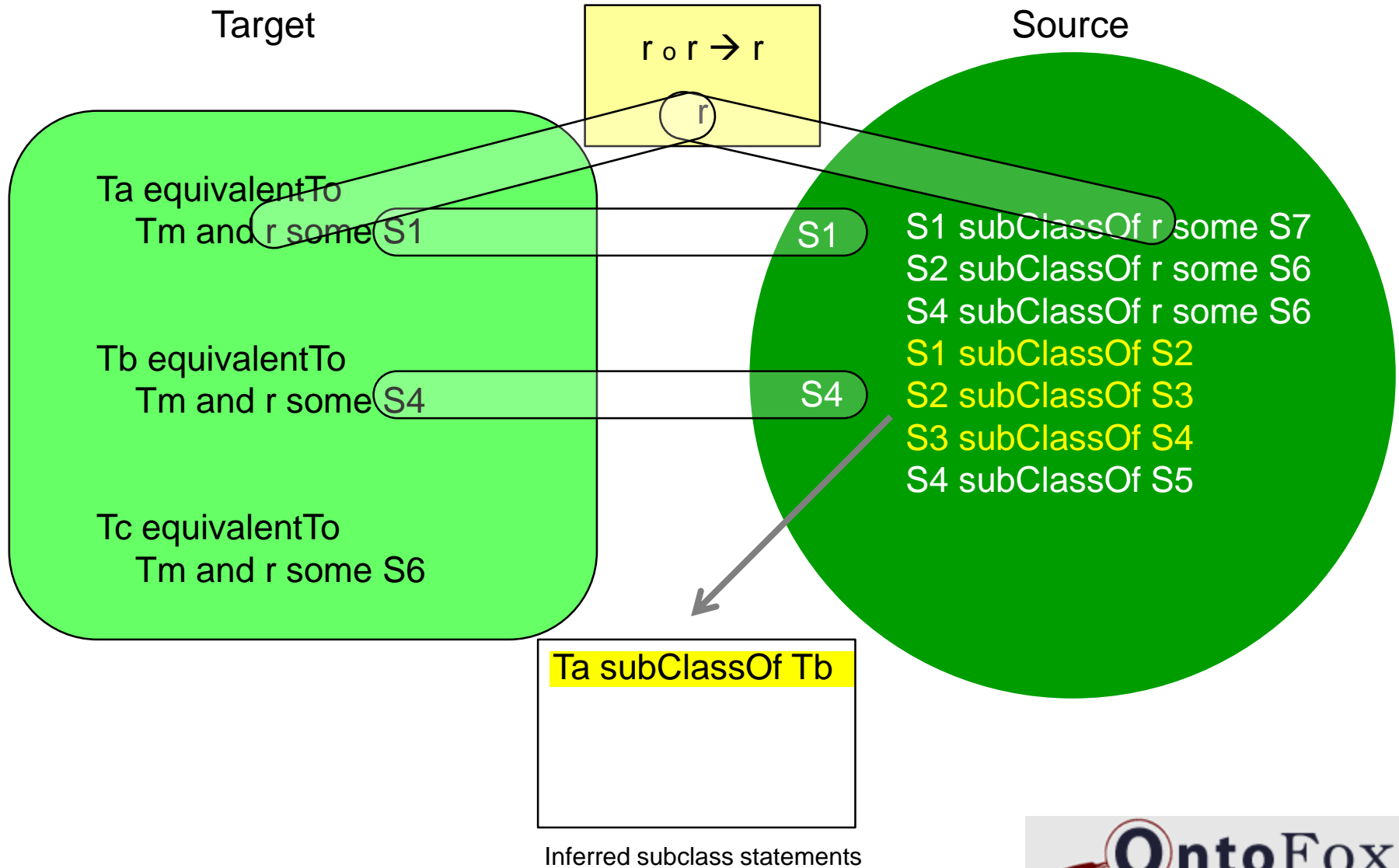
# MIREOT: minimum information to reference an external ontology term

- Guidelines for *manually* importing required terms from an external resource into a target ontology
- Minimal set to unambiguously identify a term: URI of class, source ontology , position in target ontology, superclasses...
- Problem (and possible challenge for WoMo):  
Given a signature (i.e. the set of references to an external ontology): extract axioms in the source ontology that produce new entailments in the target ontology
- Supported by OntoFox (<http://ontofox.hegroup.org/>)

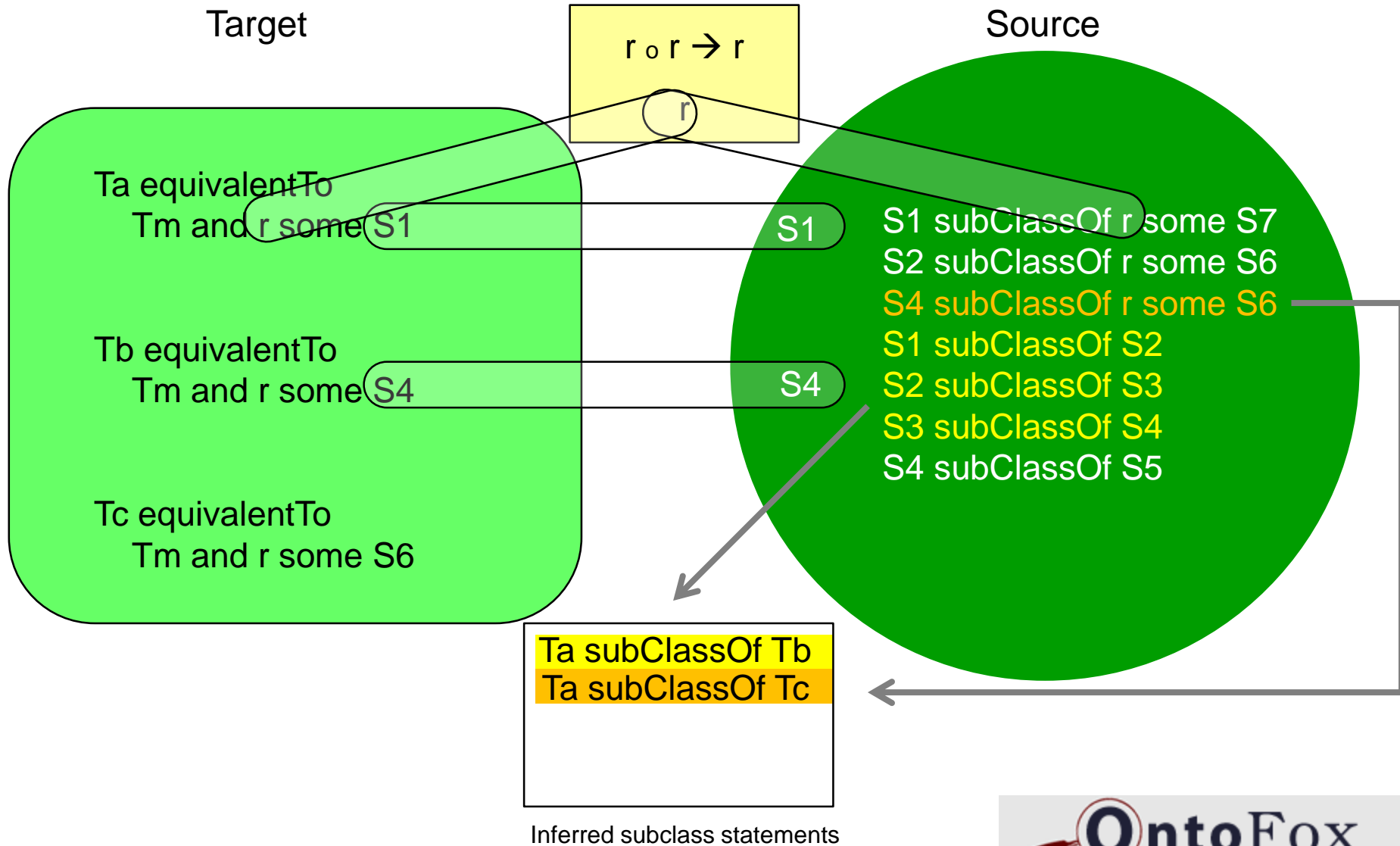
# Ontofox implementation of MIREOT: extracting significant axioms from source



# Ontofox implementation of MIREOT: extracting significant axioms from source



# Ontofox implementation of MIREOT: extracting significant axioms from source



# Structure of Presentation

- Overview of large ontology projects in the biomedical domain
  - OBO Foundry
  - NCI Thesaurus
  - SNOMED CT
- Study on purpose-driven modularization of SNOMED CT

# The NCIthesaurus

The screenshot shows the NCIthesaurus website interface. At the top, there is a red header with the National Cancer Institute logo and the text "National Cancer Institute" and "U.S. National Institutes of Health | www.cancer.gov". Below the header, the page is divided into several sections. On the left, there is a large "NCIthesaurus" logo. On the right, there is a search bar with a "Search" button and a "?". Below the search bar, there are radio buttons for search criteria: "Exact Match", "Begins With", "Contains", "Name/Code", "Property", and "Relationship". Below the search bar, there are links for "Home", "View Hierarchy", "Subsets", and "Help". Below the search bar, there is a "Quick Links" dropdown menu. Below the search bar, there is a "Welcome" section with a "Version: June 27, 2011 (11.06d)" label. The "Welcome" section contains a paragraph describing the NCIthesaurus and a list of features. Below the "Welcome" section, there is a paragraph describing the NCIthesaurus and its use. Below the "Welcome" section, there are three links: "Source Home Page", "Download", and "caBIG VKC Link". On the right side of the page, there are three boxes with logos and text: "EVS NCI Enterprise Vocabulary Services", "NCIm NCI Metathesaurus", and "cancer.gov NCI Terminology Resources".

**National Cancer Institute** U.S. National Institutes of Health | www.cancer.gov

**EVS** Enterprise Vocabulary Services **NCI** Term Browser

**NCIthesaurus**

Search ?

Exact Match  Begins With  Contains  
 Name/Code  Property  Relationship

Home | View Hierarchy | Subsets | Help

Quick Links

**Welcome** Version: June 27, 2011 (11.06d)

NCI Thesaurus (NCIt) provides reference terminology for many NCI and other systems. It covers vocabulary for clinical care, translational and basic research, and public information and administrative activities.

**NCIt features:**

- Stable, unique codes for biomedical concepts;
- Preferred terms, synonyms, definitions, research codes, external source codes, and other information;
- Links to [NCI Metathesaurus](#) and other information sources;
- Over 200,000 cross-links between concepts, providing formal logic-based definition of many concepts;
- Extensive content integrated from NCI and other partners, much available as separate NCIt [subsets](#)
- Updated frequently by a team of subject matter experts.

NCIt is a widely recognized standard for biomedical coding and reference, used by a broad variety of public and private partners both nationally and internationally including the Clinical Data Interchange Standards Consortium Terminology (CDISC), the U.S. Food and Drug Administration (FDA), the Federal Medication Terminologies (FMT), and the National Council for Prescription Drug Programs (NCPDP).

Source Home Page: <http://ncit.nci.nih.gov>

Download: [http://evs.nci.nih.gov/ftp1/NCI\\_Thesaurus](http://evs.nci.nih.gov/ftp1/NCI_Thesaurus)

caBIG VKC Link: [https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/NCI\\_Thesaurus](https://cabig-kc.nci.nih.gov/Vocab/KC/index.php/NCI_Thesaurus)

**EVS** [NCI Enterprise Vocabulary Services](#): Terminology resources and services for NCI and the biomedical community.

**NCIm** [NCI Metathesaurus](#): Comprehensive database of 3,600,000 terms from 76 terminologies.

**NCI Term Browser** [NCI Term Browser](#): NCI and other terminologies in an integrated environment.

**cancer.gov** [NCI Terminology Resources](#): More information on NCI dictionaries and resources.

<http://ncit.nci.nih.gov/ncitbrowser/>

# The NCI Thesaurus (NCIT) – a (N)ontology for cancer research

- Reference terminology of the National Cancer Institute (clinical care, translational and basic research, public and administrative information)
- Using Ontylog (Apelon, Inc. Terminology Development Environment (TDE))
- 90k classes, 11k eq axioms, 110k subclass axioms, 23k classes interpreted as individuals ("punning") for enabling non-quantified triplets
- Upper-level partition (disjoint axioms at 1<sup>st</sup> hierarchical level)
- Domain / range restrictions of relations
- Distributed in OWL (SH(D)) as stated and inferred version
- Reasoning use cases unknown
- Statement: "*NCI Thesaurus has some ontology-like features but NCI Thesaurus is not an ontology and is not designed or intended to one.*"

# Ontolog language used in the construction of NCIT and SNOMED CT

| <i>Constructor</i>           | <i>Syntax</i>  | <i>Semantics</i>   |
|------------------------------|----------------|--|
| <i>Concept name</i>          | $C$            | $C^I$ (where $C^I \subseteq \Delta^I$ )                        |
| Top                          | $\top$         | $\Delta^I$   |
| Bottom                       | $\perp$        | $\emptyset$  |
| Conjunction                  | $C \sqcap D$   | $C^I \cap D^I$   |
| Disjunction                  | $C \sqcup D$   | $C^I \cup D^I$   |
| <i>Universal restriction</i> | $\forall R.C$  | $\{x \mid \forall y : R^I(x,y) \rightarrow C^I(y)\}$           |
| Existential restriction      | $\exists R.C$  | $\{x \mid \exists y : R^I(x,y) \wedge C^I(y)\}$                |
| Modal restriction            | $\diamond R.C$ | $\{x \mid \text{Pr}(\exists y : R^I(x,y) \wedge C^I(y)) > 0\}$ |
| Role name                    | $R$            | $R^I$ (where $R^I \subseteq \Delta^I \times \Delta^I$ )        |

| <i>Definitional or Axiomatic Constraint</i> | <i>Syntax</i>        | <i>Semantic Constraint</i> |
|---|----------------------|----------------------------|
| <i>Concept definition</i>                   | $C \doteq D$         | $C^I \equiv D^I$           |
| Concept subsumption axiom                   | $C \subseteq D$      | $C^I \subseteq D^I$        |
| Role subsumption axiom                      | $R \subseteq S$      | $R^I \subseteq S^I$        |
| Right identity axiom                        | $R \circ S \doteq R$ | $(R \circ S)^I \equiv R^I$ |

<sup>1</sup> Note that disjunction, a feature currently under development, can only be used in role values



# The NCIT upper level

**Class hierarchy: Abnormal\_Cell**

- Thing
  - Abnormal\_Cell**
  - Activity
  - Anatomic\_Structure\_System\_or\_Substance
  - Biochemical\_Pathway
  - Biological\_Process
  - Chemotherapy\_Regimen
  - Conceptual\_Entities
  - DIRECTED-BINARY-RELATION
  - Diagnostic\_Therapeutic\_and\_Research\_Equipment
  - Diagnostic\_or\_Prognostic\_Factor
  - Diseases\_Disorders\_and\_Findings
  - Drugs\_and\_Chemicals
  - Experimental\_Organism\_Anatomical\_Concepts
  - Experimental\_Organism\_Diagnoses
  - Gene
  - Gene\_Product
  - Molecular\_Abnormality
  - NCI\_Administrative\_Concepts
  - Organisms
  - PAL-CONSTRAINT
  - Properties\_or\_Attributes
  - Retired\_Concepts

**Annotations: Abnormal\_Cell**

Annotations +

**DEFINITION**

"<ncicp:ComplexDefinition><ncicp:def-definition>An abnormal human cell type which can occur in either disease states or disease models.</ncicp:def-definition><ncicp:def-source>NCI</ncicp:def-source></ncicp:ComplexDefinition>"<sup>M</sup>XMLLiteral

**FULL\_SYN**

"<ncicp:ComplexTerm><ncicp:term-name>Abnormal Cell</ncicp:term-name><ncicp:term-group>PT</ncicp:term-group><ncicp:term-source>NCI</ncicp:term-source></ncicp:ComplexTerm>"

**Description: Abnormal\_Cell**

Members +

Keys +

Disjoint classes +

- Diagnostic\_or\_Prognostic\_Factor**
- Biochemical\_Pathway**
- Diseases\_Disorders\_and\_Findings**
- NCI\_Administrative\_Concepts**
- Activity
- Diagnostic\_Therapeutic\_and\_Research\_Equipment
- Conceptual\_Entities
- Retired\_Concepts
- Anatomic\_Structure\_System\_or\_Substance
- Gene\_Product
- Experimental\_Organism\_Anatomical\_Concepts
- Gene
- Drugs\_and\_Chemicals
- Experimental\_Organism\_Diagnoses
- Biological\_Process
- Chemotherapy\_Regimen
- Molecular\_Abnormality
- Organisms
- Properties\_or\_Attributes

# NCIT Content problems

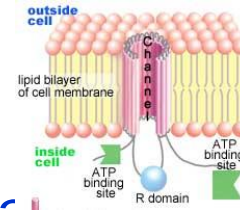
Mismatch between the intended meaning of labels and DL semantics:

- *Ureter\_Small\_Cell\_Carcinoma* subclassOf **Disease\_May\_Have\_Finding** some *Pain*



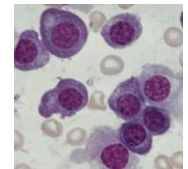
False encoding of distributive statements

- *Calcium-Activated\_Chloride\_Channel-2* subclassOf **Gene\_Product\_Expressed\_In\_Tissue** some *Lung* and **Gene\_Product\_Expressed\_In\_Tissue** some *Mammary\_Gland* and **Gene\_Product\_Expressed\_In\_Tissue** some *Trachea*



Existential quantification over parts instead of wholes

- *Antibody\_Producing\_Cell* subclassOf **Part\_Of** some *Lymphoid\_Tissue*



# Structure of Presentation

- Overview of large ontology projects in the biomedical domain
  - OBO Foundry
  - NCI Thesaurus
  - SNOMED CT
- Study on purpose-driven modularization of SNOMED CT

# SNOMED CT

INTERNATIONAL HEALTH TERMINOLOGY  
STANDARDS DEVELOPMENT ORGANISATION



[About IHTSDO](#) **[SNOMED CT](#)** [Join Us](#) [News](#) [Projects](#) [Publications](#) [Members](#) [Contact Us](#) [Home](#)

[Search](#) [Print](#) [FAQ](#)

[About SNOMED CT](#)

[Value Proposition for SNOMED CT](#)

[SNOMED CT Lookup Service](#)

[Release of SNOMED CT](#)

[Licensing](#)

[Namespaces](#)

[History of SNOMED CT](#)

[Request Submission](#)

[Member Exchange](#)

[Quality Assurance](#)

[Tooling](#)

## SNOMED CT

SNOMED CT (**S**ystematized **N**omenclature of **M**edicine-**C**linical **T**erms) is considered to be the most comprehensive, multilingual clinical healthcare terminology in the world. Each year, avoidable deaths and injuries occur because of poor communication between healthcare practitioners, or because busy practitioners forget or neglect to follow their own criteria for best practices. The delivery of a standard clinical terminology for use across the world's health information systems can therefore make a significant contribution towards improving the quality and safety of healthcare.

SNOMED CT aims to contribute to the improvement of patient care through underpinning the development of systems to accurately record health care encounters and to deliver decision support to health care providers. Ultimately, patients will benefit from the use of SNOMED CT to more clearly describe and accurately record their care, in building and facilitating better communication and interoperability in electronic health record exchange, and in creating systems that support health care decision making.

SNOMED CT intellectual property rights were transferred to the SNOMED SDO® in the formal creation of the IHTSDO.

SNOMED CT was originally created by the College of American Pathologists by combining SNOMED RT and a computer based nomenclature and classification known as Clinical Terms Version 3, formerly known as Read Codes Version 3, which was created on behalf of the UK Department of Health and is Crown copyright.

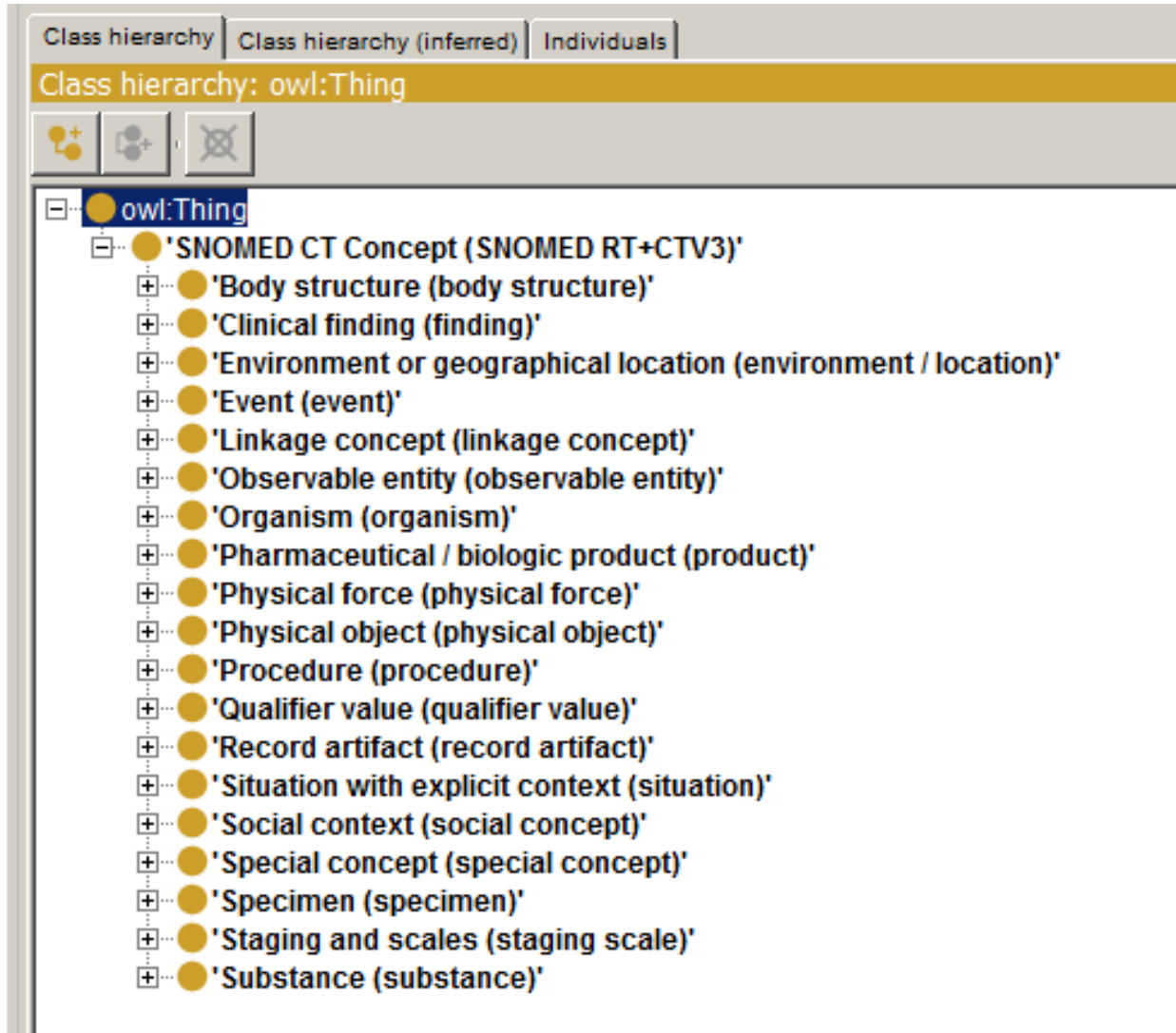
For examples of SNOMED CT implementation experience, click [here](#)

IHTSDO®, SNOMED® and SNOMED CT® are registered trademarks of the International Health Terminology Standards Development Organisation.

# SNOMED CT – universal clinical terminology with ontological foundations

- Systematized Nomenclature of Medicine - Clinical Terms
- Terminology for clinical data covering diseases, findings, procedures, organisms, substances etc.
- Using Ontylog (Apelon, Inc. Terminology Development Environment (TDE)), distribution only of the inferred version in tabular format
- Extensive taxonomy with than 311, 000 concepts, connected by 1,360,000 relational expressions
- Script creates OWL version (EL++); no disjointness statements
- OWL version not used in routine

# The SNOMED CT upper level



# SNOMED CT content problems (frequent)

- Anatomy-related entailments:

*AmputationOfTheFoot* equivalentTo *rg* some  
(**method** some *Amputation* and **procedureSiteDirect** some *FootStructure*)

*AmputationOfToe* equivalentTo *rg* some  
(**method** some *Amputation* and **procedureSiteDirect** some *ToeStructure*)

*ToeStructure* subclassOf *FootStructure*

---

*AmputationOfToe* subclassOf *AmputationOfTheFoot*

- Negations

*ThumbAbsent* subclassOf *VenousFinding* and  
**hasFindingSite** some *ThumbStructure*

- Confusion Epistemology / Ontology

*PresumedViralAgent* subclassOf *Virus*

# General observations on NCIT and SNOMED CT

- Thesaurus / Ontology hybrids
  - NCIT: "no ontology"
  - SNOMED CT: ontology-aware redesign projects underway, increasing use OWL for prototyping
- Both OWL versions are logically consistent
- Used primarily for semantic annotations
- DL reasoning used in design process, but not really in applications
- Numerous entailments are unreliable, due to plain design errors, but also due to "workarounds" to express negation or probabilistic knowledge
- Browsers can deal with huge amount of terms, reduction in size not a prime desideratum for
- Truth-maintaining modularization not (yet) seen as a requirement



# Modularity in design of NCIT / SNOMED CT

- SNOMED CT:
  - Subhierarchies are meant to be disjoint (implicitly)
  - Subhierarchies are separately maintained.
  - Strict control of links between subhierarchies by Domain / Range restriction of relations (in editing tool, not in OWL version)
    - **hasFindingSite** relates *Finding* with *Body Structure*
    - **hasProcedureSite** relates *Procedure* with *Body Structure*
- NCIT:
  - Stated disjointness of top level categories
  - Domain / Range of relations specific for subhierarchies

# Modularity in use (SNOMED CT)

- SNOMED CT: Cross-hierarchy modularity
  - Manually derived problem list subset for summary level clinical documentation:
    - maximize data interoperability among institutions and facilitate the use of SNOMED CT as the primary coding system
    - focus on redesign and QA in content development process
  - SNOMED CT module as ontological basis of the planned version ICD-11 of the International Classification of Diseases
    - manual alignment of SNOMED CT disease classes to ICD classes
    - consistency of taxonomic links
    - inclusion of content referred to in the SNOMED definitions (body parts, qualities, microorganisms)

[http://www.nlm.nih.gov/research/umls/Snomed/core\\_subset.html](http://www.nlm.nih.gov/research/umls/Snomed/core_subset.html)

[http://www.ihtsdo.org/fileadmin/user\\_upload/Docs\\_01/About\\_IHTSDO/Harmonization/Final\\_agreement\\_2010\\_WHO\\_and\\_IHTSDO.pdf](http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/About_IHTSDO/Harmonization/Final_agreement_2010_WHO_and_IHTSDO.pdf)

# Structure of Presentation

- Overview of large ontology projects in the biomedical domain
  - OBO Foundry
  - NCI Thesaurus
  - SNOMED CT
- Study on purpose-driven modularization of SNOMED CT

# High granularity of SNOMED CT is only partly of interest...

- cardiac catheterisation
  - catheterisation of left heart
    - catheterisation of both left and right heart
      - cardiac catheterisation. right heart and retrograde left

- burn of upper limb
  - third degree burn of upper limb
    - third degree burn of wrist AND/OR hand
      - third degree burn of hand
        - deep third degree burn of hand
          - deep third degree burn of hand with loss of body part
            - deep third degree burn of finger. not thumb with loss of body part
              - deep third degree burn of two OR more fingers not including thumb

# High granularity of SNOMED CT is only partly of interest...

- cardiac catheterisation
  - catheterisation of left heart
    - catheterisation of both left and right heart
      - cardiac catheterisation. right heart and retrograde left

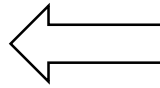


Cardiologist

- burn of upper limb
  - third degree burn of upper limb
    - third degree burn of wrist AND/OR hand
    - ~~third degree burn of hand~~
    - ~~deep third degree burn of hand~~
    - ~~deep third degree burn of hand with loss of body part~~
    - ~~deep third degree burn of finger. not thumb with loss of body part~~
      - ~~deep third degree burn of two OR more fingers not including thumb~~

# High granularity of SNOMED CT is only partly of interest...

- cardiac catheterisation
- ~~catheterisation of left heart~~
- ~~catheterisation of both left and right heart~~
- ~~cardiac catheterisation, right heart and retrograde left~~



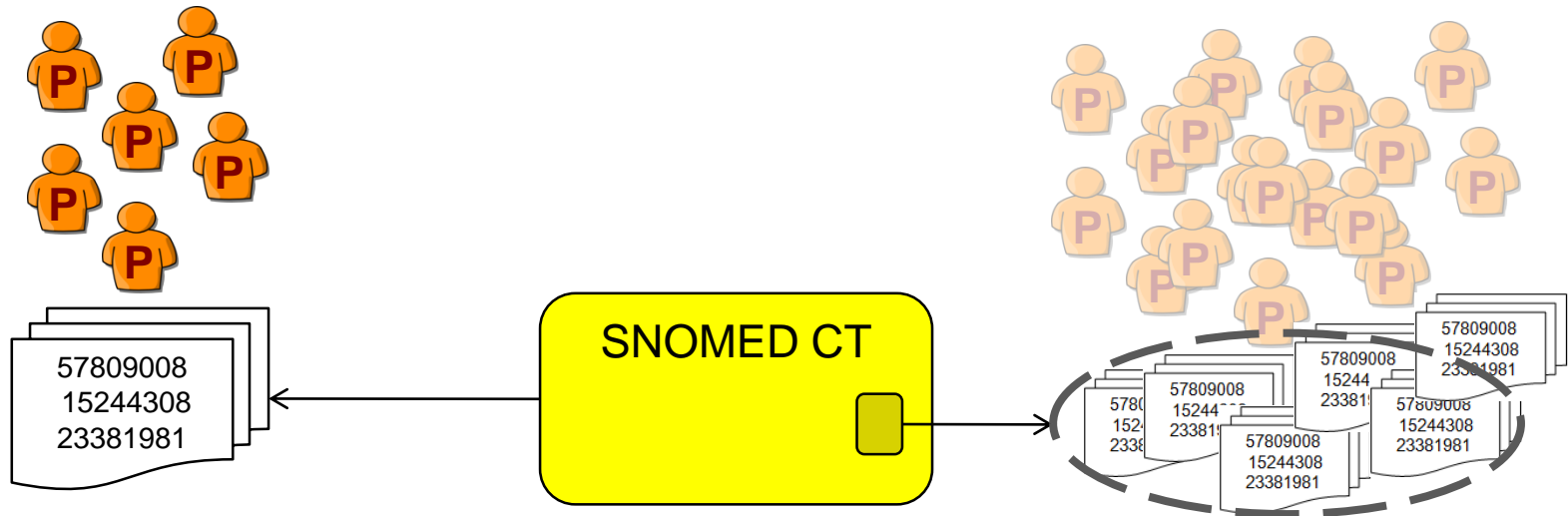
- burn of upper limb
- third degree burn of upper limb
- third degree burn of wrist AND/OR hand
- third degree burn of hand
- deep third degree burn of hand
- deep third degree burn of hand with loss of body part
- deep third degree burn of finger, not thumb with loss of body part
- deep third degree burn of two OR more fingers not including thumb



Surgeon

# Experiment on SNOMED CT modularization

- Given
  - an inpatient population of a specific clinical discipline (cardiology)
  - a collection of fully SNOMED coded medical records
- Create a compact SNOMED module suited to contain address the coverage and granularity requirements of the subdomain



# Sample fragment of an annotated discharge summary (manually coded)

| Original Text                           | SNOMED CT Term                                      | SNOMED CT Concept ID |
|---|---|----------------------|
| Masculino,                              | Male (finding)                                      | 248153007            |
| 43 anos,                                | Aging (finding)                                     | 248280005            |
| hipertenso,                             | Hypertensive disorder, systemic arterial (disorder) | 38341003             |
| tabagista,                              | Tobacco user (finding)                              | 110483000            |
| etilista,                               | Current drinker of alcohol (finding)                | 219006               |
| interna                                 | Hospital admission (procedure)                      | 32485007             |
| por infarto agudo do miocárdio          | Acute myocardial infarction (disorder)              | 57054005             |
| sem supradesnívelmaneto de segmento ST. | ST segment elevation (finding)                      | 76388001             |



# Graph traversal heuristics

- SNOMED CT as a graph with
  - taxonomic links                    A subClassOf B
  - nontaxonomic links                A subClassOf r some B  
(A subClassOf rg some (r some B))
- Basic approach (Seidenberg & Rector):
  - for each signature concept:
    - builds a set containing the complete hierarchy of the node
    - recursively follow links for every node in the set.
    - property filtering and depth limiting to limit the size of the target ontology

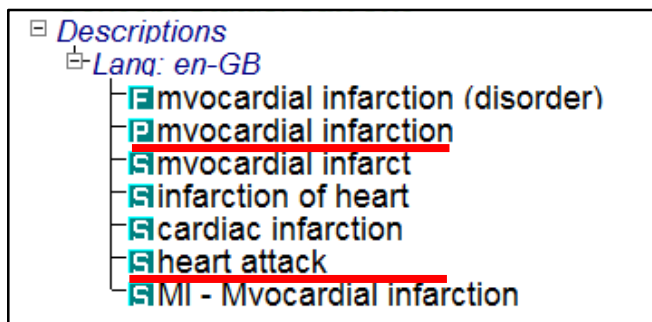
# Graph traversal heuristics: Four variations

- Four strategies
  - Upwards Segmentation: does not follow links in the subtree of the input concept
  - S-Heuristic: adding all sibling nodes (from all parents) and their *is-a* links.
  - ST-Heuristic: additionally include the complete subtree of each sibling.
  - IL-Heuristic: Each node connected by an incoming link is used as an extra seed.

# Relevance assessment of each SNOMED CT concepts based on corpus statistics

- Basic principle: score each SNOMED CT concept by relevance according to the occurrences of the attached terms in a reference corpus (MEDLINE)
- Several thresholds

SNOMED CT  
Concept 22298006



[J Cardiovasc Pharmacol Ther.](#) 2011 Sep;16(3-4):260-6.

## Perconditioning and postconditioning: current know

[Vinten-Johansen J, Shi W.](#)

1Division of Cardiothoracic Surgery, The Cardiothoracic Research Laboratory, Ca

### Abstract

The broad definition of "conditioning" is the application of a series of alte ischemic event in ischemic preconditioning, it is applied during the even **heart attack** victims to the hospital before percutaneous coronary interv enthusiastic attention from scientists that have done much to demonstra including gradual normalization of tissue pH, reduction in generation of r postconditioning does not exert cardioprotection in experimental models unknown, as is the interrelationship between the many molecular, cellul conditioning from another are unanswered questions. Yet, the translatio developments from industry.

[PLoS One.](#) 2011;6(7):e22693. Epub 2011 Jul 28.

## Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and

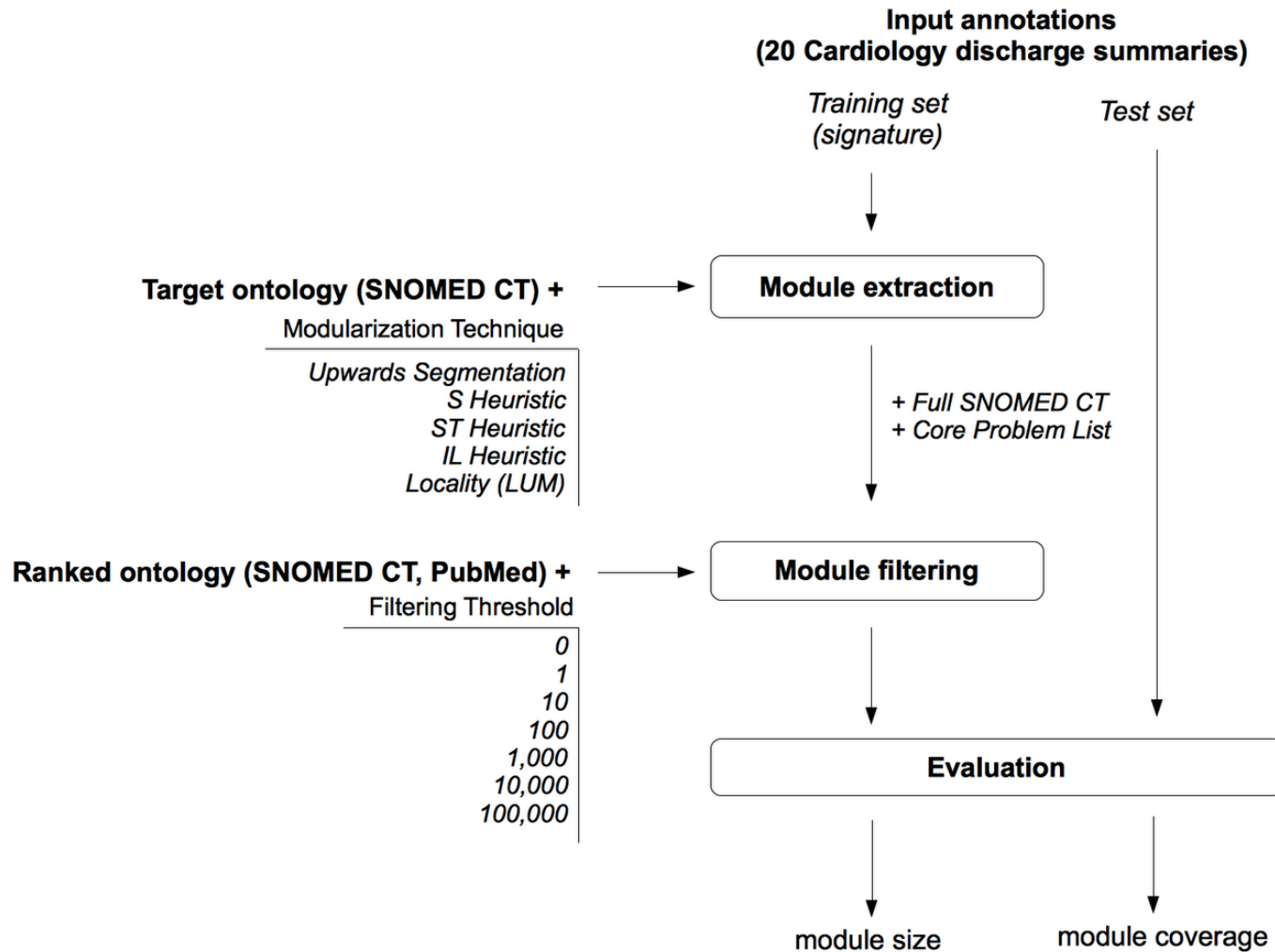
[Pedigo A, Seaver W, Odoi A.](#)

Department of Comparative Medicine, The University of Tennessee, Knoxville, Tennessee, United States of America.

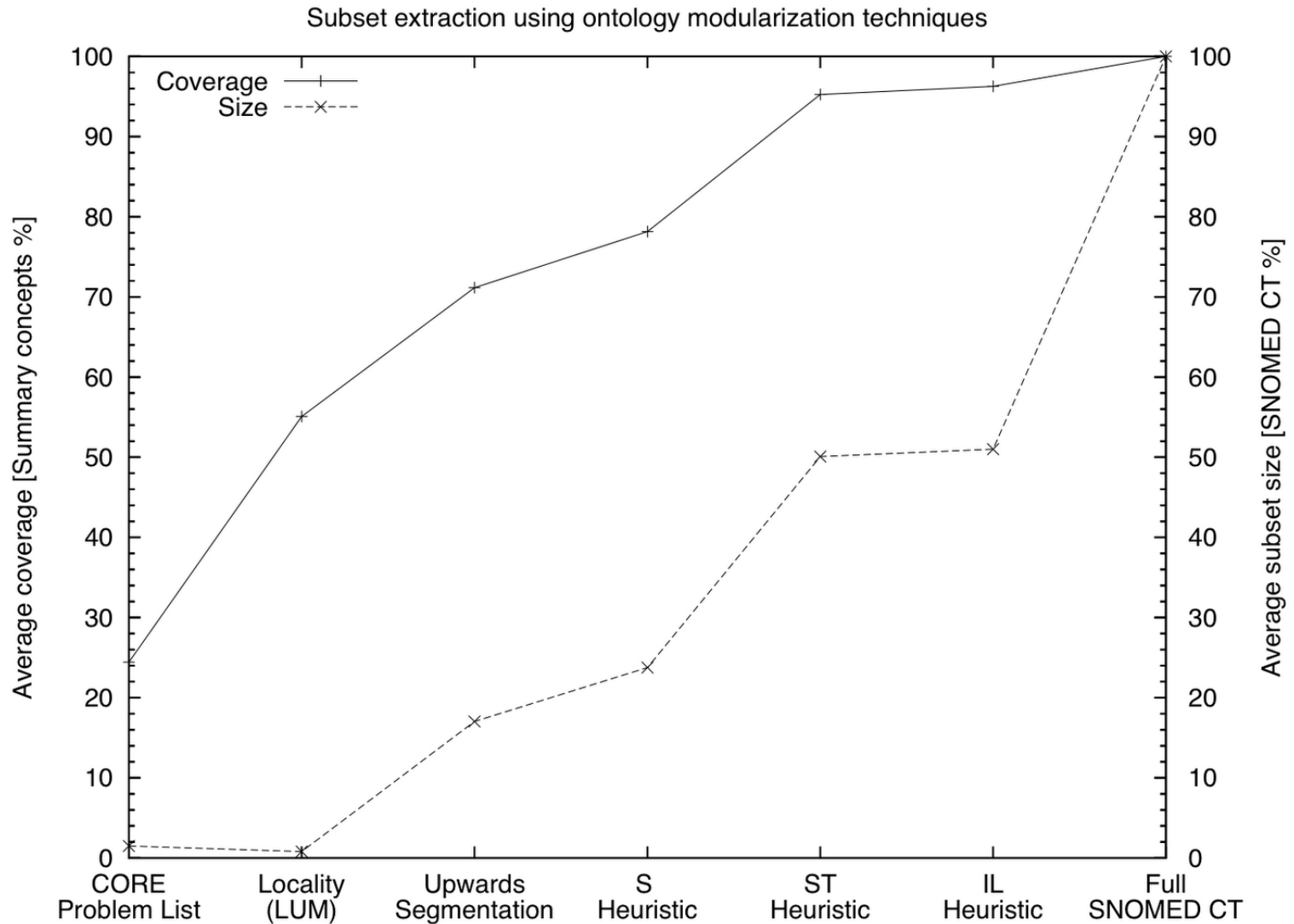
### Abstract

**BACKGROUND:** Socioeconomic, demographic, and geographic factors are known determinants of stroke and **myocardial infarction** (MI) risk. Clustering of these factor programs intended to reduce disparities. Given the complex and multidimensional nature of these factors, multivariate methods are needed to identify neighborhood cl

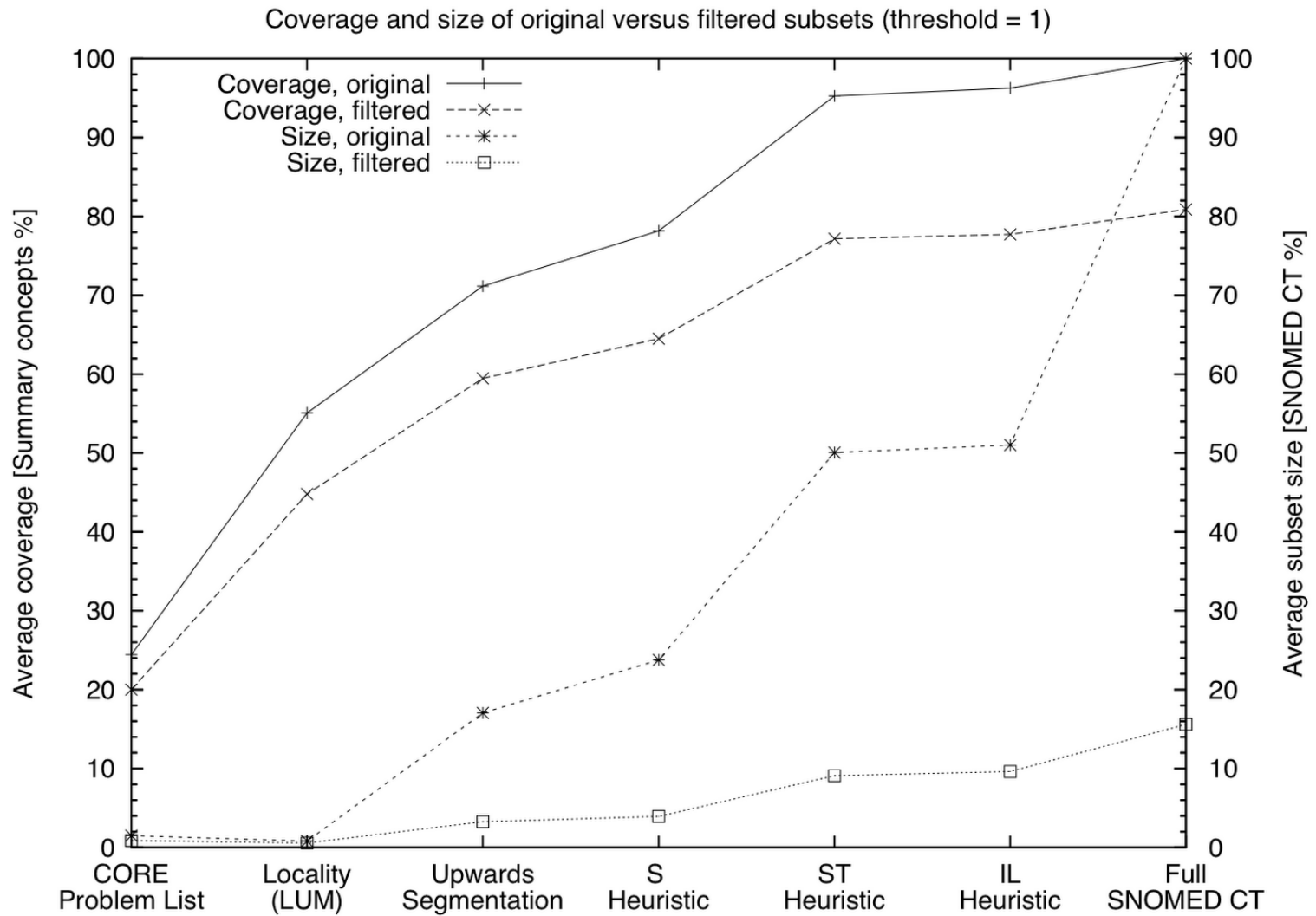
# Experimental design, parameters and measured variables



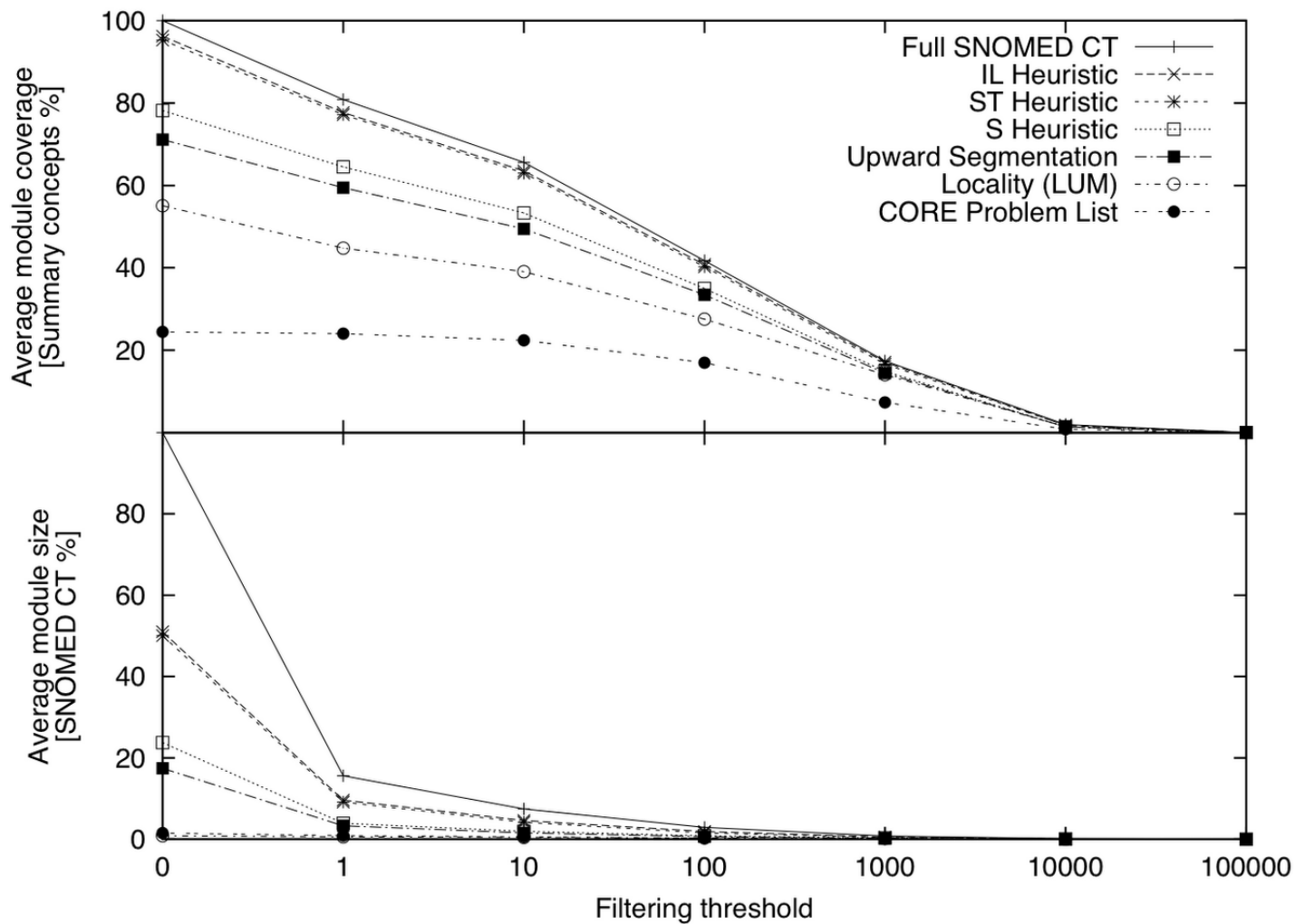
# Results: Coverage and Size



# Results: Coverage and Size



Influence of filtering in module coverage and size



# Discussion

- Module creation
  - If a coverage of 80% is considered, S-heuristics method of choice
  - Possible improvement: use of different approaches in different SNOMED CT subhierarchies
- Filtering
  - With complete SNOMED CT only 80%
  - Problem: vocabulary mismatch trainings corpus / test corpus. Improvement expected using clinical text and looser string matching
- Work in progress
- Testing / adapting other segmentation methods



# Conclusion

- Module creation in large biomedical ontologies:
  - Main interest in domain-specific subset
  - Module should cover must much more than the signature nodes
  - Logical properties of modules are still secondary, since logical entailments of both SNOMED CT and NCIT are not yet reliable
- Module creation in the context of (already modularly structured) OBO Foundry
  - Main interest in enrichment of the axioms in target module by re-using content from source module(s)
  - Finding additional axioms from source to compute new entailments in target

# Acknowledgements

Organizers of WoMo 2011

Pablo López

(PhD candidate at University of the Basque Country,  
Spain, currently at Stanford University)

pablo\_lopez@ehu.es