Using Abstraction Network in Complement to Description Logic for Quality Assurance Purposes in SNOMED CT

Presented by: Duo Wei

Mentor: Olivier Bodenreider

NATIONAL LIBRARY OF MEDICINE

Outline

- Introduction to SNOMED and Auditing
- Description Logic & Abstraction Network
- Abstraction Network Helps to Detect Errors
- How Abstraction Network in Complement DL for Quality Assurance

Quality Assurance for SNOMED

- Ontological & linguistic based techniques (*Ceusters, et al.*)
- Lexical approach (*Pathak, et al.*)
- Description-logic (DL) based techniques (*Cornet, et al.*)
- Formal concept analysis (FCA) method (*Jiang, et al.*)

Description Logic

- Description Logic (DL): A family of logic based Knowledge Representation formalisms
- DL not only provides the facilities for defining concepts , but it also provides methods for reasoning service
- Basic Inference on concept expression
  - Sumsumption
  - Satisfiability

DL & SNOMED

- It classifies SNOMED terminologies; it identifies inconsistencies
- Stated view: native view that only defining relationships that an author has explicitly stated to be true
- Inferred view: the distributed version generated by DL reasoner automatically
- Inferred view is derived from the stated view

Motivation

- DL serves as the basis of terminology classification and consistency checking
- DL alone is not sufficient to create an error-free system
- Errors can be divided into two categories

Motivation (cont'd)

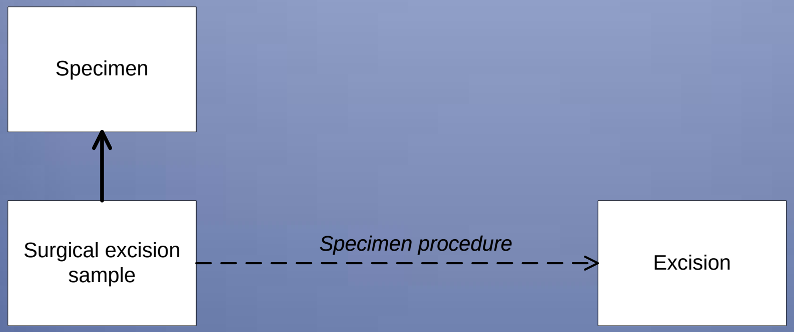
- Type I errors are those might cause logic contradictory in the system
  - Can be easily detected by DL
- Type II errors are those will not generate logic conflicts of the system
  - Independent auditing should be involved

Abstraction Networks

- **Structural abstraction network**
- **Hierarchical abstraction network**
- **Both derived automatically**

Specimen Hierarchy

- Concepts e.g. Surgical excision sample
- Relationships e.g. Specimen procedure



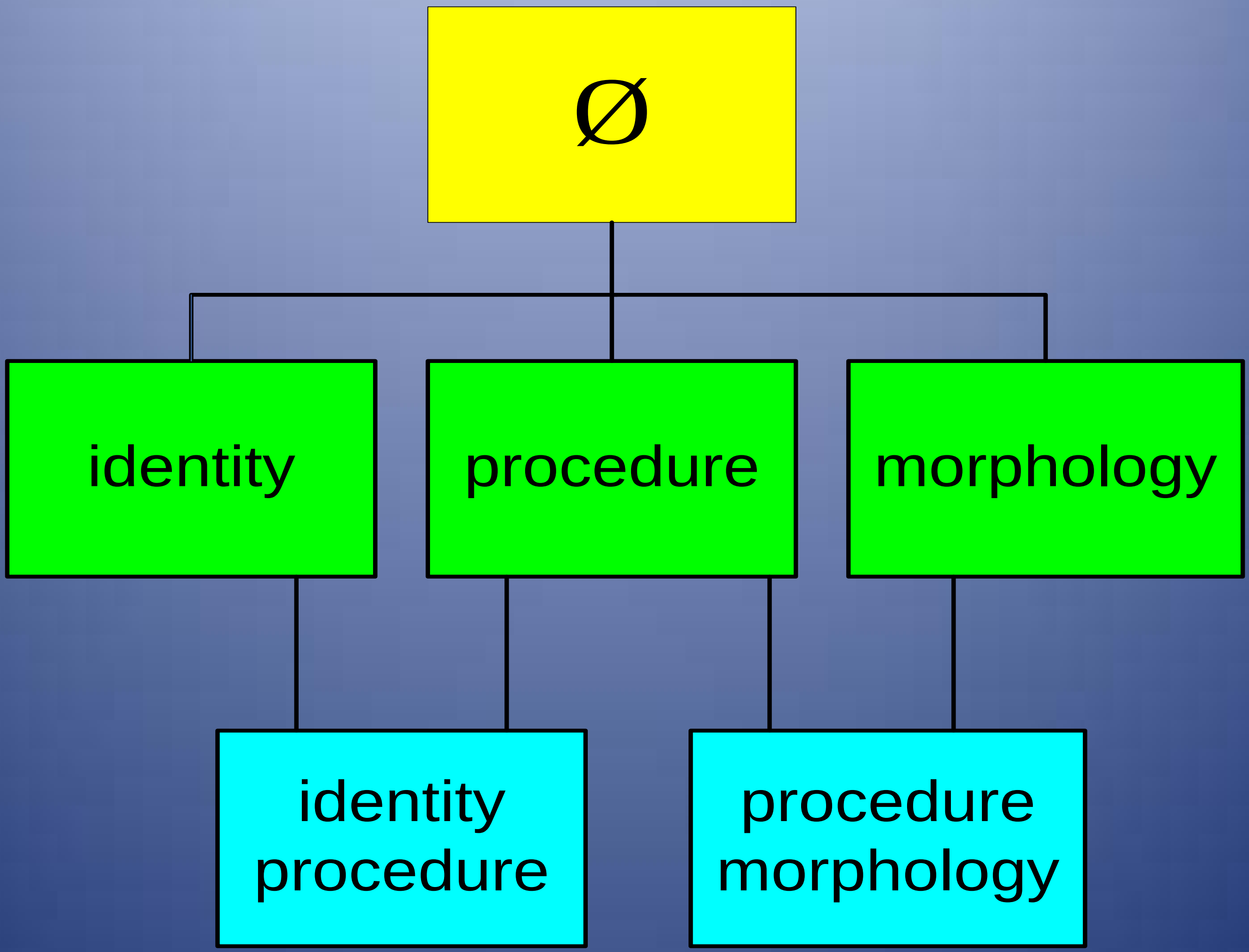
Full relationship name	Short form
Specimen source morphology	morphology
Specimen procedure	procedure
Specimen source identity	identity
Specimen substance	substance
Specimen source topography	topography

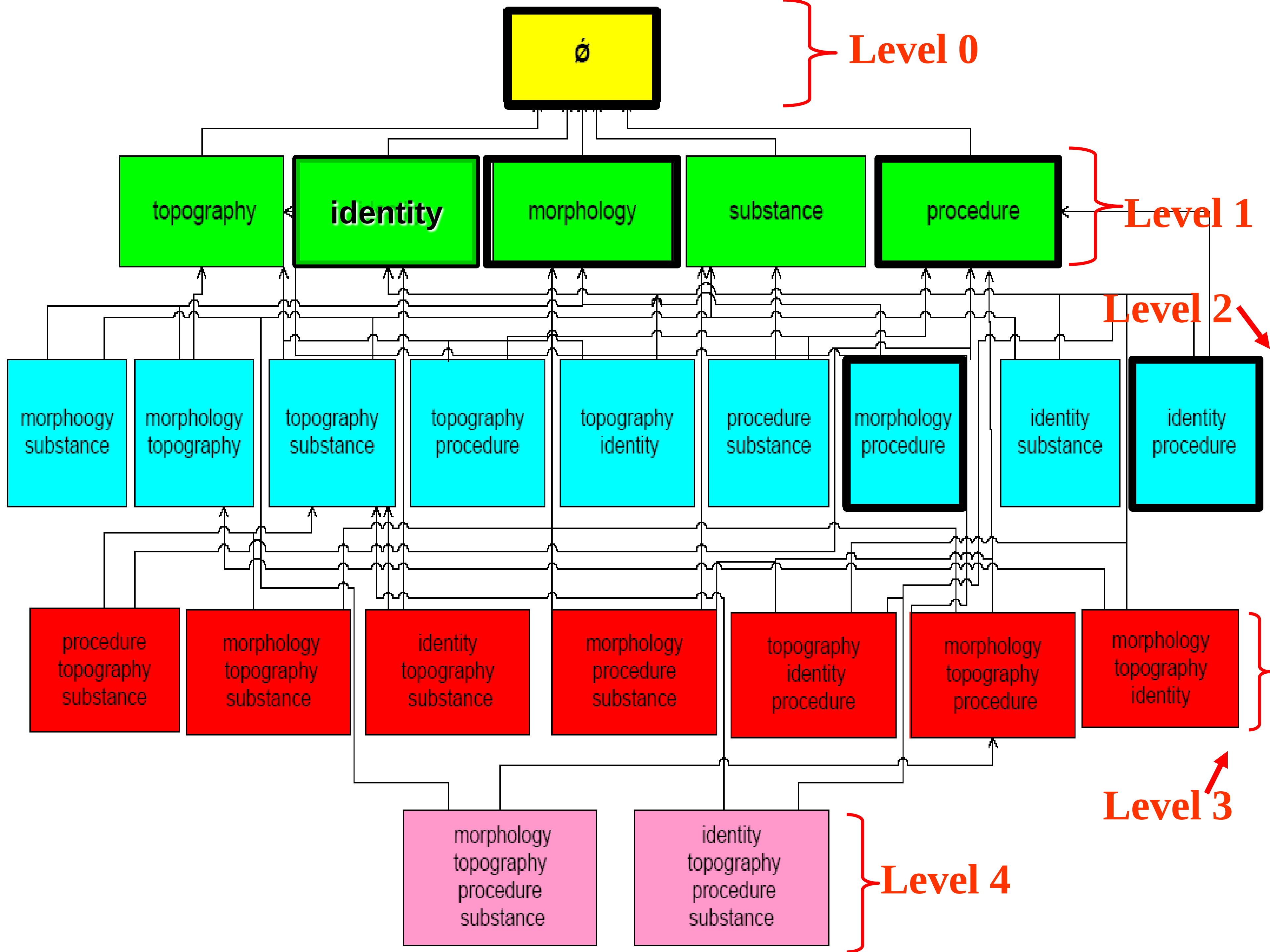
Structural Abstraction Network

Structure Uniformity Group (strUG) is a collection of all concepts with exactly same set of associative relationships (Structure).

Structure Uniformity Taxonomy is a hierarchical graph structure that consists of only the strUGs (as nodes) and hierarchical *child-of* relationships.

Excerpt of Structure Uniformity Taxonomy for Specimen Hierarchy (2007)





ø
Level 0
topography
identity
morphology
substance
procedure
Level 1
Level 2
morphoogy substance
morphology topography
topography substance
topography procedure
topography identity
procedure substance
morphology procedure
identity substance
identity procedure
procedure topography substance
morphology topography substance
identity topography substance
morphology procedure substance
topography identity procedure
morphology topography procedure
morphology topography identity
Level 3
morphology topography procedure substance
identity topography procedure substance
Level 4

Concepts in strUG({identity})

Arterial cannula tip
Arteriovenous shunt tip
Blood bag specimen
Blood bag specimen, from blood product
Blood bag specimen, from patient
Cannula specimen
Cannula tip
Catheter and cannula tip
Catheter specimen
Catheter tip specimen
Device specimen
Drain device specimen
Drain tip
Electrode specimen
Implantable venous catheter specimen
Plasma bag specimen
Specimen from patient
Tube specimen
Vascula cannula tip
Venous cannula tip

identity

Hierarchical Abstraction Network

A **Root** of an strUG is a concept, of the strUG, whose parents all reside in other strUGs.

Semantic Uniformity Group (smtUG) is a set of concepts comprising a single root and all its descendants within the strUG.

strUG({Identity})

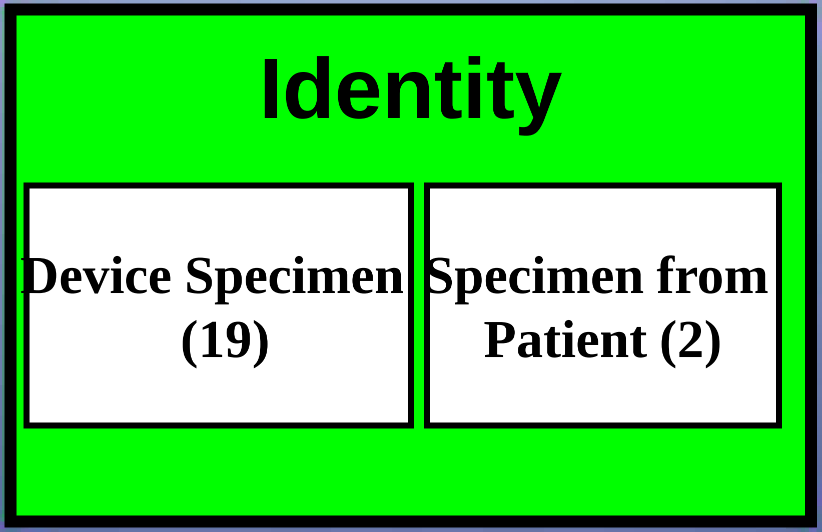
Device specimen
- Blood bag specimen
  - Blood bag specimen , from blood product
  - Blood bag specimen, from patient
- Cannula specimen
  - Catheter and cannula tip
    - Cannula tip
      - Vascular cannula tip
        - Arterial cannula tip
      - Arteriovenous shunt tip
      - Venous cannula tip
- Catheter specimen
  - Implantable venous catheter specimen
- Catheter tip specimen
- Drain device specimen
- Drain tip
- Electrode specimen
- Plasma bag specimen
- Tube specimen

Specimen from patient
- Blood bag specimen, from patient

All descendants of a root are subsumed by the root concepts.

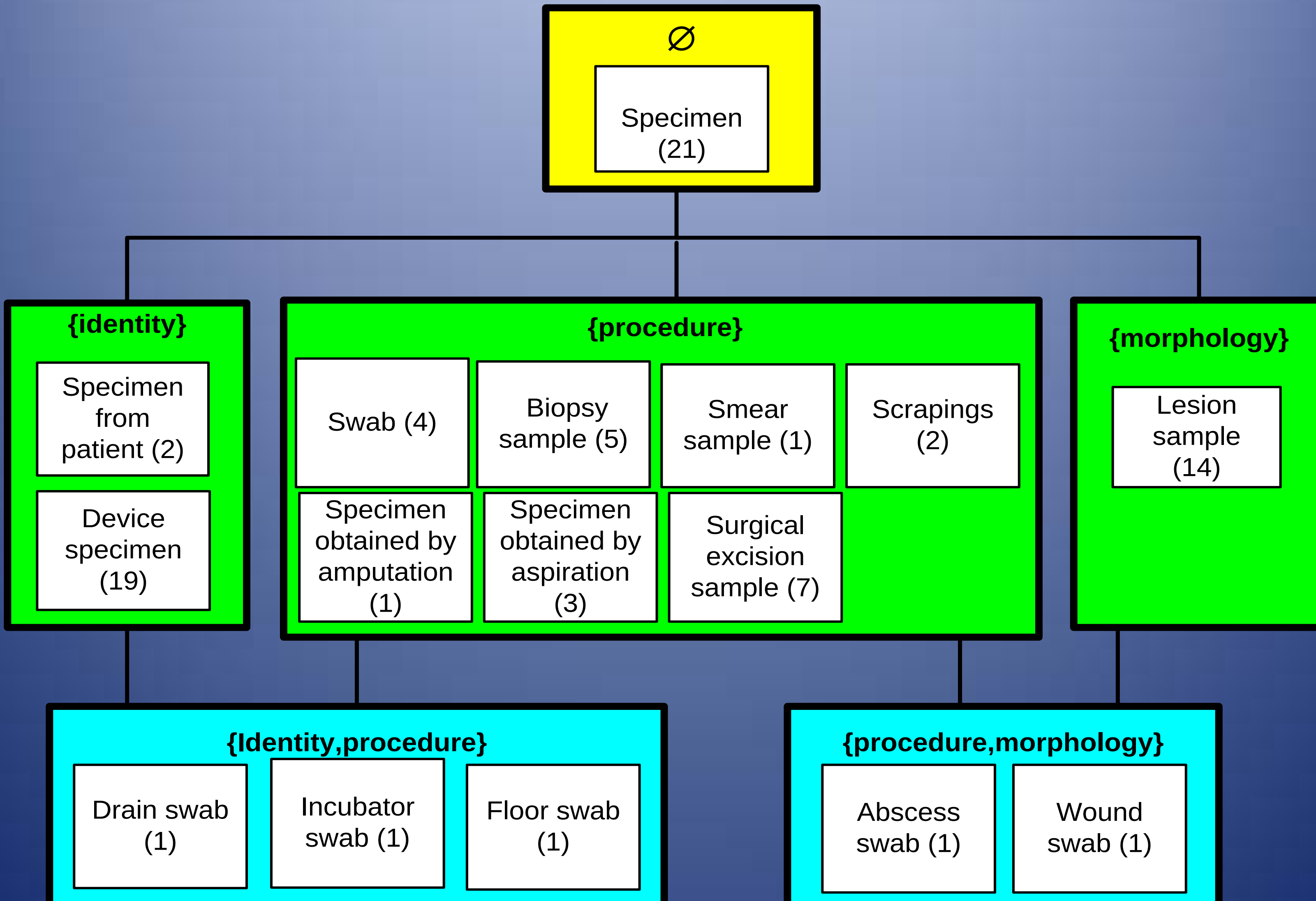
identity

Semantic Uniformity Groups in strUG({Identity})



This strUG({identity}) has concepts of 2 kinds: Device Specimen and Specimen from Patient.

Semantic Uniformity Taxonomy

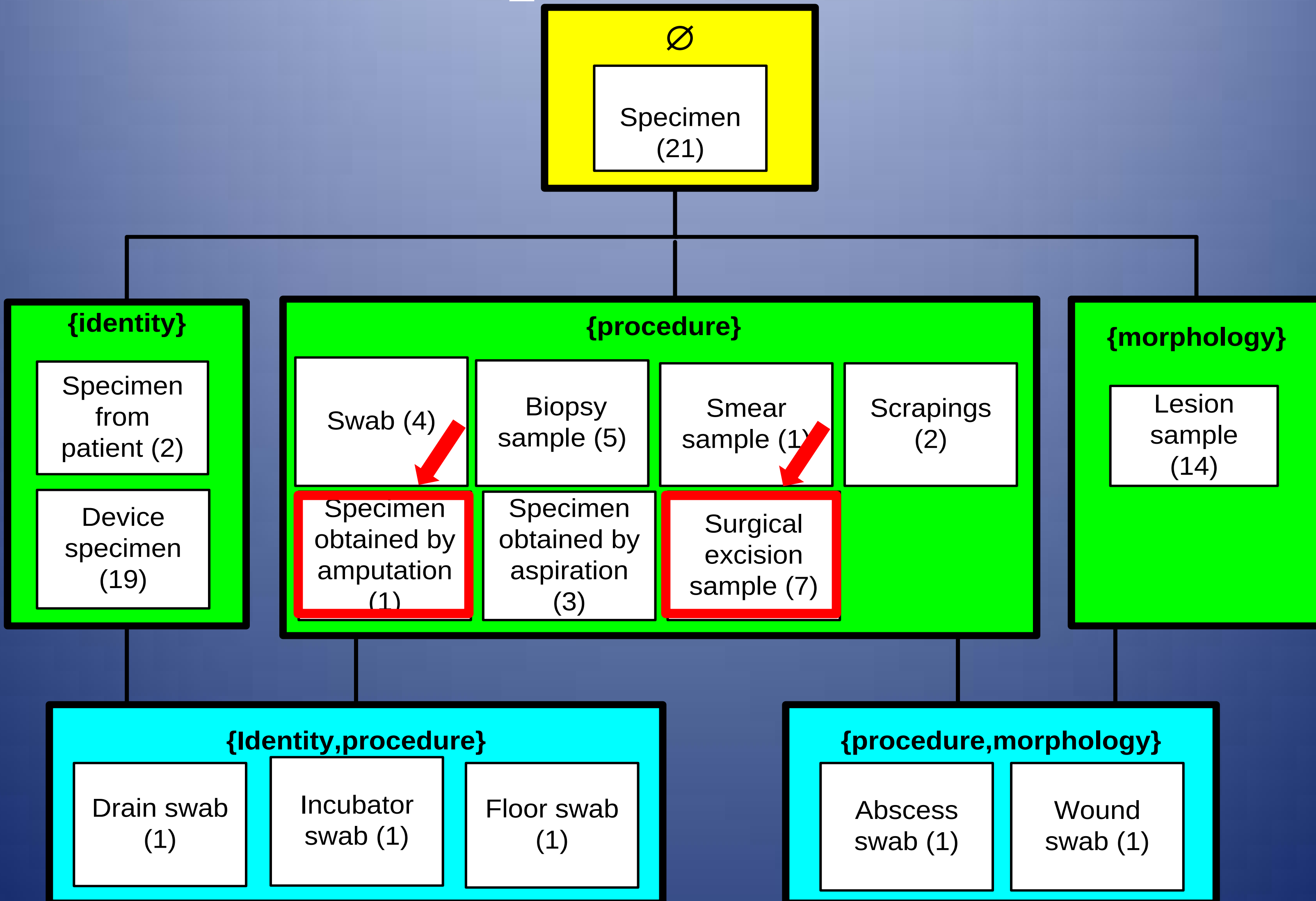


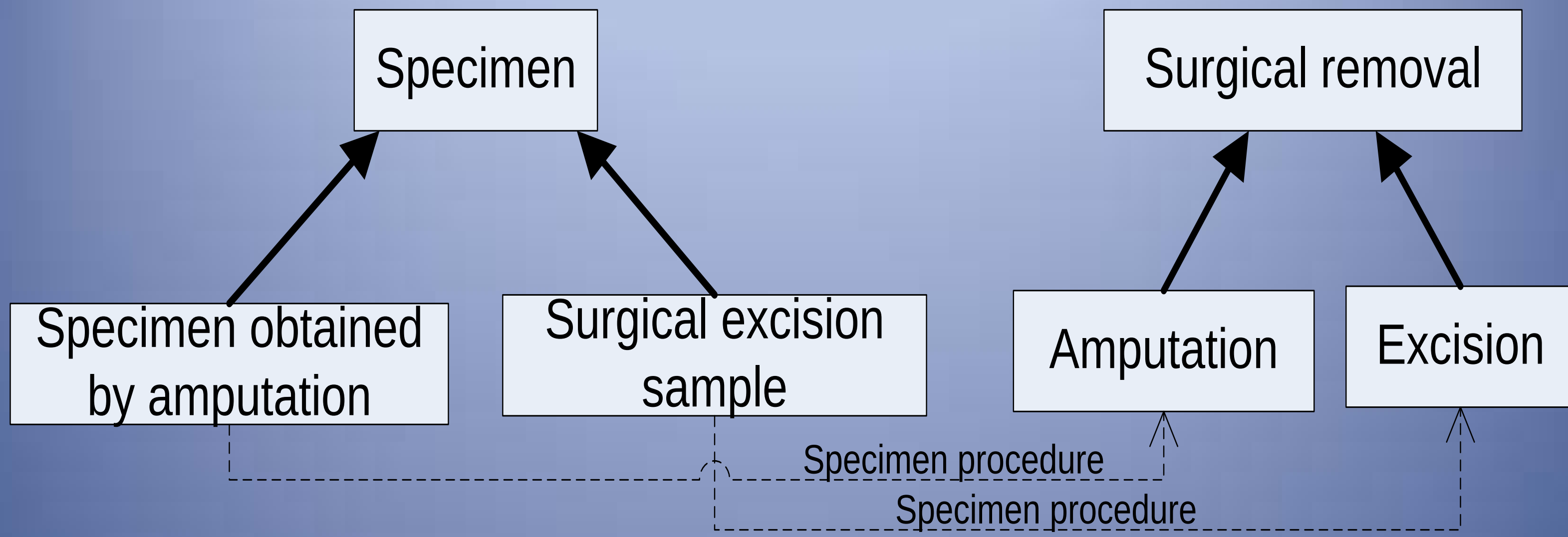
How AN helps error detection

Rule 1(***Disparity***) **In the same structural uniformity group, different semantic uniformity groups are not hierarchically related and they are referring to different semantic domain.**

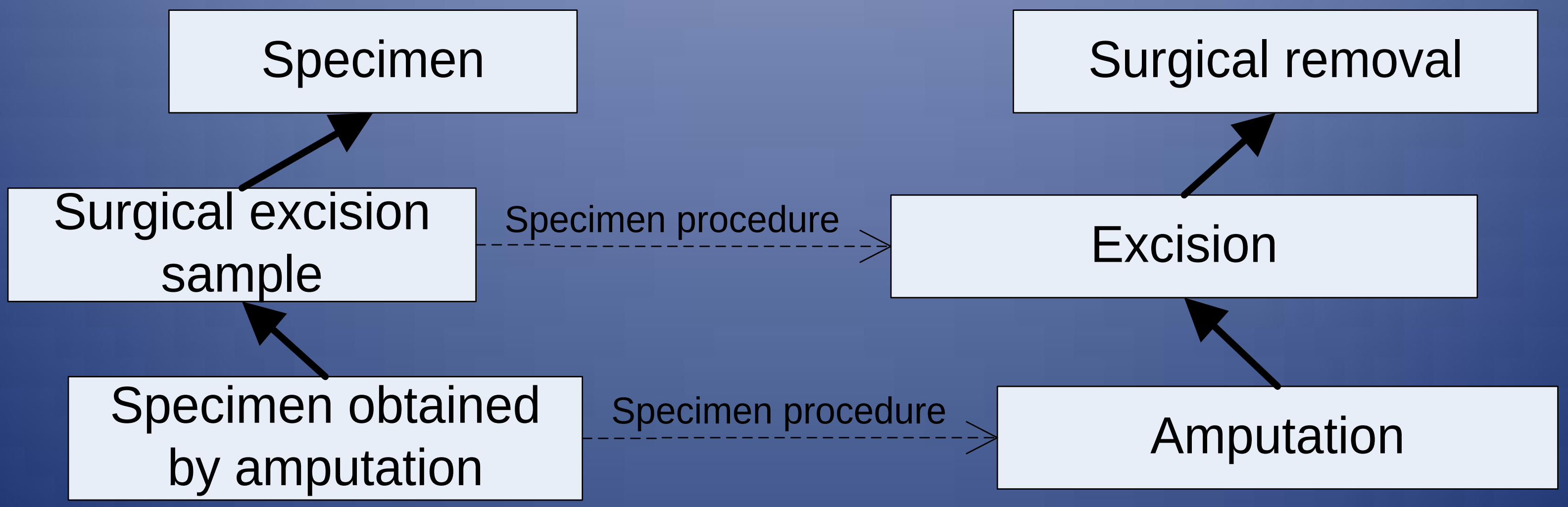
Interpretation: Roots within the same strUG are expected to have the same strucutre , but with great difference by nature.

Example for Rule 1





(a) Before correction (July 2008)



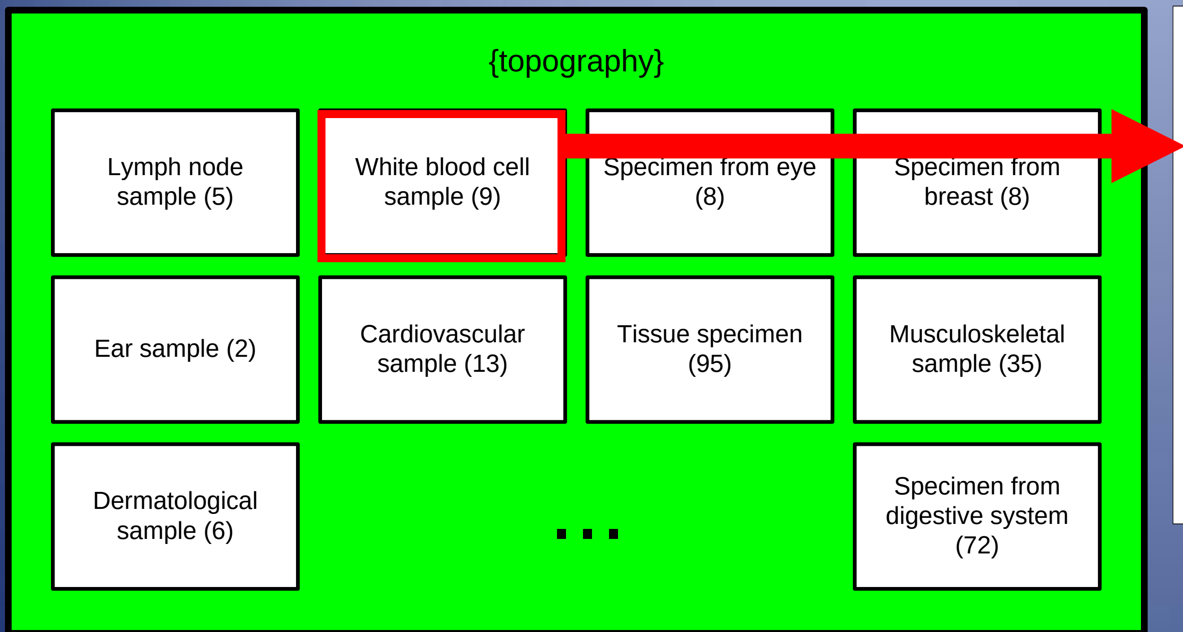
(a) After correction (July 2009)

How AN helps error detection (cont'd)

Rule 2(*Similarity*) In the same structural uniformity group, concepts within the same semantic uniformity group are hierarchically related.

***Interpretation*:** Concept in the same smtUG inherit from a common ancestor. This compact view helps the auditor to detect irregularities such as duplicate or missing.

Example for Rule 2



White blood cell sample
- Macrophage specimen
- Leukocyte specimen
  - Basophil specimen
  - Buffy coat
  - Lymphcyte specimen
  - Eosinophil specimen
  - Polymorphonuclear neutrophil specimen
  - Leukocyte specimen from control

Error: White blood cell sample and Leukocyte specimen are identical concepts

Proposed solution: one of which should be synonym of the other.

Auditing methods based on AN

- **Group-based auditing (in Wang, et al.)**

  Concepts of a semantic uniformity group were reviewed at the same time, so that concepts in the same group will provide contexts for each other

- **Error concentration based auditing (in Halper, et al.)**

  The method quantified the correlation between the size of the semantic uniformity group and the error concentration

- **Auditing "complex" concepts (in Wang, et al.)**

  A "complex" concepts is a concept inherited from parents belonging to more than one semantic uniformity group

Results for Group-based Auditing

- Fifty-four errors of different kinds were found
- These errors were reviewed by Chief Terminologist of SNOMED CT.
- All but four of the errors were confirmed and corrected in the subsequent release of the SNOMED CT

Results for Error Concentration Based Auditing

Table 1. Errors across ranges of semantic uniformity group size (# of concepts)

Size of the smtUG	# smtUG	#Concepts	#Errors	%Errors
1-7	427	646	69	10.68
8 or more	25	410	28	6.83
Total	**452**	**1,056**	**97**	**9.19**

- Semantic Uniformity Group in the former range are deemed to be “small”; while in the latter, large.
- As is seen from the table, the number is only 6.83% for large semantic uniformity group.

Results for Auditing “Complex” Concepts

Table 2. Errors and their counts by auditing “complex” concepts

Kinds of Errors	#
Ambiguous concept	1
Missing child	48
Missing parent	30
Missing relationship	21
Missing sibling	4
Incorrect child	5
Incorrect parent	44
Incorrect target of relationship	5
Total	**158**

Testing the Consistency of Suggested Changes

- Corrected the proposed errors and transformed the modification to owl format
- Reclassified the modified SNOMED owl with DL based reasoner FaCT++
- Reclassification took 17 min

  No warning or errors reported during loading and classification process

Technical Challenge: SNOMED owl is very large

So we use 64-bit Microsoft Windows OS with 4 GB of RAM

Discussion & Future Work

- Abstraction Network is based on the distributed version (inferred view) of SNOMED CT. AN can detect those errors that DL cannot identify.
- AN is independent of DL. It provides another framework for human reviewers to identify other kinds of inconsistencies.

Discussion & Future Work

- As for the future work, efforts should be made to minimize both Type I and Type II errors
- Type I error: it is hard to improve Type I because it related not only to the performance of DL classifier, but also the expressiveness of DL to represent concepts.
- Type II error: independent method such as Abstraction Network need further investigation.

Conclusion

- DL formalism is a necessary but not sufficient tool for quality assurance purposes
- Abstraction Network is one of the in independent methods in complement to DL to detect Type II errors

Reference

- Wang Y, Halper M, Min H, Perl Y, Chen Y, Spackman KA. *Structural methodologies for auditing SNOMED.* JBI. 2007; 40(5): 561-581

- Bodenreider O, Smith B, Kumar A, Burgun A. *Investigating subsumption in DL-based terminologies: a case study in SNOMED CT.* In: Hahn U, Schulz S, Cornet R, eds. Proc. KR-MED 2004. Whistler, Canada; 2004. p. 12-20

- Wang Y, Wei D, Xu Junchuan, Elhanan G, Perl Y, Halper M, Chen Y, Spackman KA, Hripcsak G. *Auditing Complex Concepts in Overlapping Subsets of SNOMED.* In J. Suermondt, R. S. Evans, and L. Ohno-Machado, editors, *Proceedings of the 2008 American Medical Informatics Association (AMIA) Annual Symposium*, pages 273-277, Washington, DC, November 2008.

Reference (cont'd)

- Halper M, Wang Y, Min H, Chen Y, Hripcsak G, Perl Y, Spackman KA. *Analysis of Error Concentrations in SNOMED.* In J. M. Teich, J. Suermondt, and G. Hripcsak, editors, *Proceedings of the 2007 American Medical Informatics Association (AMIA) Annual Symposium*, pages 314-318, Chicago, IL, November 2007.

Acknowledgement

- Dr. Bodenreider
- Dr. Fung
- Dr. Perl
- Dr. McDonald
- Ms. May Cheh
- LHNCBC
- All my colleagues and friends


Thank you!