Leveraging SNOMED CT Relationships for Optimizing Electronic Medical Records Data

Sina Madani MD, PhD[1] , Jerry Henderson, MD[2]

Shari Just, RN, BSN[1], Scott Nelson, PharmD, MS[1], Asli Weitkamp, PhD[1]

[1]Vanderbilt University Medical Center, Department of HealthIT

[2]MD Anderson Cancer Center, Department of Institutional Analytics & Informatics

Agenda

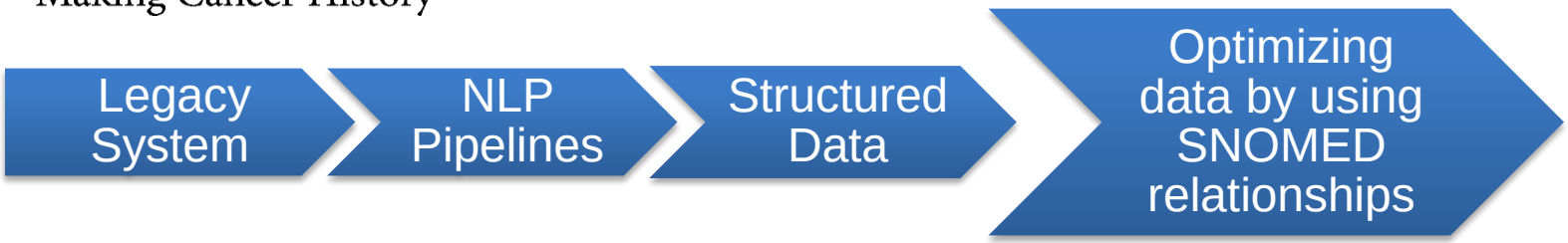
- Describe how we used SNOMED CT ontological structure to optimize patient data in 2 different settings

1. Structured data generation and optimization at MD Anderson Cancer Center

1. Structured data conversion between two EMR systems at Vanderbilt University Medical Center

Settings

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®



VANDERBILT UNIVERSITY
MEDICAL CENTER

The University of Texas MD Anderson Cancer Center

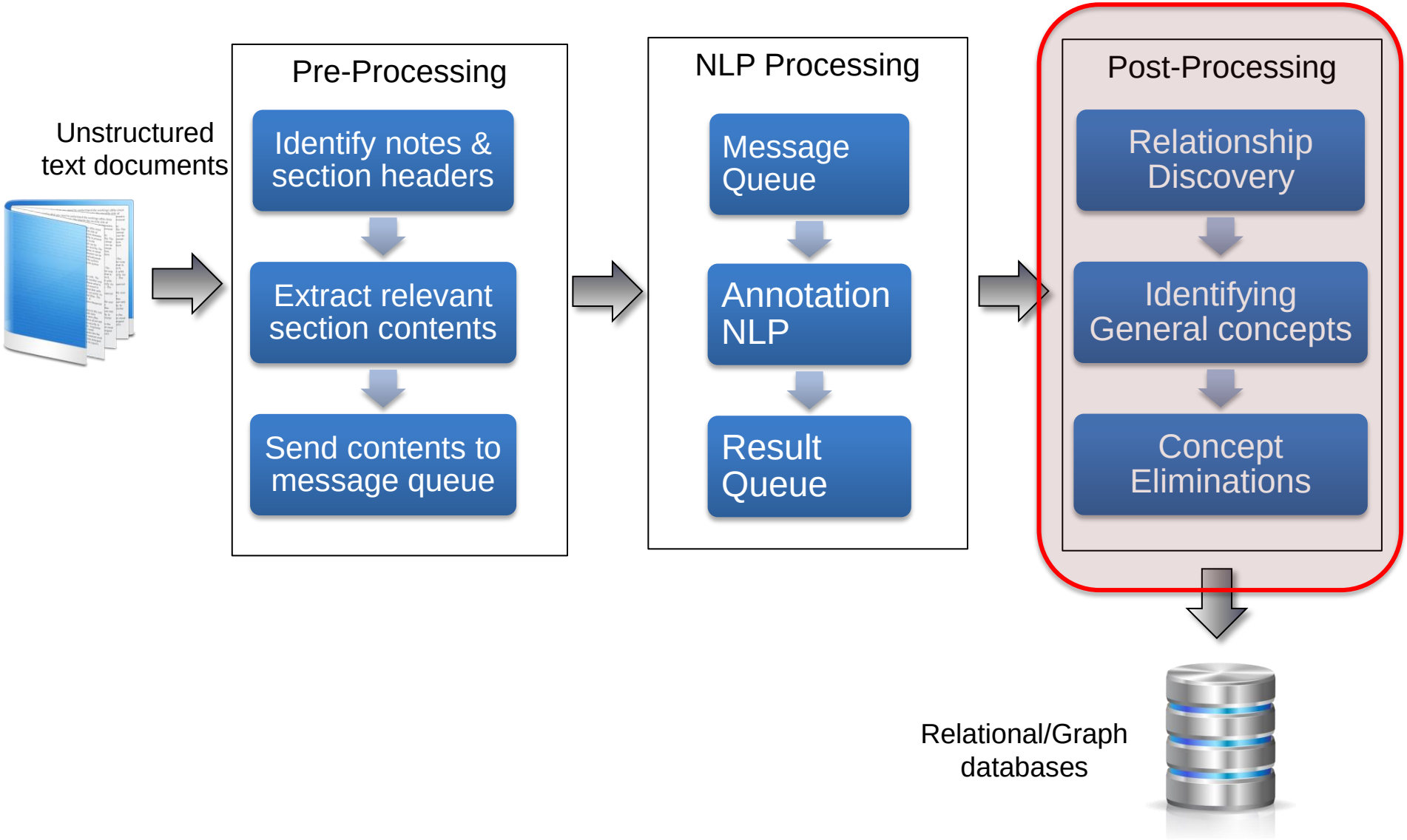
- One of the largest cancer centers in the country
- The legacy EMR system contained > 17 million transcribed documents
- ~160,000 notes/month enters into the EMR
- Dictation (free text) is the input format of choice
- March 2016: Transition from a homegrown EMR to Epic

Patient Problem List

- *Objective:* to extract active problem list for each patient using existing text documents

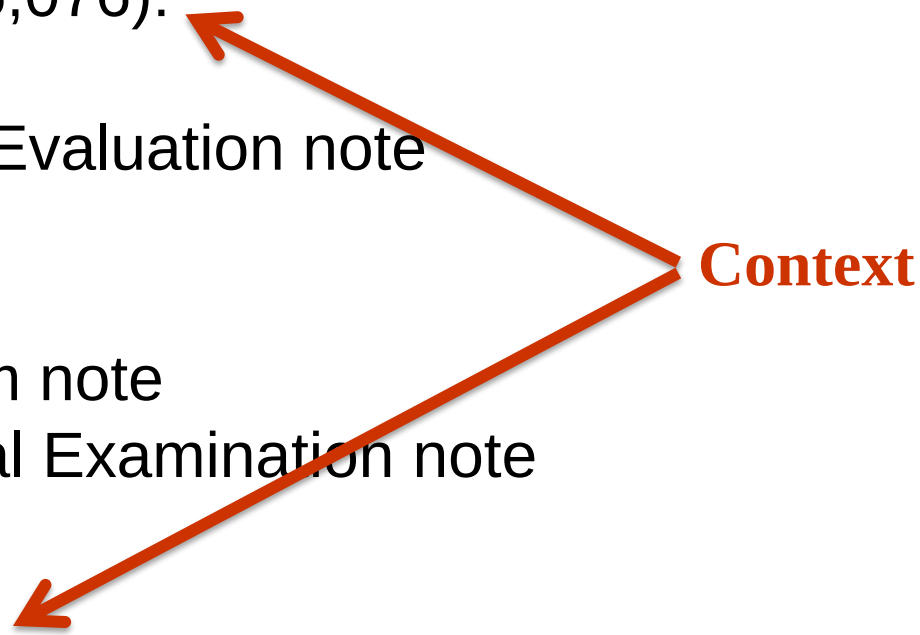
- Import *problem list* for each patient into Epic EMR system
    - Structured
    - Mapped to SNOMED CT codes
    - Maintained as an institutional master problem list
    - Customizable with institutional specific *terms/synonyms*

NLP Framework at MD Anderson

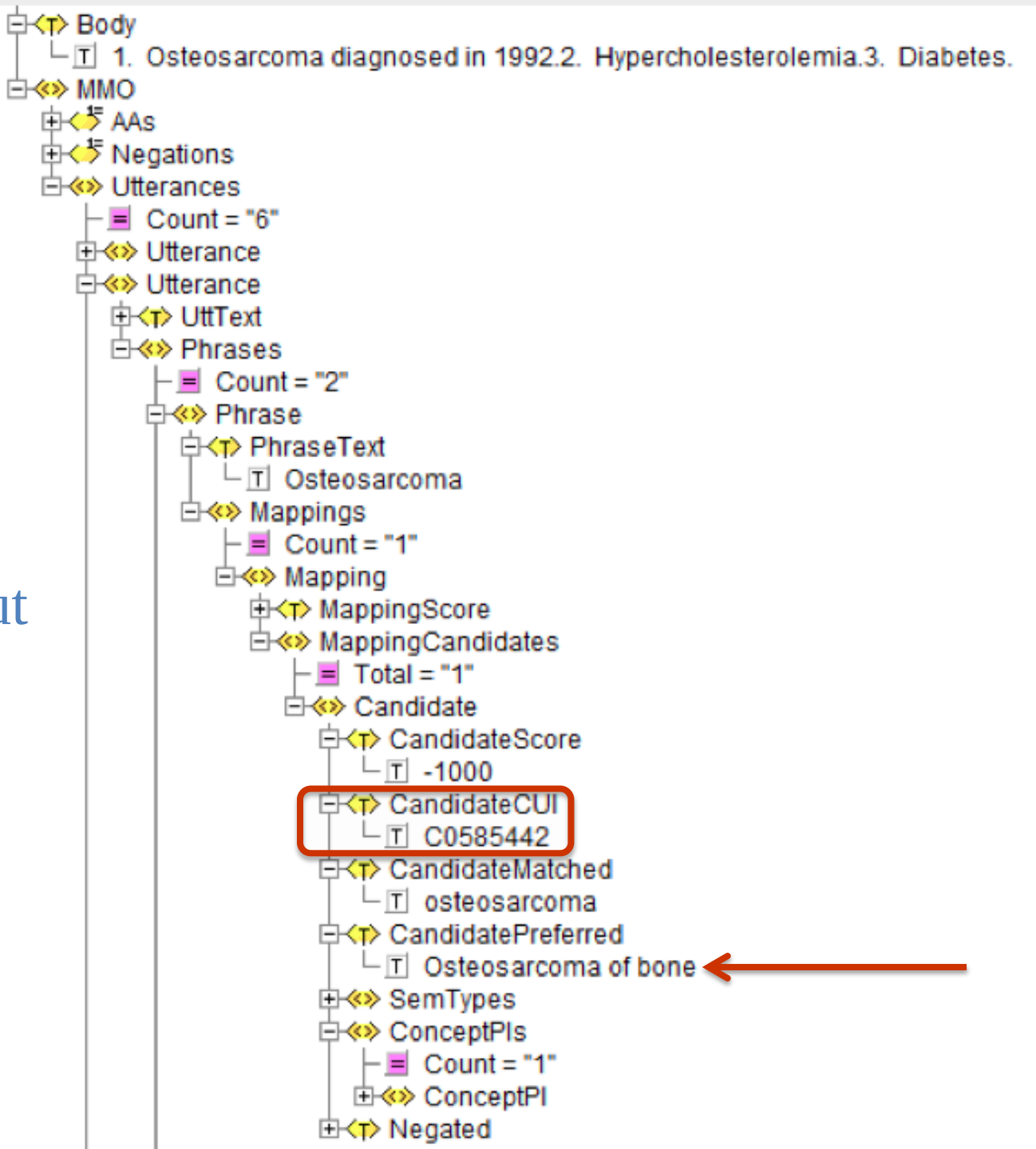


Extraction Methods

- ~ 17 Million unstructured free-text patient notes
  - Target note types (5,636,076):
    - Primary Medical Evaluation note
    - Progress note
    - Consult note
    - Emergency Room note
    - History & Physical Examination note
  - Target section headers:
    - Assessment & Plan

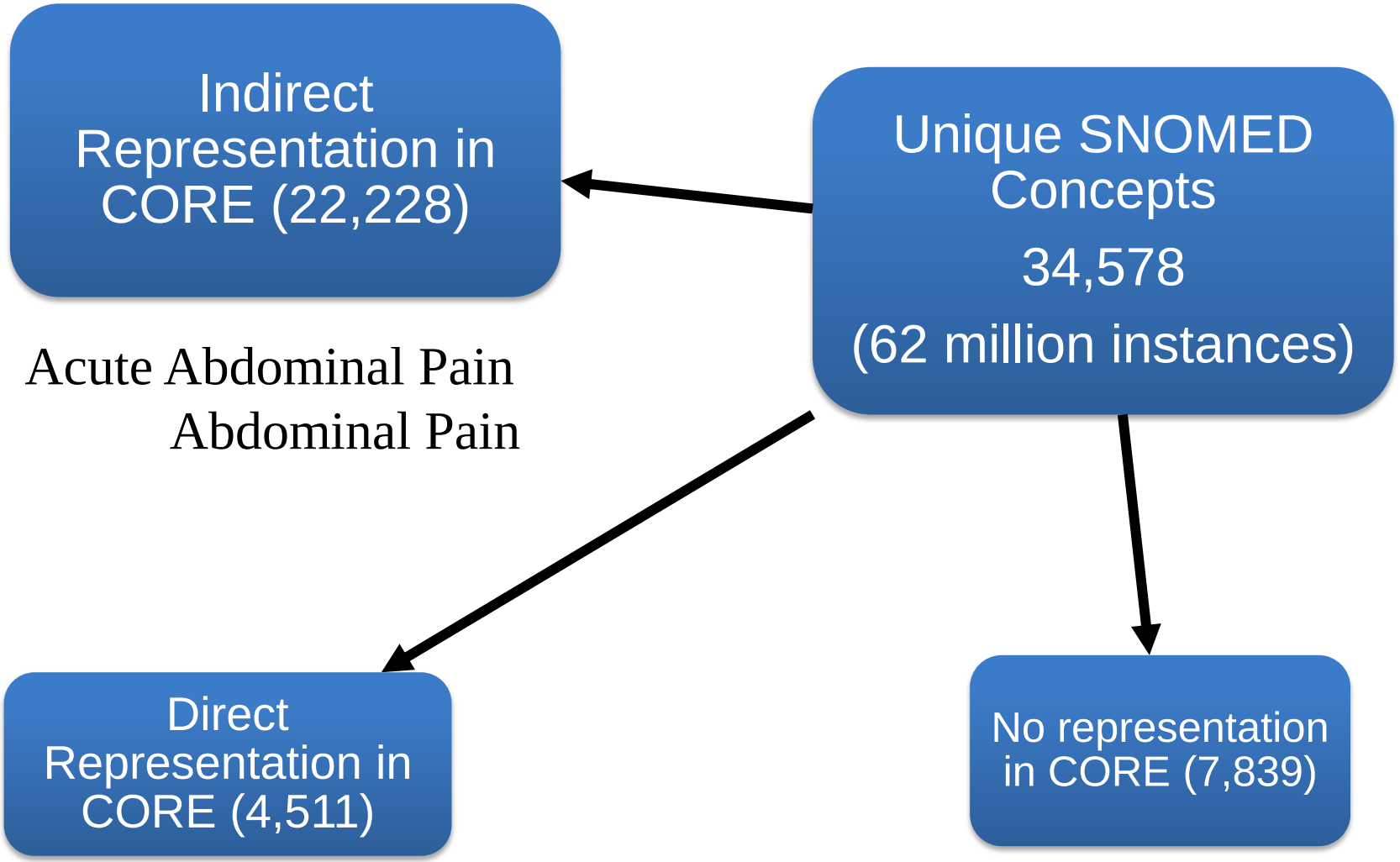


Sample NLP Output



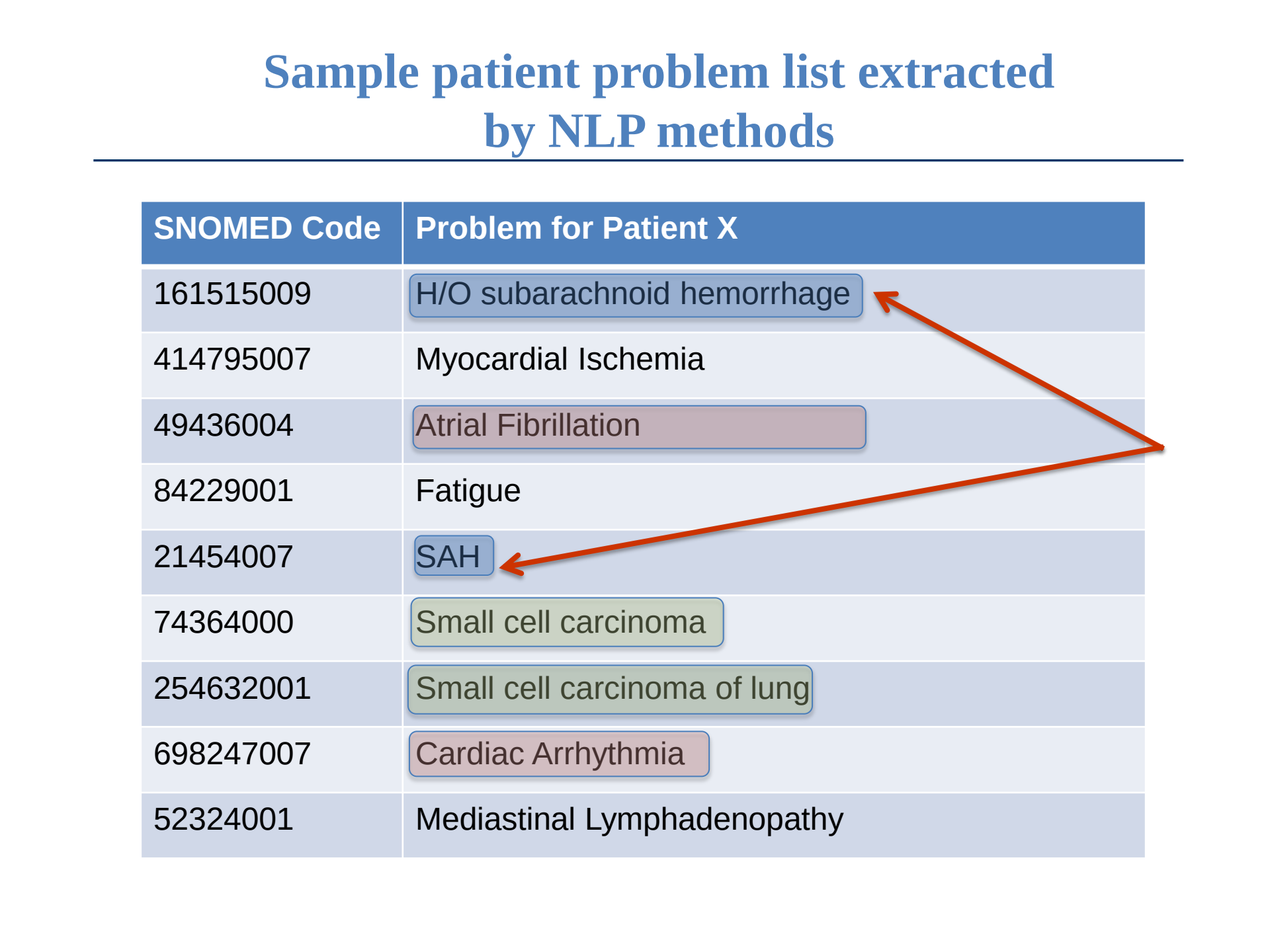
Clinical Observation Recording and Encoding (CORE)

Created to facilitate the use of SNOMED CT for problem list and other documentation purposes – project researcher : Dr. KW Fung

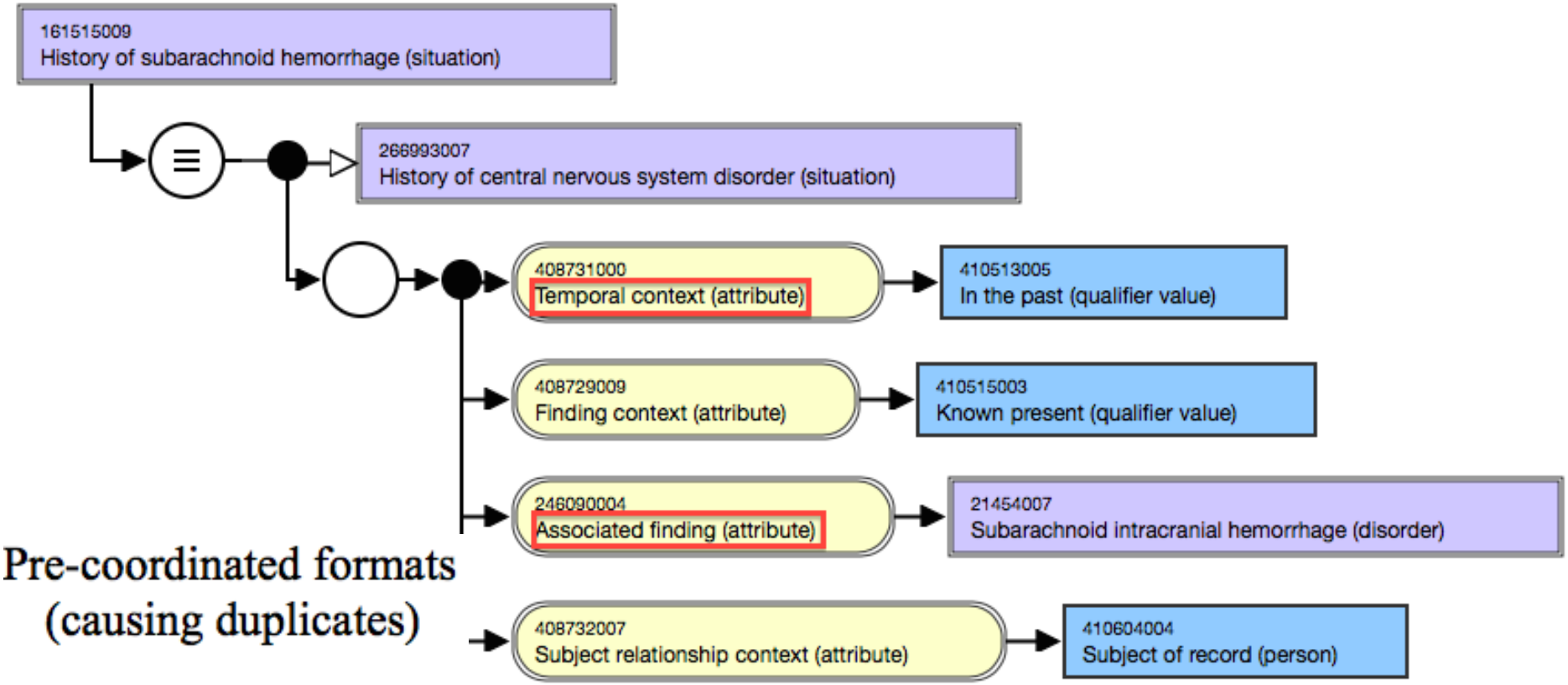


Sample patient problem list extracted by NLP methods

SNOMED Code	Problem for Patient X
161515009	H/O subarachnoid hemorrhage
414795007	Myocardial Ischemia
49436004	Atrial Fibrillation
84229001	Fatigue
21454007	SAH
74364000	Small cell carcinoma
254632001	Small cell carcinoma of lung
698247007	Cardiac Arrhythmia
52324001	Mediastinal Lymphadenopathy



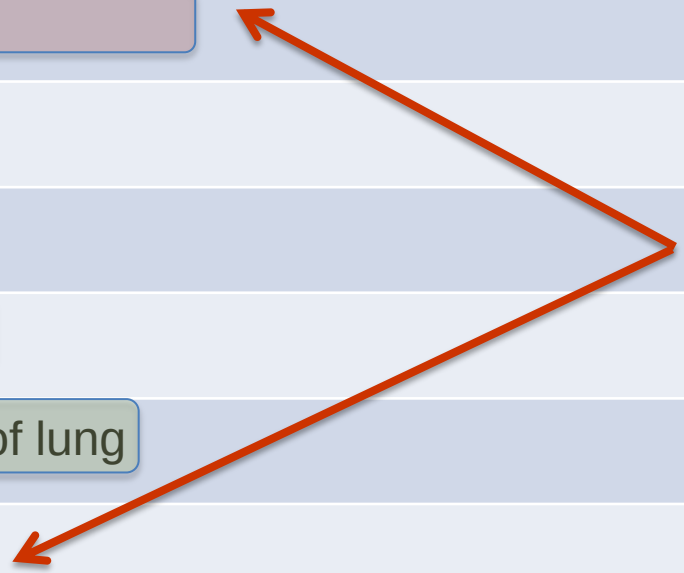
Eliminating similar concepts in patient problems (post-processing)



Source Concept CUI	Source Concept Name	target Concept CUI	target UMLS Name	Target SNOMED Code
C0455618	H/O cardiac surgery	C0018821	Operation on heart	64915003
C0457969	H/O Malignant melanoma	C0025202	Malignant melanoma	372244006
C0581908	H/O orchidectomy	C0029189	Testis excision	120001005
C0455613	H/O Spinal surgery	C0185908	Operative procedure on spinal structure	392236004
C0581077	H/O splenectomy	C0037995	Splenectomy	234319005
C0475872	H/O subarachnoid hemorrhage	C0038525	Subarachnoid hemorrhage	21454007

Sample patient problem list
Extracted by NLP methods

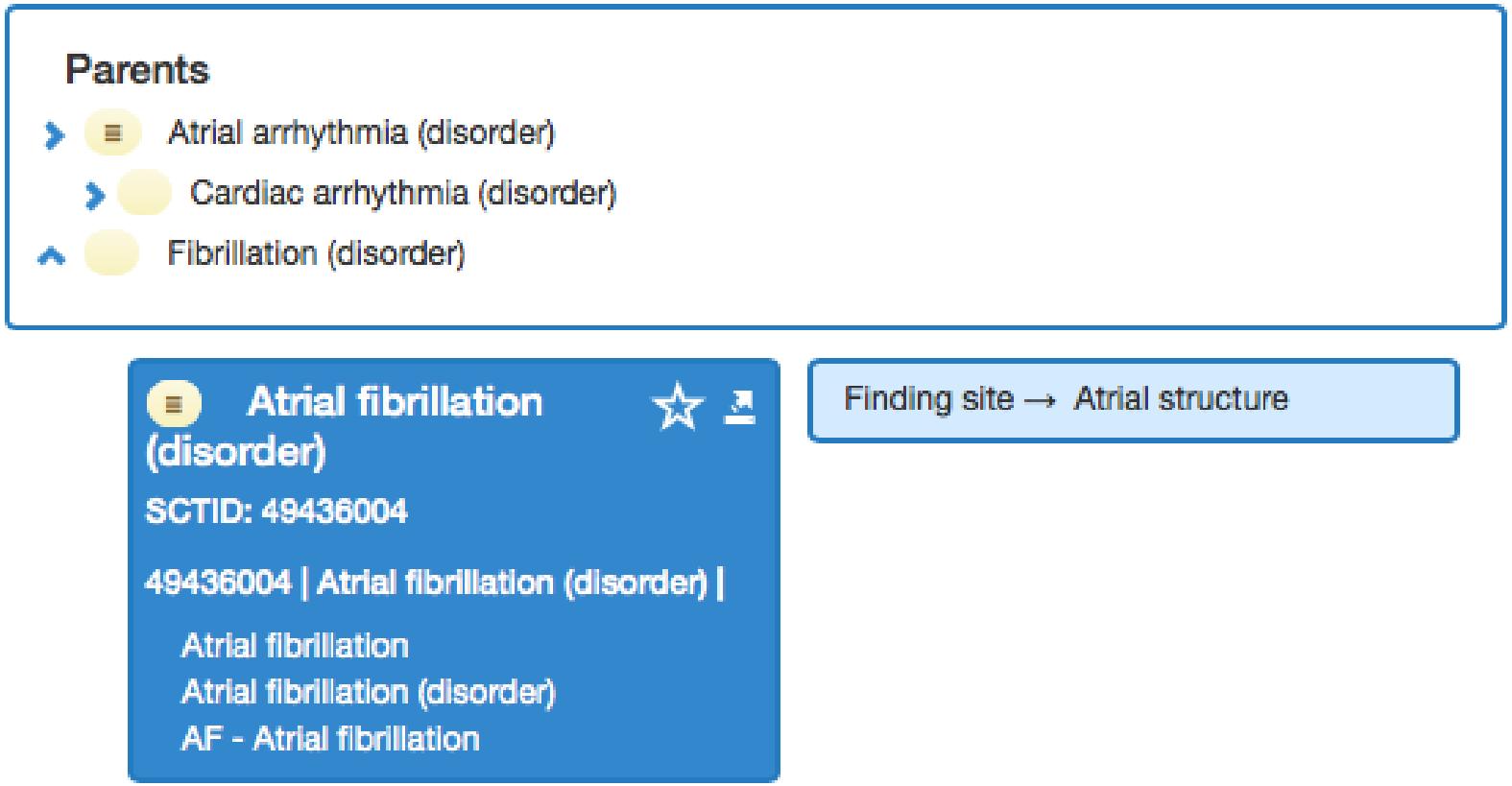
SNOMED Code	Problem for Patient X
161515009	H/O subarachnoid hemorrhage
414795007	Myocardial Ischemia
49436004	Atrial Fibrillation
84229001	Fatigue
21454007	SAH
74364000	Small cell carcinoma
254632001	Small cell carcinoma of lung
698247007	Cardiac Arrhythmia
52324001	Mediastinal Lymphadenopathy



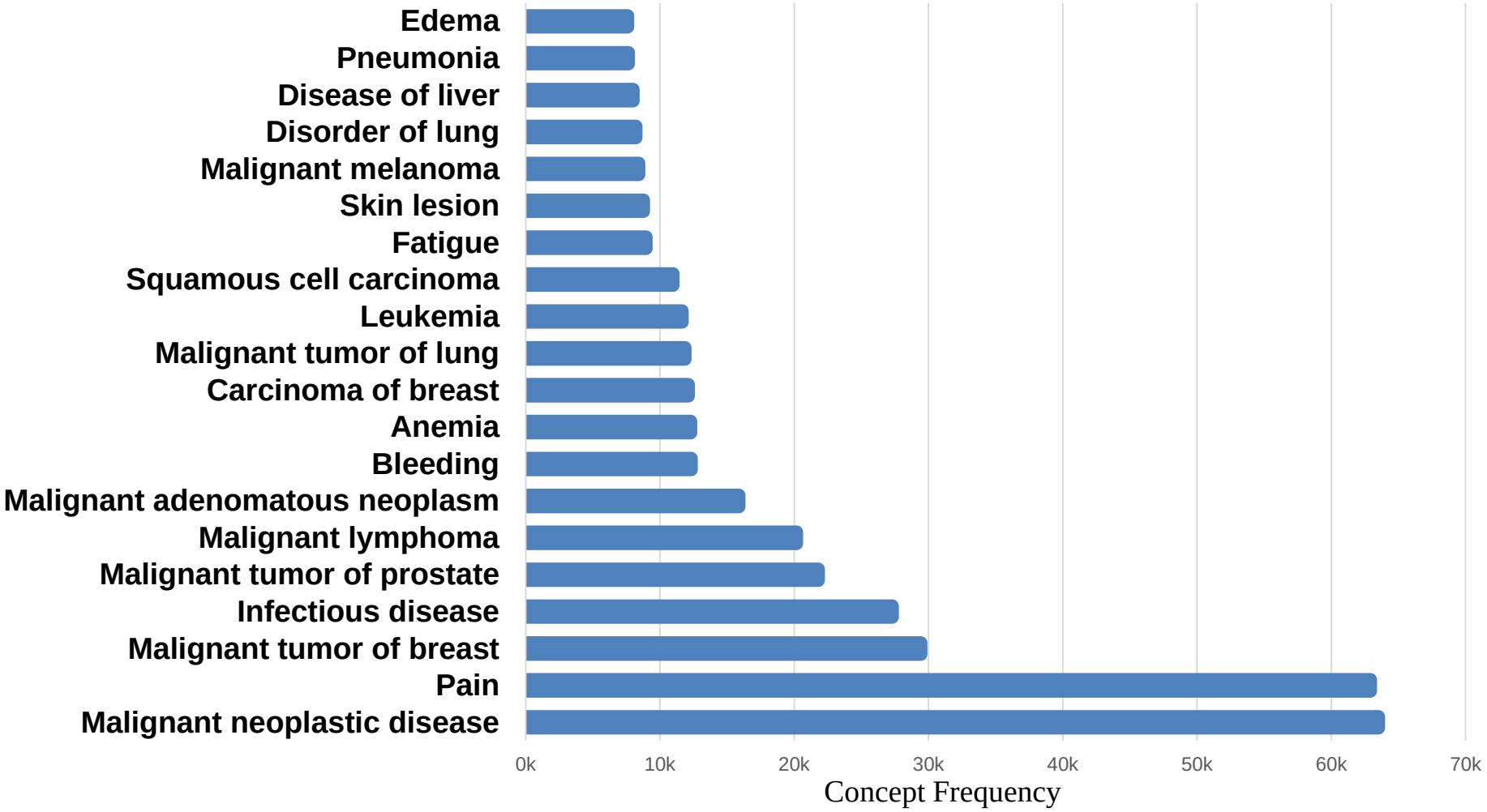
Eliminating more general concepts by using *Parent - Child relationship*

Recommended practice for documenting problem list:

Document the most detailed form of the problem



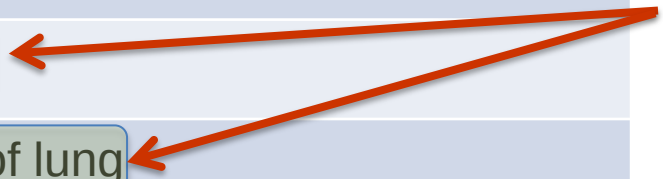
Top 20 concepts that were eliminated based on Parent-Child relationship



*SNOMED concepts eliminated based on parent-child relationships:*803,730

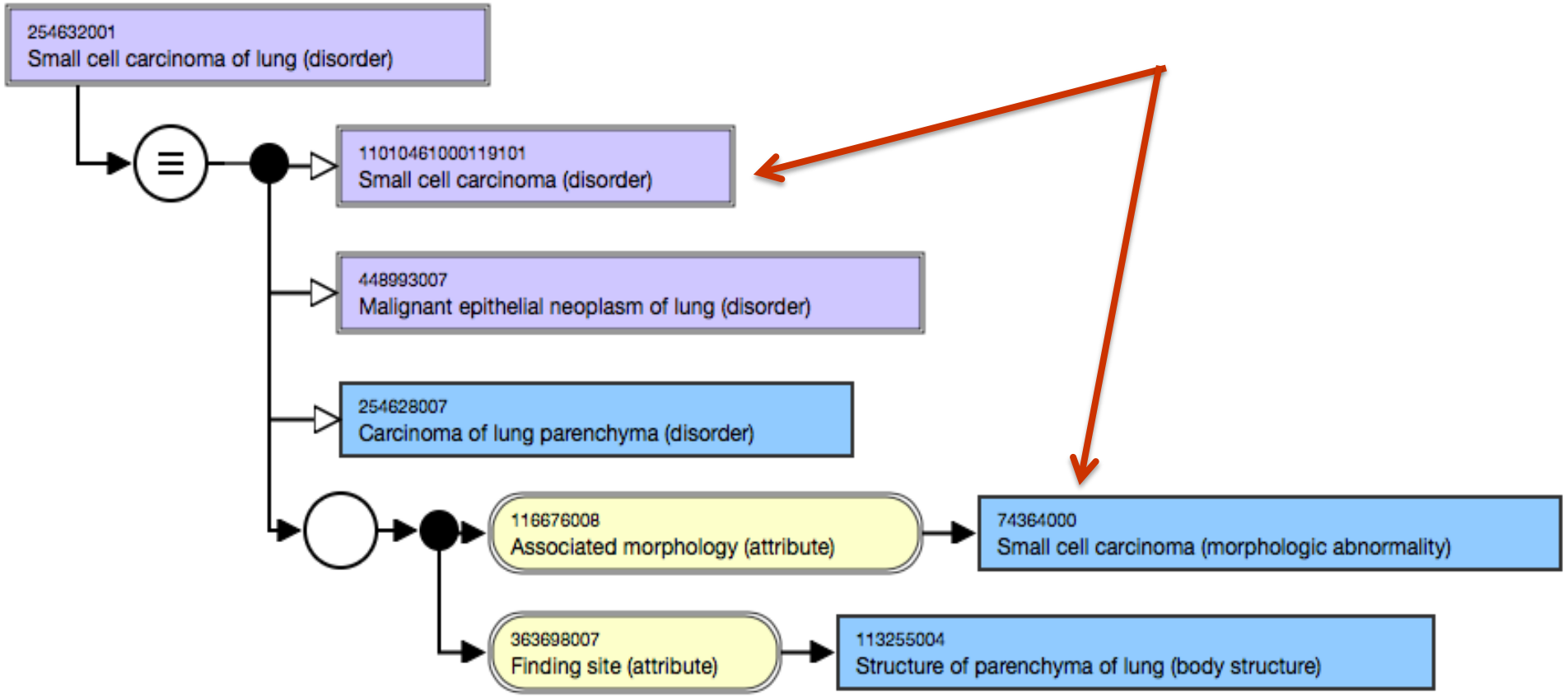
Sample patient problem list extracted by NLP methods

SNOMED Code	Problem for Patient X
161515009	H/O subarachnoid hemorrhage
414795007	Myocardial Ischemia
49436004	Atrial Fibrillation
84229001	Fatigue
21454007	SAH
74364000	Small cell carcinoma
254632001	Small cell carcinoma of lung
698247007	Cardiac Arrhythmia
52324001	Mediastinal Lymphadenopathy

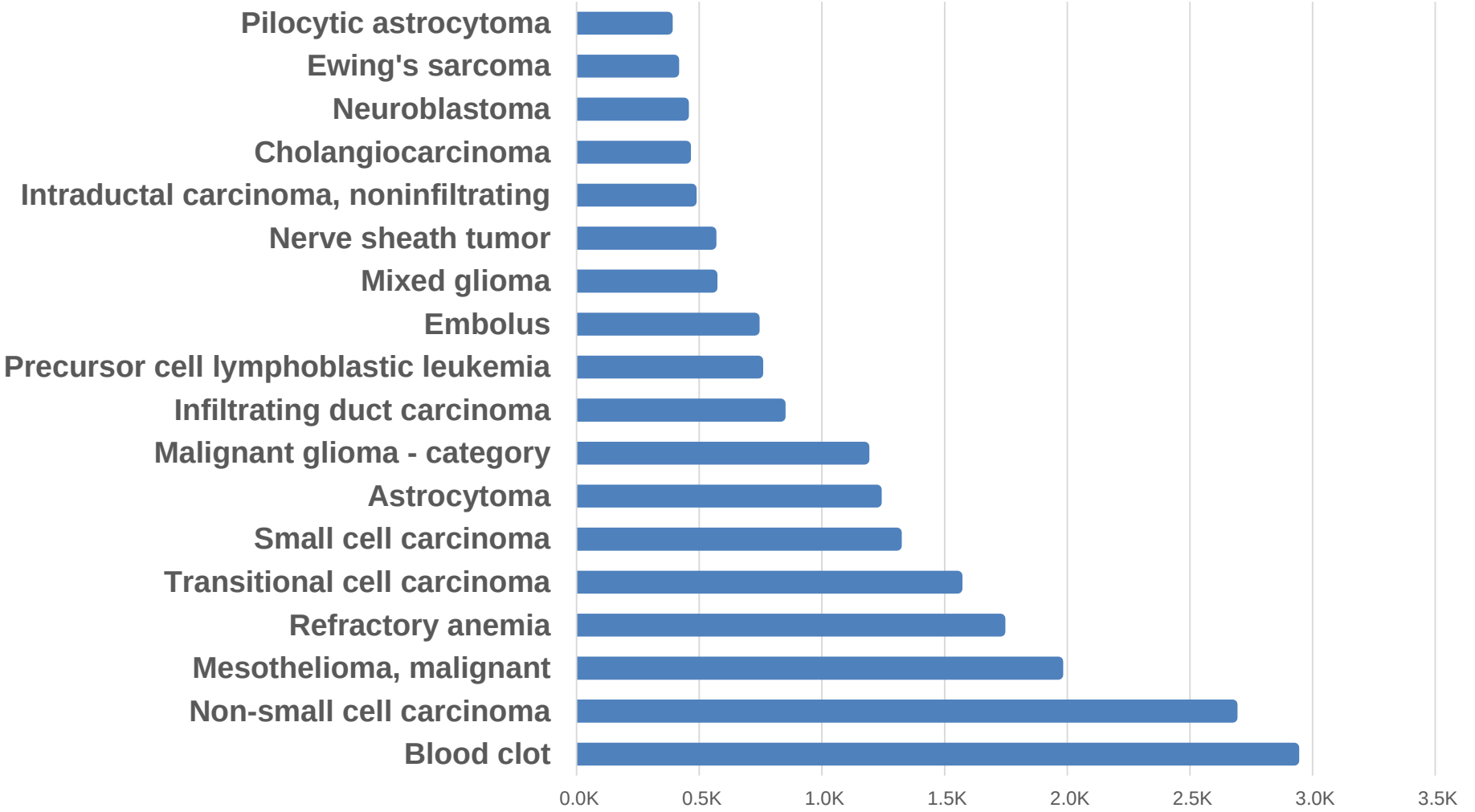


Eliminating more general concepts with *horizontal/attribute relationships*

Disease – Morphology relationship in SNOMED



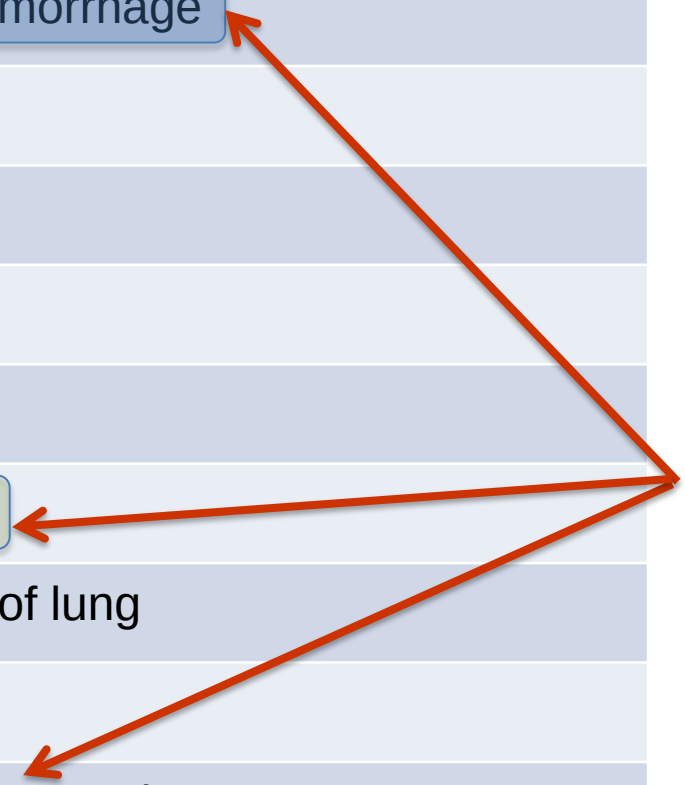
Top 20 Morphologic Abnormalities that were eliminated based on Disease-Morphology link



SNOMED Concepts eliminated based on disease-morphology relationship: 26,906

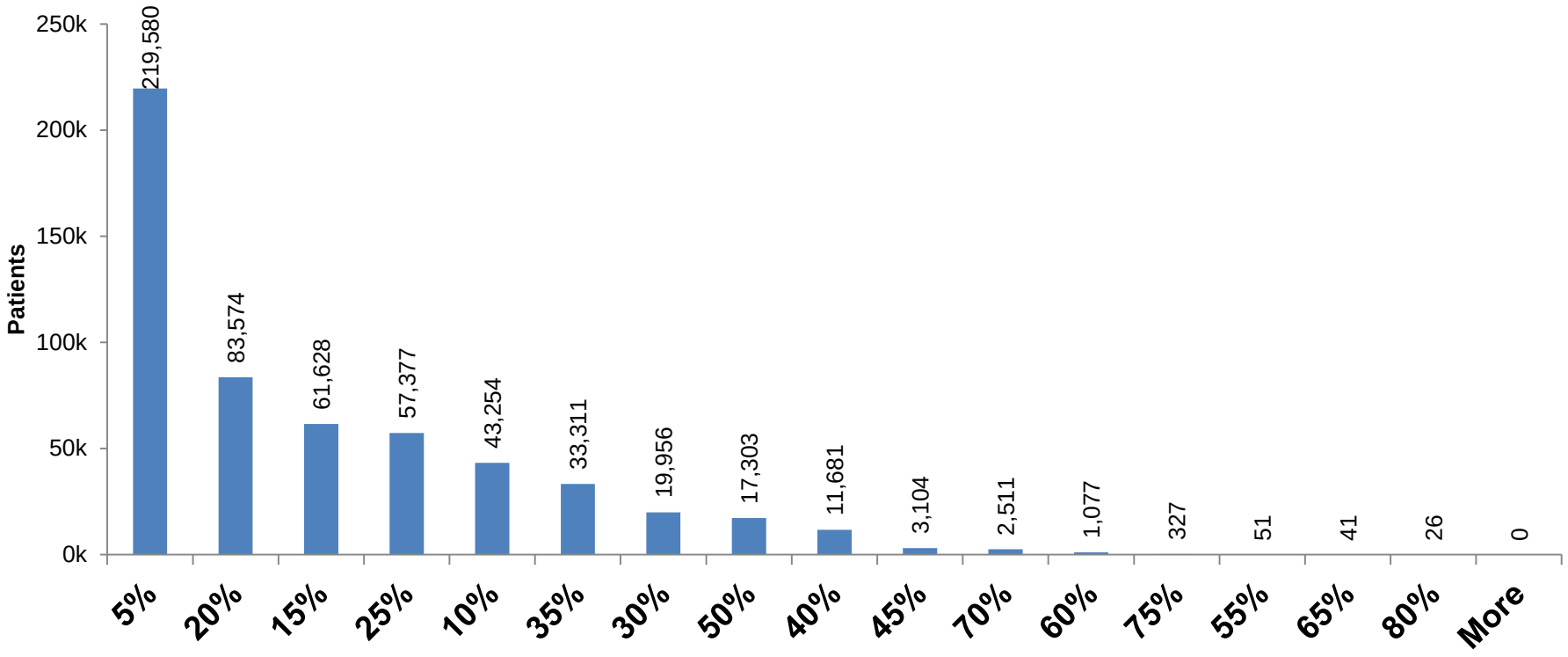
Eliminating redundant (more general) concepts

SNOMED Code	Problem for Patient X
161515009	H/O subarachnoid hemorrhage
414795007	Myocardial Ischemia
49436004	Atrial Fibrillation
84229001	Fatigue
21454007	SAH
74364000	Small cell carcinoma
254632001	Small cell carcinoma of lung
698247007	Cardiac Arrhythmia
52324001	Mediastinal Lymphadenopathy



33 % reduction in the length of the Problem List

Reduction (%) in the length of the problem list

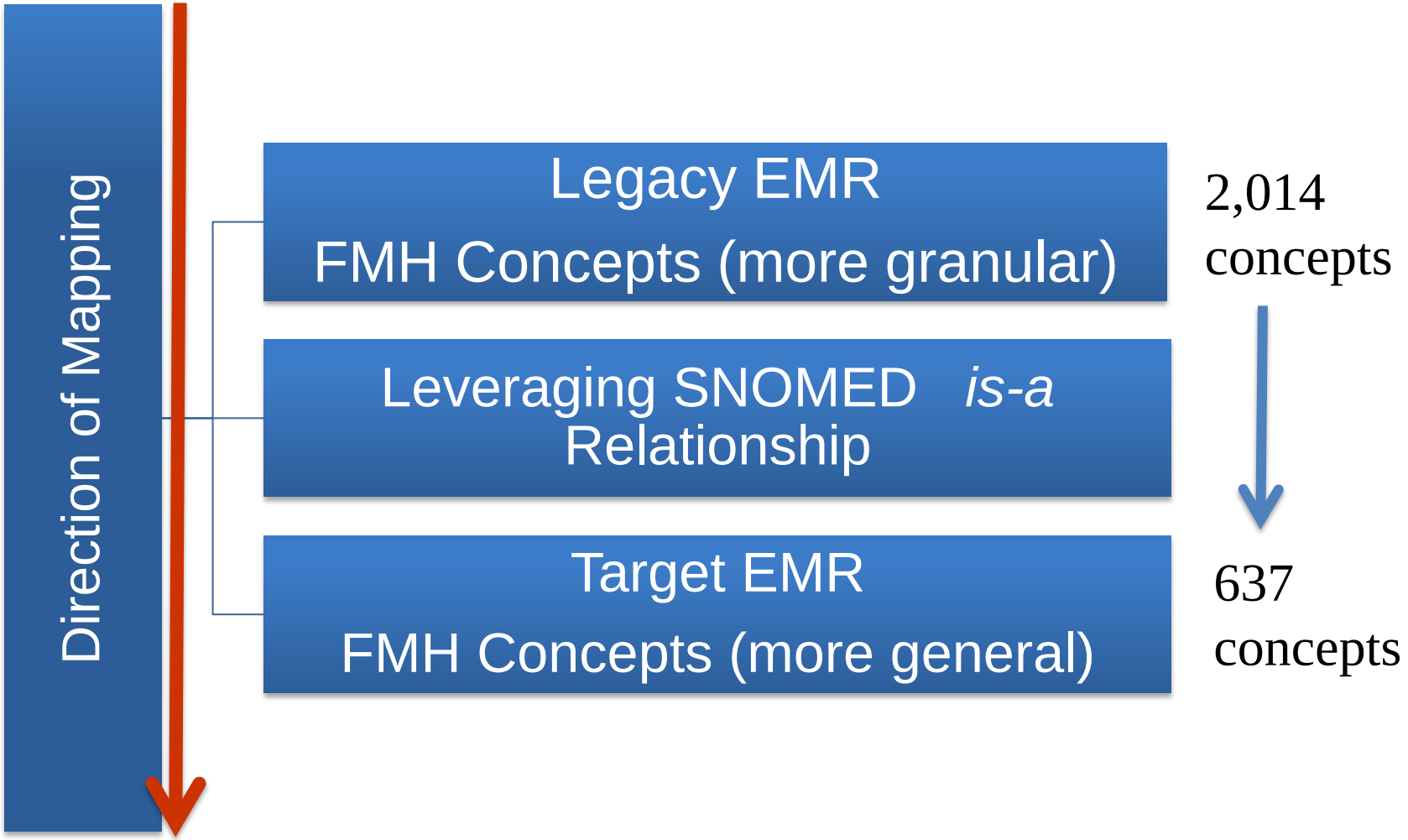


- Number of patients with more than 2 problems: 496,901
- Number of problem lists with a reduction between 25%-75%: 146,739

Vanderbilt University Medical Center

- Transition from a home grown EMR system
- Rapid and consistent conversion of clinical data is needed
- The legacy EMR system contains structured data
  - All concepts under family medical history module are mapped to 2,014 unique SMOMED codes
  - No direct match to the target EMR concepts
- Target EMR system contains limited SNOMED codes with lower granularities (637 concepts)
- **Objective**: *To identify relevant concepts between two data sets with different levels of granularities*

Mapping Schema between legacy and new EMR systems

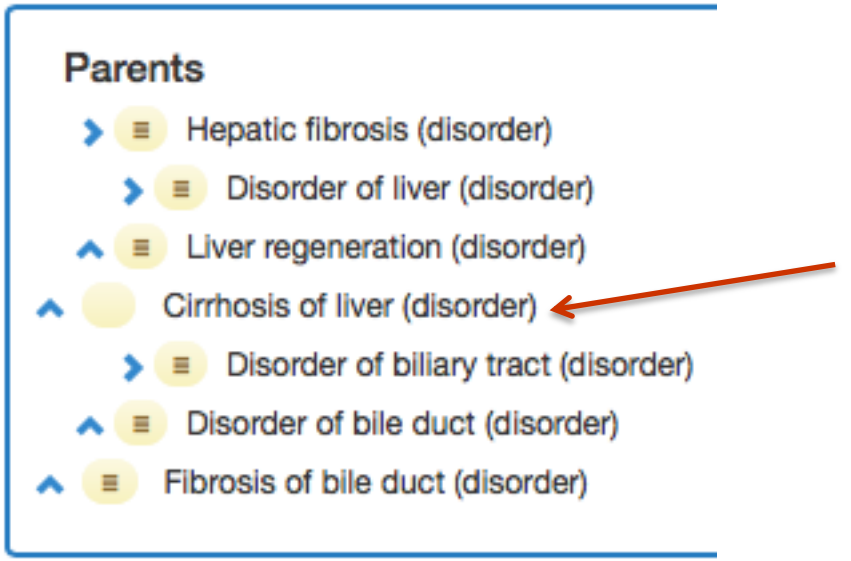


Methods

- A graph version of the SNOMED CT was used (20166AB) to extract all parents (super classes) of 2,014 SNOMED concepts in the legacy EMR system
- The distance between a given concept and its direct/indirect parent(s) was calculated
- Retired concepts were excluded or replaced with active concepts
- Extracted parent concept codes from the legacy dataset were matched to the equivalent concepts in the target dataset
- Subject matter experts evaluated matching based on the distance

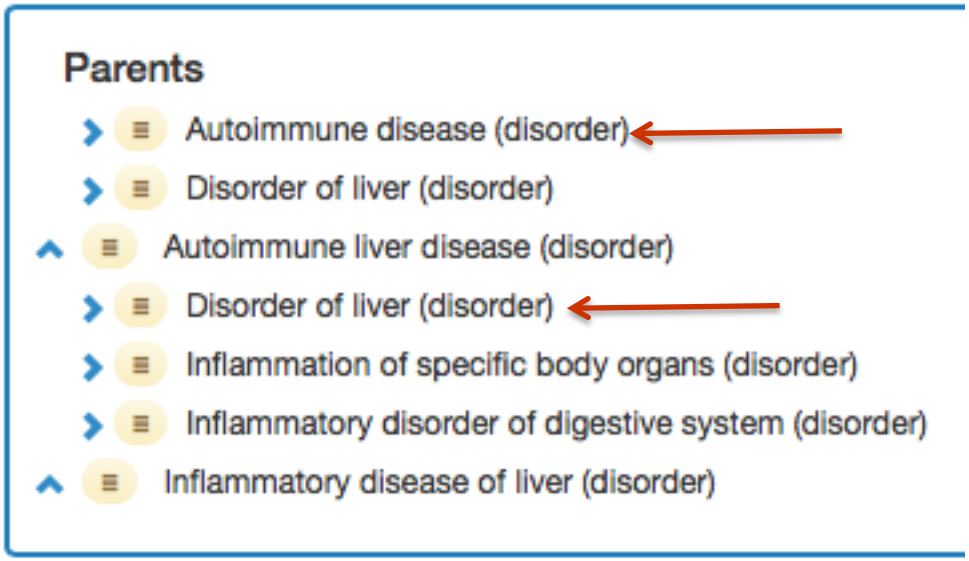
Closest parent match: Accept

Legacy EMR System			Target EMR System Match	
SNOMED Concept	SNOMED Parents	Distance	SNOMED Concept	SME Response
Biliary cirrhosis	Cirrhosis of liver	1	Cirrhosis of liver	Accept
Biliary cirrhosis	Fibrosis of bile duct	1		
Biliary cirrhosis	Hepatic fibrosis	2		
Biliary cirrhosis	Disorder of bile duct	2		
Biliary cirrhosis	Liver regeneration	2		
Biliary cirrhosis	Disorder of biliary tract	3		
Biliary cirrhosis	Lesion of liver	3		
Biliary cirrhosis	Disease of liver	3	Liver disease	Reject



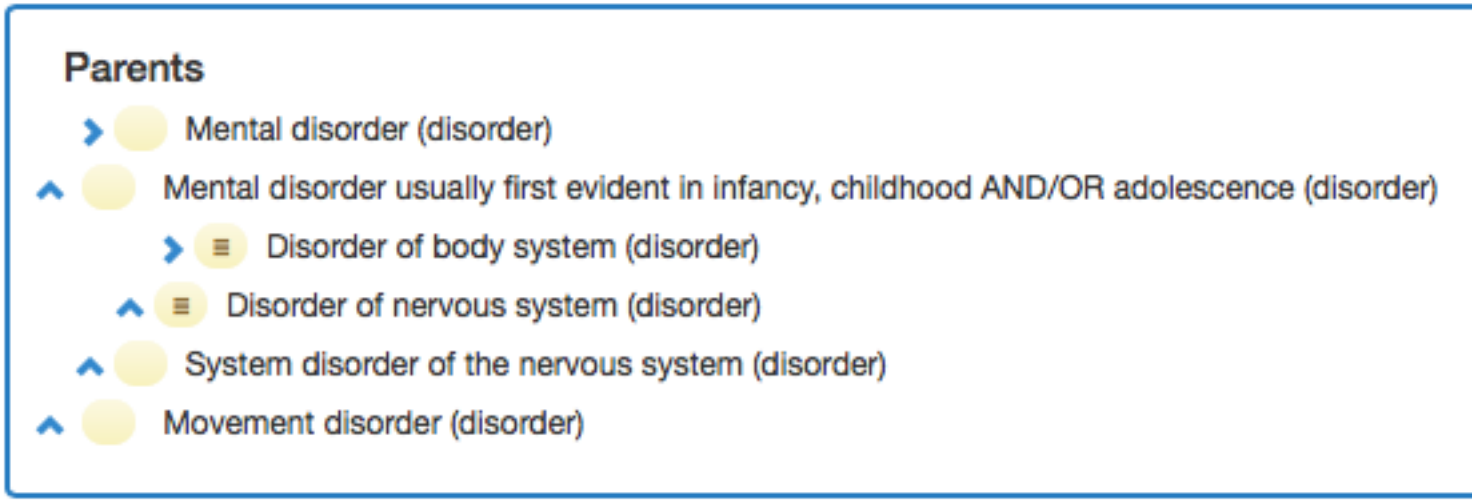
Closest parent match: Reject

Legacy EMR System			Target EMR System Match	
SNOMED Concept	SNOMED Parents	Distance	SNOMED Concept	SME Response
Autoimmune hepatitis	Autoimmune disease	2	Autoimmune disease	Accept
Autoimmune hepatitis	Inflammatory disease of liver	1	Inflammatory disease of liver	Reject
Autoimmune hepatitis	Autoimmune liver disease	1		
Autoimmune hepatitis	Inflammatory disorder of digestive system	2		
Autoimmune hepatitis	Inflamation of specific body organ	2		
Autoimmune hepatitis	Disease of liver	2	Disease of liver	Reject



None of the matched parent concepts accepted

Legacy EMR System			Target EMR System Match	
SNOMED Concept	SNOMED Parents	Distance	SNOMED Concept	SME Response
Tic disorder	Mental disorder usually first evident in infancy	1		
Tic disorder	Movement disorder	1	Movement disorder	Reject
Tic disorder	Mental disorder	2	Mental disorder	Reject
Tic disorder	System disorder of the nervous system	2		
Tic disorder	Disorder of nervous system	3		
Tic disorder	Disorder of body system	4		



- Another concepts (from a different SNOMED hierarchy) was selected from the target EMR data set *Tic* (Finding)

Results

- We have verified that 88% of the mappings based on the shortest distance were acceptable
- We didn't find any acceptable match for 4% of the concepts

Shortest Distance Match			
Accept		Reject	
Immediate parent match	> 1 parents match	Immediate parent match	> 1 parents match
891	893	124	34
1784		158	

Conclusions

- Identification of concepts with lexical similarities between clinical data sets is an important contributory factor in interoperability and re-usage of data

- Our approach has greatly facilitated the migration and re-usage of patient data during migrating from a legacy EMR system to a new EHR system

- This framework can be used for similar environments where concepts with higher granularities should be mapped to the relevant but less granular concepts based on SNOMED CT *is-a* relationship

Acknowledgement

MD Anderson Cancer Center

Department of Institutional Analytics & Informatics

Samuel Camp, Trey Kelly, Jerry Henderson, MD

Vanderbilt University Medical Center

Department of HealthIT

Shari Just, RN, Scott Nelson Pharm D

Asli Weitkamp, PhD

National Library of Medicine

MetaMap team

Francois Lang, Alan Aronson, PhD

Dina Demner-Fushman, MD, PhD

Questions?