Using SNOMED CT Concept Definitions for Natural Language Processing

April Russell, Hua Cheng, James Chan, Thomas Polzin, Gilan Saadawi

The Promise of SNOMED

- Comprehensive
- Widely Adopted
- Multinational and multilingual
- Extendable
- Description Logic
- Fully defined concepts

The Challenge
Mapping Narrative to Clinical Definitions

Concept: 75835007
Laparoscopic-assisted vaginal hysterectomy (procedure)

It was elected to perform a laparoscopic-assisted vaginal hysterectomy.

Strict String Matching

The laparoscope was inserted through the sheath. …
It was elected to remove the uterus.

?

The patient had a history of vaginal bleeding for which she had an open abdominal hysterectomy, she is now coming with acute cholecystitis and will be scheduled for a laparoscopic cholecystectomy.

Loose String Matching

The Challenge

- SNOMED concept descriptions are clinical definitions
- Concept definitions can be very complex
- Very loose correlation between a concept description and its actual reference within medical documentation

Concept 336651000
Extraction of primary membranous cataract by mechanical fragmentation

Concept 373161008
Metastasis in 1 to 3 axillary lymph nodes and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent (breast)

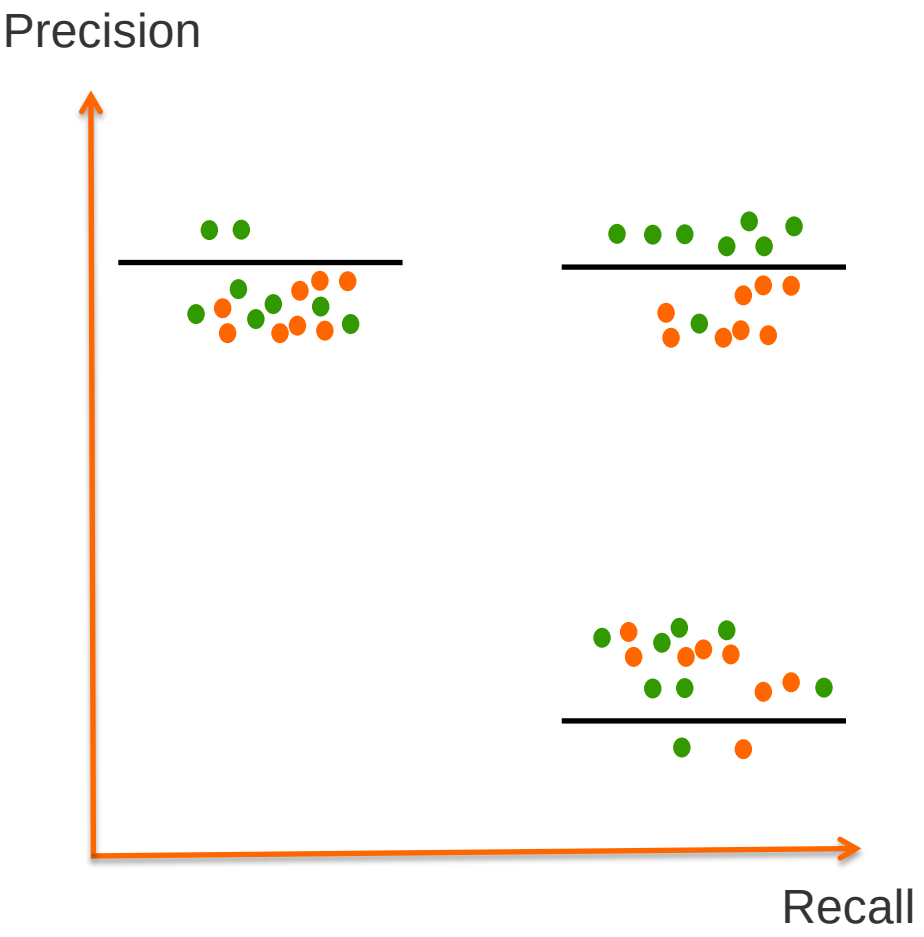
How do we map clinical narrative accurately to SNOMED CT concepts?

The Accuracy Challenge: High Precision and High Recall

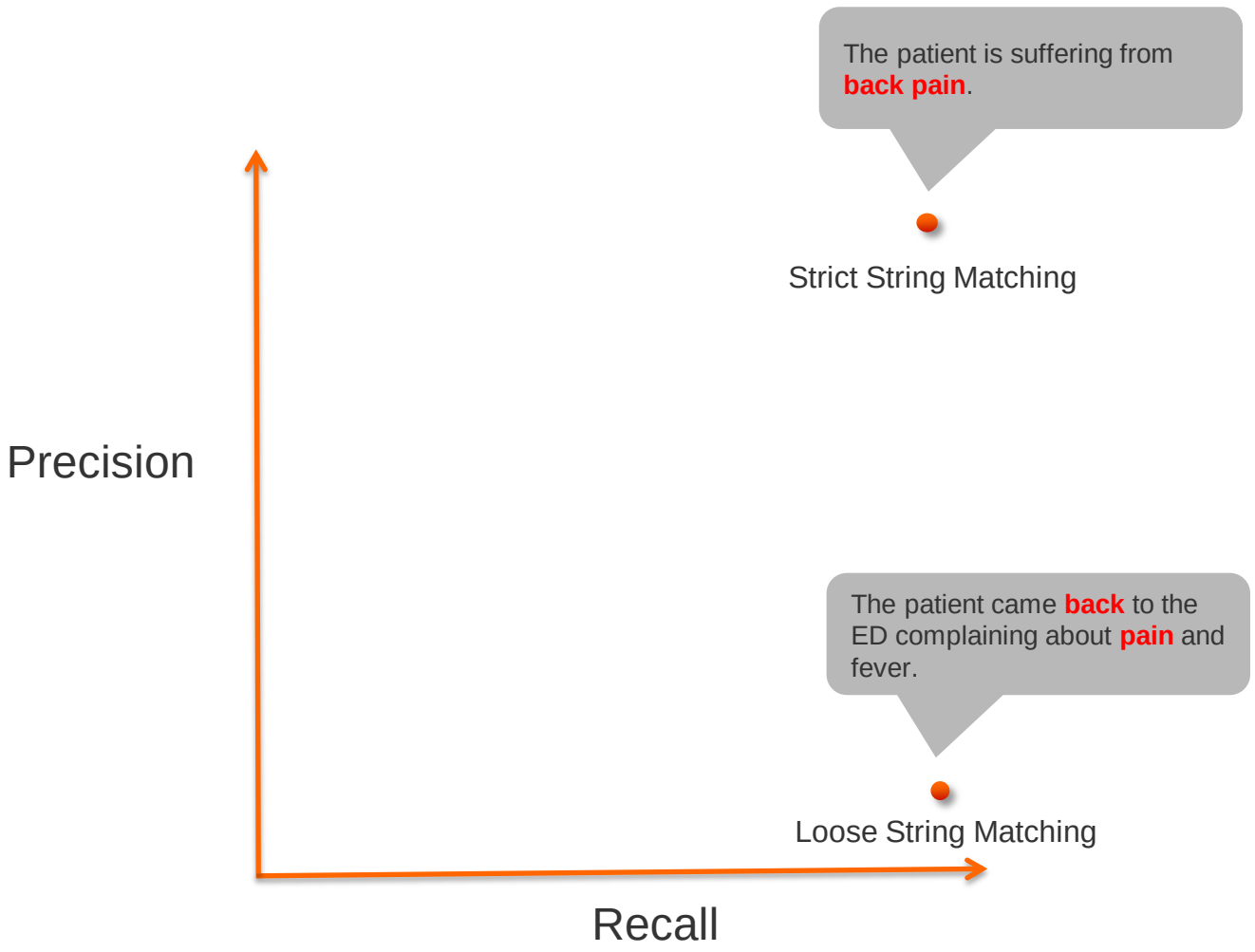
High Precision: Mapped concepts are correct

High Recall: Most relevant concepts are mapped

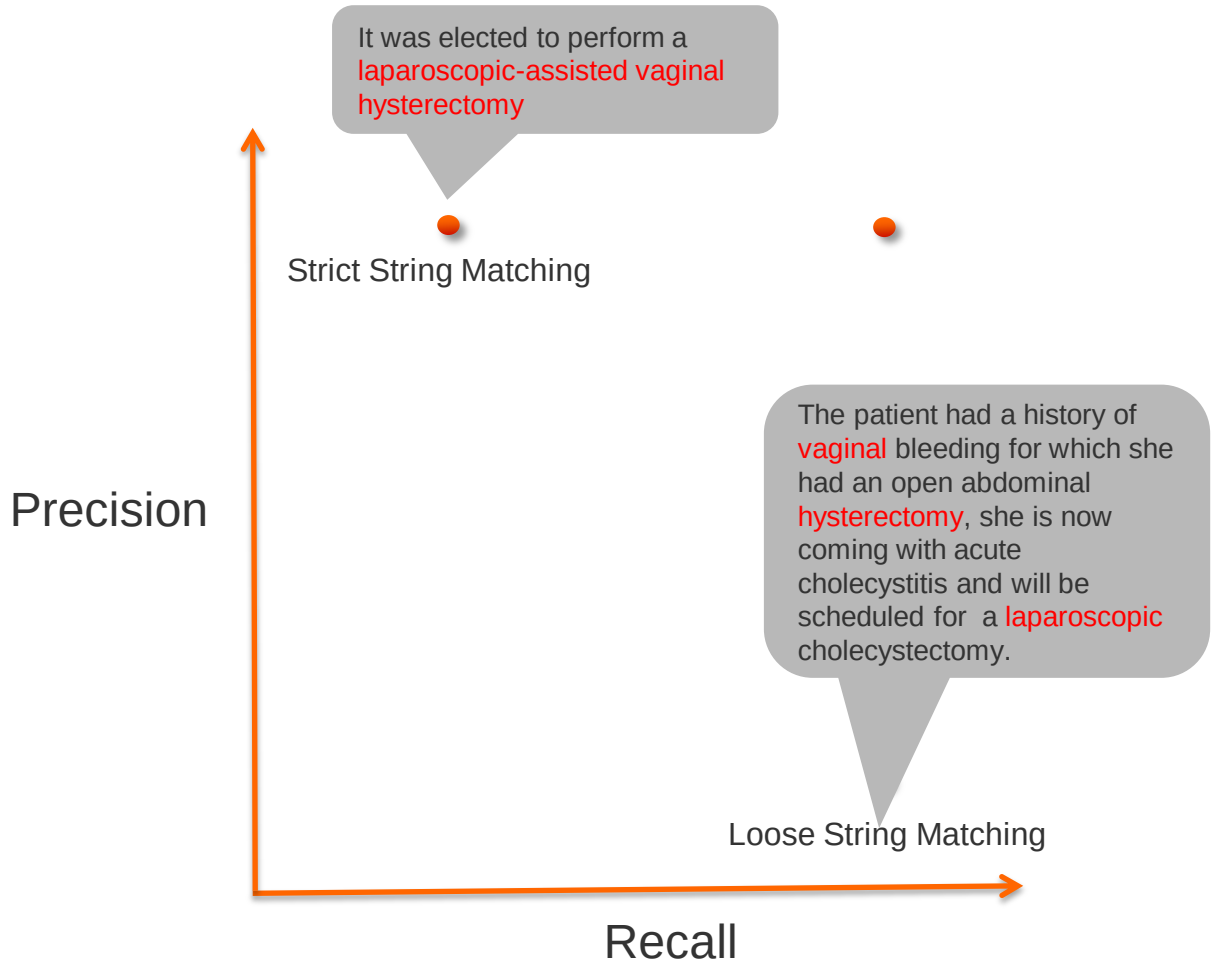
High Precision and High Recall:
- Mapped concepts are correct
- Most relevant concepts are mapped



The Challenge: Simple Concepts



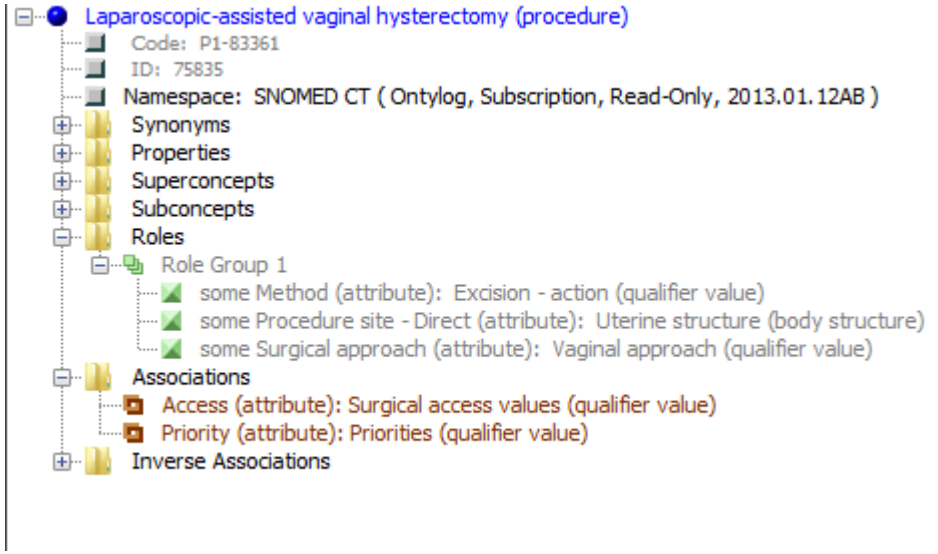
The Challenge: Complex Concepts



Two Observations

- Items that are related semantically tend to occur in close proximity.
- A significant number of SNOMED concepts are defined: Necessary (and sometimes sufficient) conditions are part of the definition

He broke his left arm and his right leg.



The Approach

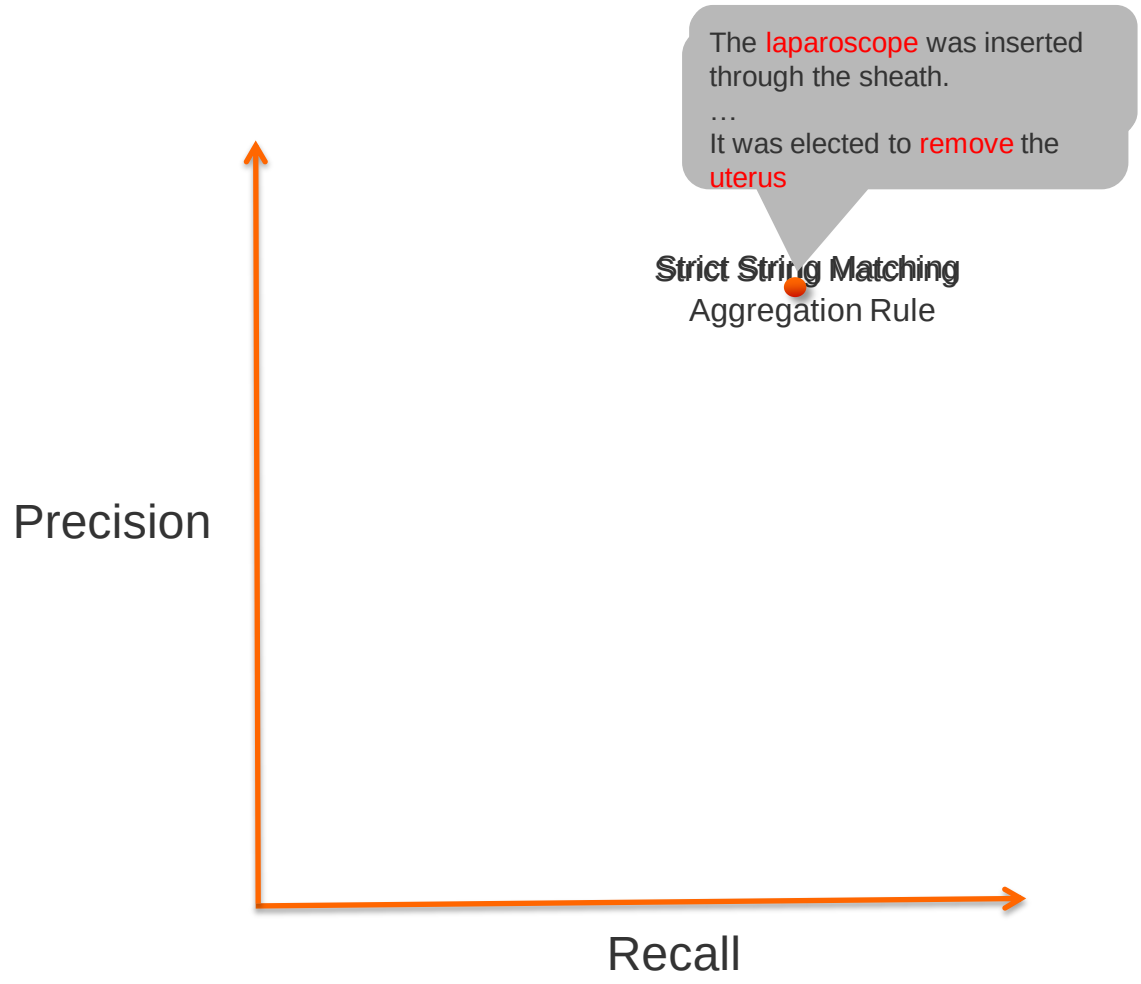
- Not trying to match a concept description but matching necessary parts of the concept definition
- Strict String Matching for concepts that are part of a definition
  - Finding site
  - Morphological abnormality
- Aggregation rules to assemble necessary parts given
  - Syntactical conditions
    - The attribute *device used* is realized as a prepositional phrase
  - Context conditions
    - Same paragraph
    - Same section
    - Particular section type

The Challenge: Complex Concepts

The laparoscope was inserted through the sheath.

…

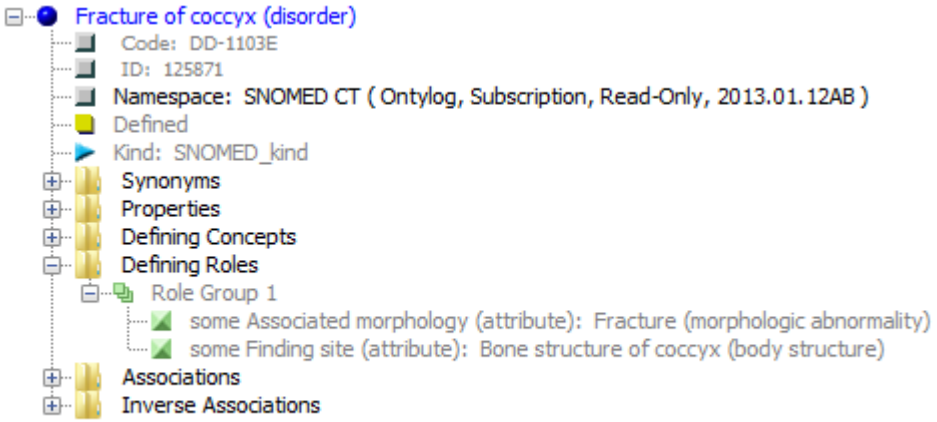
It was elected to remove the uterus



Method 1: Automatic Derivation of Aggregation Rules

- Fully Defined Concepts in certain subsets within SNOMED
- Merge role groups for conjunct concepts
- Use Strict String Matching for matching relevant parts of the definition:
  - Associated morphology
  - Finding site
- Use Aggregation Rules to assemble concept from its part

Method 1: Example

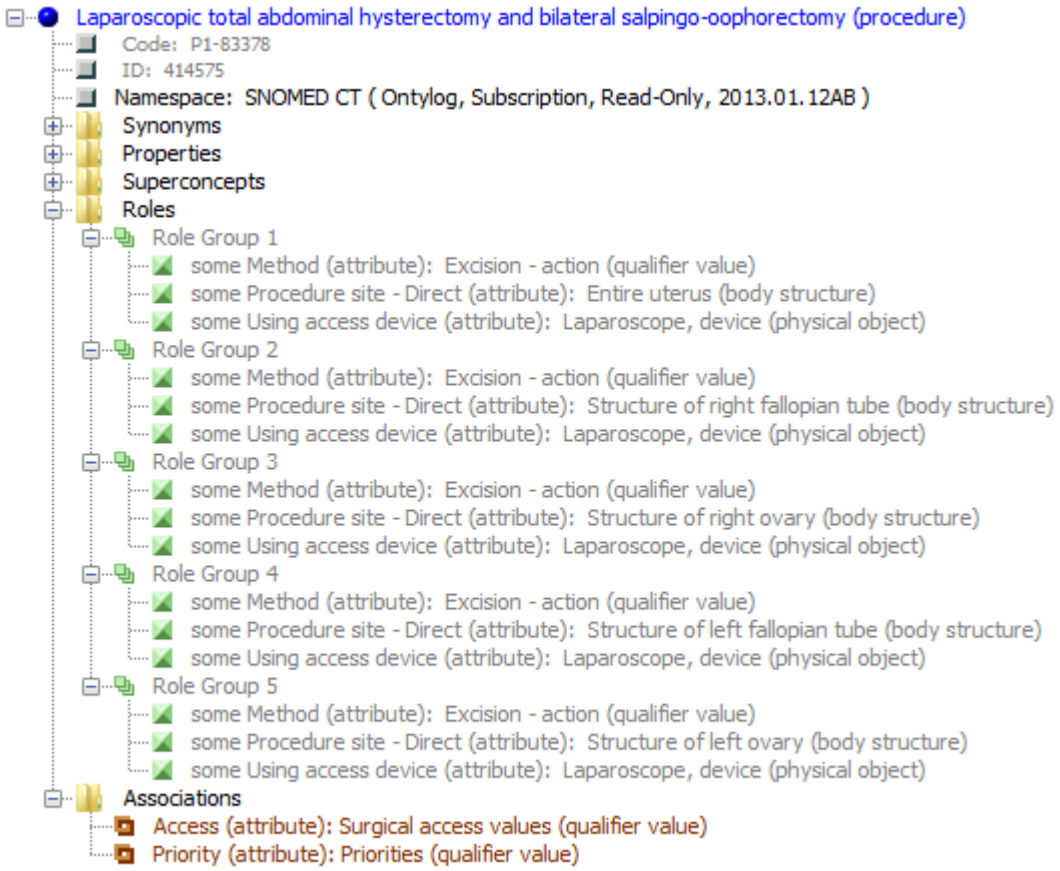


Role	Value	Scope
Finding Site	Bone structure of coccyx (body structure)	Sentence
Associated Morphology	Fracture (morphological abnormality)	Target

Method 2: Manual Derivation of Aggregation Rules

- Significant number of SNOMED concepts lack appropriate necessary and sufficient conditions
  - No aggregation rules
  - Add relationships by hand
- Significant number of SNOMED concepts are too detailed with respect to necessary and sufficient conditions
  - Aggregation rules require information that the clinical narrator typically does not document
  - Remove relationships

Method 2: Example



Method 2: Example continued

Laparoscopic total abdominal hysterectomy and bilateral salpingo-oophorectomy

Role	Value	Scope
Surgical approach	Laparoscopy	section

isa isa

Role	Value	Scope
Procedure site	Uterine structure	sentence

Hysterectomy

Role	Value	Scope
Procedure Site	Abdominal structure	section
Using device	Trocar device	section

Laparascopy

Some Numbers

- Automatically derived aggregation rules:
  - Body structure concepts with laterality information: 17,439
  - Dislocation of Joint, Fracture of Bone, Neoplasm of Colon, and Neoplasm of Breast: 2307
- Manually built post-coordination rules: 3082
- Significant improvement in CAC
- Used in CDI
- Used in Debridement

Summary

- Promising approach that improves overall accuracy
- Mapping narrative to SNOMED is not all
  - Certainty
  - Subject
  - Temporality
  - Identity
- Learning Aggregation Rules
- Languages other than English