

July 24, 2003

# Exploring Methods for Improving the Integration of LOINC in the UMLS



Holli Huff, NLM Fellow  
Olivier Bodenreider MD PhD, Preceptor



# Metathesaurus Concepts

- Two structure types
  - Natural language
  - Fielded data – LOINC
- These structure types are used to encode similar or identical information in different ways.
- How can we integrate the two structure types?

# Introduction to LOINC

## Logical Observation Identifier Names and Codes

- Medical coding system used for the identification of laboratory and clinical observations
- Regenstrief Institute
- Financial support from many agencies
- Publicly available\*

# LOINC Laboratory Observations

Component

Timing

Scale

SODIUM : SCNC : PT : UR : QN

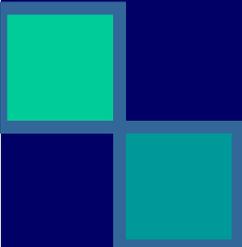
Property

System

Method

not mandatory

# Natural Language Examples

Two overlapping squares, one light blue and one teal, are positioned to the left of the text.

serum glucose

urinary sodium increased

blood protein electrophoresis

CSF cell count and differential

drug of abuse screen on urine

Two overlapping squares, one light blue and one teal, are positioned to the right of the text.

# Comparison: LOINC to Natural Language

[ Serum Creatinine Tests  
CREATININE:SCNC:PT:SER/PLA:QN

# Comparison: LOINC to Natural Language

[ Serum Creatinine Tests  
CREATININE:SCNC:PT:SER/PLA:QN

# Comparison: LOINC to Natural Language

[ Serum Creatinine Tests

[ CREATININE:SCNC:PT:SER/PLA:QN

[ Urine Cannabinoids Screen

[ CANNABINOIDS:ACNC:PT:UR:ORD:SCREEN

# Comparison: LOINC to Natural Language

[ Serum Creatinine Tests

CREATININE:SCNC:PT:SER/PLA:QN

[ Urine Cannabinoids Screen

CANNABINOIDS:ACNC:PT:UR:ORD: SCREEN

They have similar constituents, but in different forms.

# Objective

- Create a method to facilitate the detection of missing relationships (including synonymy) between natural language and LOINC terms in the UMLS based on the order of constituents

# Methods

- Target natural language concepts that potentially match LOINC concepts
  - Using semantic types
  - Using hierarchies
- Use two strategies for comparing natural language and LOINC concepts
  - Generative approach
  - Analytical approach

# Target Concepts Using Semantic Type

- Clinical Attribute
  - Laboratory Procedure
  - Laboratory or Test Result
- 
- 10,788 concepts / 20,903 strings

# Target Concepts Using Hierarchy

- Descendants of
  - C0427351 Laboratory Test Observations
  - C0600201 Laboratory Techniques and Procedures
- 54,792 concepts / 131,744 strings

# Comparison Strategy

- Generative Approach
  - Generate natural language terms from LOINC decompositions
- Analytical Approach
  - Analyze existing natural language terms with reference to LOINC decompositions

# Generative Approach

- 4 Parts
  - Grammar
  - Lexicon
  - String Generation
  - Matching
- Purpose: Identify patterns in natural language strings

# 1. Grammar

Build a grammar by inspection of natural language terms

<C>   <S>  
sodium urine

<S>   <C>   <Q>  
urine glucose decreased

## 2. Lexicon

Create a lexicon from the values of LOINC axes, augmented by missing concepts

<C>  
component

glucose

date last dose

sodium

*Na*

<S>  
system

CSF

saliva

urine

<Q>  
qualifier

abnormal

decreased

high

### 3. String Generation

Generate strings using the lexicon and grammar.

<C> <S>

glucose CSF

glucose saliva

glucose urine

date last dose CSF

date last dose saliva

date last dose urine

sodium CSF

sodium saliva

sodium urine

Na CSF

Na saliva

Na urine

## 4. Matching

Compare the generated strings to the natural language target terms.

glucose CSF

glucose saliva

glucose urine

date last dose CSF

date last dose saliva

date last dose urine

sodium CSF

sodium saliva

sodium urine

Na CSF

Na saliva

Na urine

## 4. Matching

Compare the generated strings to the natural language target terms.

glucose CSF

glucose saliva

glucose urine

date last dose CSF

date last dose saliva

date last dose urine

sodium CSF

sodium saliva

sodium urine

Na CSF

Na saliva

Na urine

# Generative Approach: Results

- 19 rules generated over 2.4 billion strings consuming 111 gigabytes of disk space and running for approximately 50 hours.
- 5252 strings (1637 concepts) were matched.
- What does this mean?  
Hold that thought . . .

# Analytical Approach

- 4 Parts
  - Natural Language Terms
  - Symbol Replacement
  - Recognize Patterns
  - Augment the Lexicon
- Purpose: Identify patterns in natural language strings

# 1. Natural Language Terms

|          |          |                            |
|----------|----------|----------------------------|
| C0004076 | S0006391 | urine glucose              |
| C0004076 | S1640009 | Glucose urine              |
| C0017746 | S0001306 | Glucose, decreased level   |
| C0202040 | S0320575 | Glucose measurement, CSF   |
| C0337438 | S0045142 | Glucose                    |
| C0430376 | S1056061 | Urine dipstick for glucose |

## 2. Symbol Replacement

<C>



|          |          |                                   |
|----------|----------|-----------------------------------|
| C0004076 | S0006391 | urine <u>glucose</u>              |
| C0004076 | S1640009 | <u>Glucose</u> urine              |
| C0017746 | S0001306 | <u>Glucose</u> , decreased level  |
| C0202040 | S0320575 | <u>Glucose</u> measurement, CSF   |
| C0337438 | S0045142 | <u>Glucose</u>                    |
| C0430376 | S1056061 | Urine dipstick for <u>glucose</u> |

## 2. Symbol Replacement

<S>  
↓

|          |          |       |                   |
|----------|----------|-------|-------------------|
| C0004076 | S0006391 | urine | <C>               |
| C0004076 | S1640009 | <C>   | urine             |
| C0017746 | S0001306 | <C>   | , decreased level |
| C0202040 | S0320575 | <C>   | measurement, CSF  |
| C0337438 | S0045142 | <C>   |                   |
| C0430376 | S1056061 | Urine | dipstick for <C>  |

## 2. Symbol Replacement

|          |          |     |                   |
|----------|----------|-----|-------------------|
| C0004076 | S0006391 | <S> | <C>               |
| C0004076 | S1640009 | <C> | <S>               |
| C0017746 | S0001306 | <C> | , decreased level |
| C0202040 | S0320575 | <C> | measurement, <S>  |
| C0337438 | S0045142 | <C> |                   |
| C0430376 | S1056061 | <S> | dipstick for <C>  |

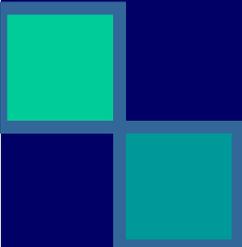
## 2. Symbol Replacement

|          |          |     |                   |     |
|----------|----------|-----|-------------------|-----|
| C0004076 | S0006391 | <S> | <C>               | <Q> |
| C0004076 | S1640009 | <C> | <S>               | ↙   |
| C0017746 | S0001306 | <C> | , decreased level |     |
| C0202040 | S0320575 | <C> | measurement, <S>  |     |
| C0337438 | S0045142 | <C> |                   |     |
| C0430376 | S1056061 | <S> | dipstick for <C>  |     |

## 2. Symbol Replacement

|          |          |     |              |       |
|----------|----------|-----|--------------|-------|
| C0004076 | S0006391 | <S> | <C>          |       |
| C0004076 | S1640009 | <C> | <S>          |       |
| C0017746 | S0001306 | <C> | , <Q>        | level |
| C0202040 | S0320575 | <C> | measurement, | <S>   |
| C0337438 | S0045142 | <C> |              |       |
| C0430376 | S1056061 | <S> | dipstick for | <C>   |

### 3. Recognize Patterns



<S> <C>

<C> <S>

<C>, <Q> level

<C> measurement, <S>

<C>

<S> dipstick for <C>

### 3. Recognize Patterns

<S> <C>

<C> <S>

<C>, <Q> level

<C> measurement, <S>

<C>

<S> dipstick for <C>

## 4. Augment the Lexicon

<C>, <Q> level

<C> measurement, <S>

<S> dipstick for <C>

---

<C>

<C> <S>

<S> <C>

# Most Frequent Patterns

<C>

<C> <T>

<S> <C> <T>

<S> <C>

<S> <C> <Q>

<C> <M>

<C> <T>, <S>

<M>

<S> <M>

<S> <T>

<T>, <C>

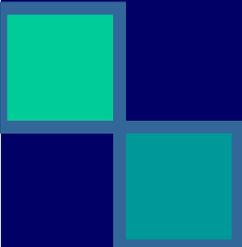
<T>; <C>

<S> <C> <M>

<C> <Q>

# Analytical Approach: Results

- Generated 458 patterns, ran faster, and used much less disk space.
- Matched 11,134 strings representing 4543 concepts.
- What does this mean?

A decorative graphic consisting of two overlapping squares, one light blue and one teal, is positioned to the left of the main title.

# Detecting Relationships Between Natural Language and LOINC Terms

A decorative graphic consisting of two overlapping squares, one light blue and one teal, is positioned to the right of the main title.

Preliminary Work



# Relationships between Natural Language and LOINC Terms

C0365737 URINE CODEINE SCREEN

<S>

<C>

<M>



# Relationships between Natural Language and LOINC Terms

C0365737 URINE CODEINE SCREEN

<S>

<C>

<M>

<C> : \* : \* : <S> : \* : <M>

# Relationships between Natural Language and LOINC Terms

C0365737 URINE CODEINE SCREEN

<S>

<C>

<M>

<C> : \* : \* : <S> : \* : <M>

# Relationships between Natural Language and LOINC Terms

C0365737 URINE CODEINE SCREEN

<S>            <C>            <M>

<C>            : \*            : \*            : <S> : \*            : <M>

CODEINE            : ACNC : PT : UR : ORD : SCREEN

CODEINE CUTOFF : MCNC : PT : UR : QN : SCREEN

# Matches



C0587865 URINE OPIATES SCREEN

|                          |      |    |    |     |                      |
|--------------------------|------|----|----|-----|----------------------|
| OPIATES                  | ACNC | PT | UR | ORD | SCREEN               |
| OPIATES                  | ACNC | PT | UR | ORD | SCREEN>2000<br>NG/ML |
| OPIATES                  | ACNC | PT | UR | ORD | SAMHSA SCREEN        |
| OPIATES<br>CUTOFF        | MCNC | PT | UR | QN  | SCREEN               |
| OPIATES<br>TESTED<br>FOR | PRID | PT | UR | NAR | SCREEN               |
| OPIATES<br>TESTED<br>FOR | PRID | PT | UR | NOM | SCREEN               |

# Future Work

- Automate the process of determining relationships.
- Correct the inconsistencies in LOINC and the UMLS that are exposed by this process.
- Use MetaMap as an alternative for the analytical approach.

# Summary

- Explored 2 approaches for identifying relations between LOINC and existing natural language terms in the Metathesaurus.
- The analytical approach produced better results.
- The detection of relationships remains to be done.

# Acknowledgements

Olivier Bodenreider

Tom Rindflesch

May Cheh

Bill Hole's Group

Alexa McCray

The Lister Hill Center Family

The Summer Rotation Students