ihtsdo · Leading healthcare terminology, worldwide

# Overview of Expression Normalization, Equivalence and Subsumption Testing

Delivering
**SNOMED CT**
The global language of healthcare

David Markwell
Head of Education

---

ihtsdo · Delivering SNOMED CT

## Overview

- Different ways to express the same meaning
- Different logical views of expressions
- Normalizing expressions
- Normal forms and canonical forms
- Testing for equivalence and subsumption
- Taking account of context
- Limitations to subsumption testing
- Optimizing expression subsumption tests
- Summary
- Links to Further Information

**Different Ways to Express the Same Meaning**

---

## Different Ways to Represent a Closed Fracture of the Shaft of the Tibia

28012007 |Closed fracture of shaft of tibia|

125605004 |Fracture of bone| :
{ 363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|,
116676008 |Associated morphology| = 20946005 |Fracture, closed | }
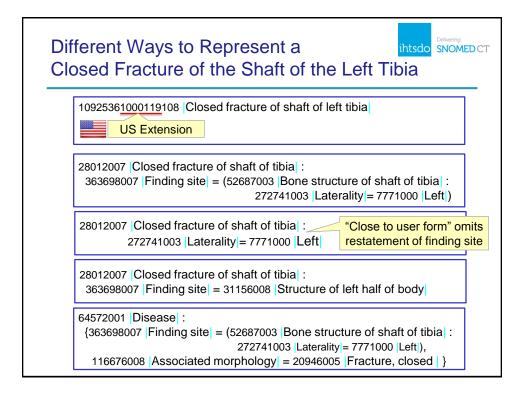
64572001 |Disease| :
{ 363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|,
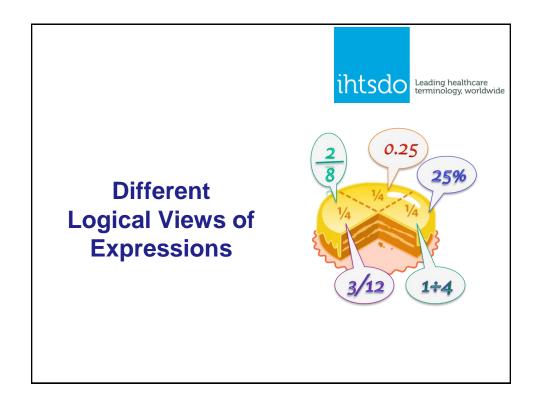116676008 |Associated morphology| = 20946005 |Fracture, closed | }
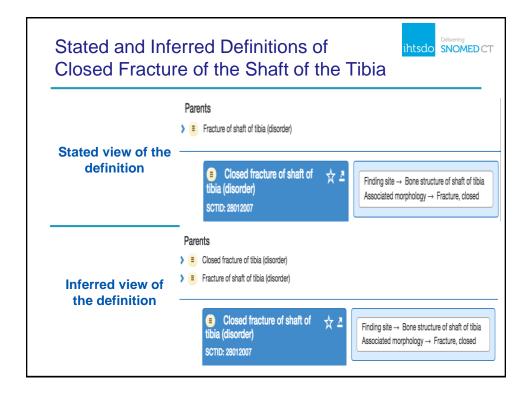
423125000 |Closed fracture of bone|:
363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|

6990005 |Fracture of shaft of tibia |:
116676008 |Associated morphology| = 20946005 |Fracture, closed |

## Different Ways to Represent a Closed Fracture of the Shaft of the Left Tibia

ihtsdo SNOMED CT
Delivering

10925361000119108 |Closed fracture of shaft of left tibia|

US Extension

28012007 |Closed fracture of shaft of tibia| :
363698007 |Finding site| = (52687003 |Bone structure of shaft of tibia| :
272741003 |Laterality|= 7771000 |Left|)

28012007 |Closed fracture of shaft of tibia| :
272741003 |Laterality|= 7771000 |Left|

"Close to user form" omits restatement of finding site

28012007 |Closed fracture of shaft of tibia| :
363698007 |Finding site| = 31156008 |Structure of left half of body|

64572001 |Disease| :
{363698007 |Finding site| = (52687003 |Bone structure of shaft of tibia| :
272741003 |Laterality|= 7771000 |Left|),
116676008 |Associated morphology| = 20946005 |Fracture, closed | }

---

ihtsdo
Leading healthcare terminology, worldwide

$\frac{2}{8}$    0.25    25%

¼
¼   ¼

**Different Logical Views of Expressions**

3/12    1÷4

3

## Stated and Inferred Definitions of Closed Fracture of the Shaft of the Tibia

ihtsdo SNOMED CT  Delivering

**Parents**
> ≣  Fracture of shaft of tibia (disorder)

**Stated view of the definition**

≣  **Closed fracture of shaft of tibia (disorder)**  ☆ ↗
SCTID: 28012007

Finding site → Bone structure of shaft of tibia
Associated morphology → Fracture, closed

**Parents**
> ≣  Closed fracture of tibia (disorder)
> ≣  Fracture of shaft of tibia (disorder)

**Inferred view of the definition**

≣  **Closed fracture of shaft of tibia (disorder)**  ☆ ↗
SCTID: 28012007

Finding site → Bone structure of shaft of tibia
Associated morphology → Fracture, closed

---

## Expression Views of Definitions of Closed Fracture of the Shaft of the Tibia

ihtsdo SNOMED CT  Delivering

Concept Details

**Concept Details**                                                                                  ⏱ ⚙

Summary   Details   Diagram   Expression   Refsets   Members   References

Pre-coordinated Expression (*)  ⬓

    28012007 |Closed fracture of shaft of tibia (disorder)|

Expression from Stated Concept Definition (*)  ⬓

    === 6990005 |Fracture of shaft of tibia (disorder)| :
            { 363698007 |Finding site (attribute)| = 52687003 |Bone structure of shaft of tibia (body structure)|,
              116676008 |Associated morphology (attribute)| = 20946005 |Fracture, closed (morphologic abnormality)| }

Expression from Inferred Concept Definition (*)  ⬓

    === 6990005 |Fracture of shaft of tibia (disorder)| +
        447139008 |Closed fracture of tibia (disorder)| :
            { 363698007 |Finding site (attribute)| = 52687003 |Bone structure of shaft of tibia (body structure)|,
              116676008 |Associated morphology (attribute)| = 20946005 |Fracture, closed (morphologic abnormality)| }

## Stated and Inferred Expressions for Closed Fracture of the Shaft of the Tibia

ihtsdo  Delivering SNOMED CT

| Precoordinated expression | 28012007 |Closed fracture of shaft of tibia| |
|---|---|
| Stated view of the definition as and expression | 6990005 |Fracture of shaft of tibia | : { 363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|, 116676008 |Associated morphology| = 20946005 |Fracture, closed| } |
| Inferred view of the definition as an expression | 6990005 |Fracture of shaft of tibia| + 447139008 |Closed fracture of tibia | : { 363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|, 116676008 |Associated morphology| = 20946005 |Fracture, closed| } |

ihtsdo  Leading healthcare terminology, worldwide

# Normalizing Expressions

*Rainbow colors*

RED  GREEN  BLUE
INDIGO  VIOLET
ORANGE  YELLOW

ihtsdo  Delivering SNOMED CT

## Comparing Fractions and Comparing Ideas



- Simplifying fractions helps us to compare them
  - 3/12 → ¼
  - 2/8 → ¼
  - 1 ÷ 4 → ¼
  - 0.25 → 0.25/1 → ¼
  - 25% → 25/100 → ¼

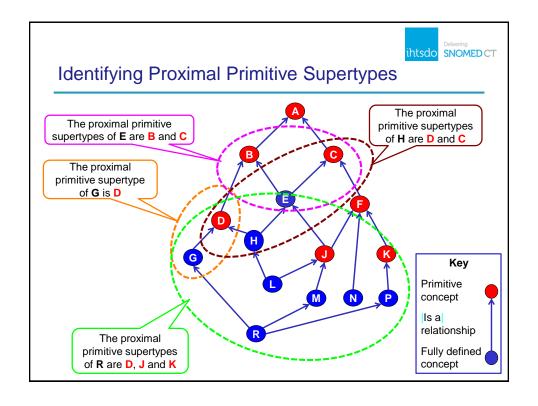**Rainbow colors**

↓

RED  GREEN  BLUE
INDIGO  VIOLET
ORANGE  YELLOW
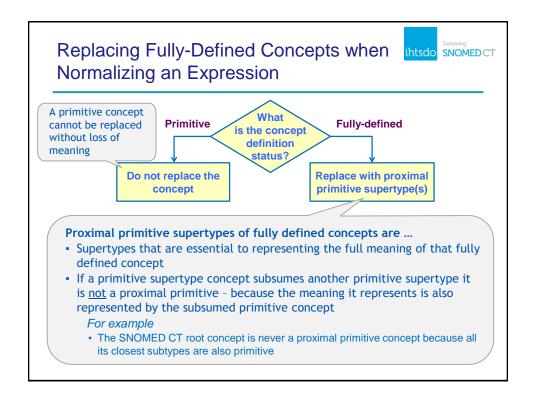
- Breaking down ideas helps us to answer questions like
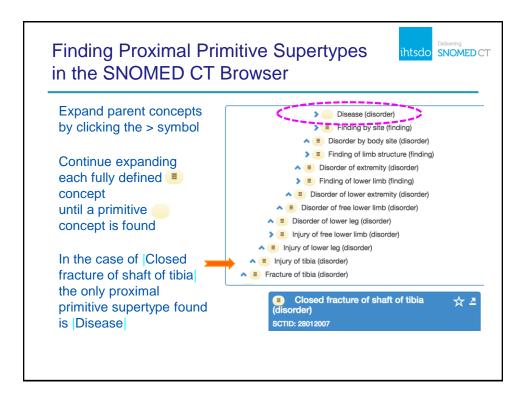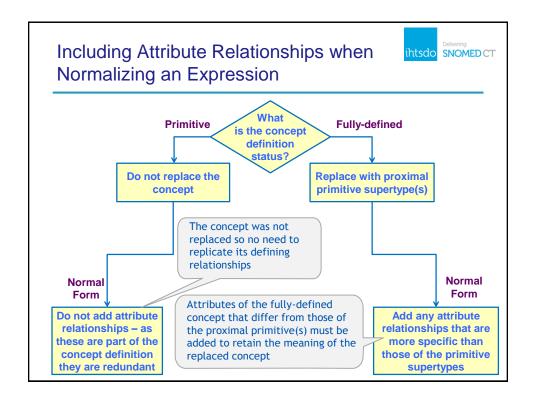  - Is red a rainbow color?
  - Does a rainbow contain blue?

---

ihtsdo  Delivering SNOMED CT

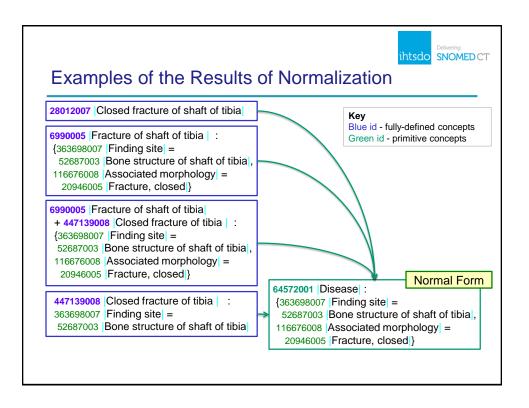## Normalization Breaks an Expression into Parts

- The process of normalization
  - Replaces each fully-defined concept with an expression representing the definition of that concept
    - The definition of a fully-defined concept has the same meaning as the concept itself
  - Parts of the resulting expression that repeat the same information are merged to reduce or remove redundancy
- The result of normalization
  - All identifiers in the expression refer to primitive concepts
    - Primitive concepts cannot be replaced, as primitive definitions do not fully represent the meaning of the concept
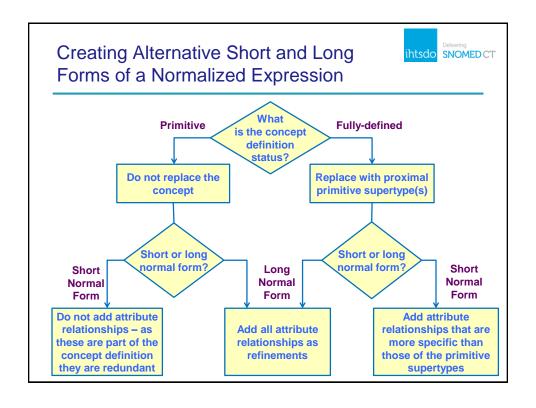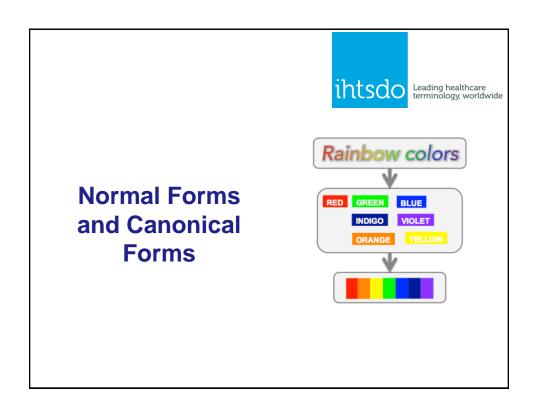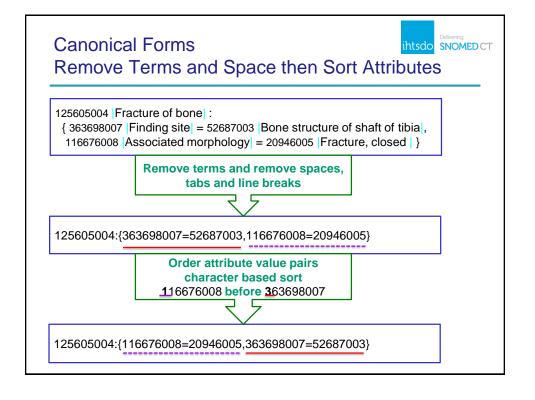
## Replacing Fully-Defined Concepts when Normalizing an Expression

ihtsdo SNOMED CT

A primitive concept cannot be replaced without loss of meaning

**What is the concept definition status?**

**Primitive**

**Fully-defined**

**Do not replace the concept**

**Replace with proximal primitive supertype(s)**

**Proximal primitive supertypes of fully defined concepts are …**
- Supertypes that are essential to representing the full meaning of that fully defined concept
- If a primitive supertype concept subsumes another primitive supertype it is not a proximal primitive – because the meaning it represents is also represented by the subsumed primitive concept
  - *For example*
    - The SNOMED CT root concept is never a proximal primitive concept because all its closest subtypes are also primitive

---

## Identifying Proximal Primitive Supertypes

ihtsdo SNOMED CT

The proximal primitive supertypes of **E** are **B** and **C**

The proximal primitive supertypes of **H** are **D** and **C**

The proximal primitive supertype of **G** is **D**

The proximal primitive supertypes of **R** are **D**, **J** and **K**

**Key**

Primitive concept

|Is a| relationship

Fully defined concept

## Finding Proximal Primitive Supertypes in the SNOMED CT Browser

ihtsdo | Delivering SNOMED CT

Expand parent concepts by clicking the > symbol

Continue expanding each fully defined ≡ concept until a primitive concept is found

In the case of |Closed fracture of shaft of tibia| the only proximal primitive supertype found is |Disease|

> Disease (disorder)
> ≡ Finding by site (finding)
∧ ≡ Disorder by body site (disorder)
> ≡ Finding of limb structure (finding)
∧ ≡ Disorder of extremity (disorder)
> ≡ Finding of lower limb (finding)
∧ ≡ Disorder of lower extremity (disorder)
∧ ≡ Disorder of free lower limb (disorder)
∧ ≡ Disorder of lower leg (disorder)
> ≡ Injury of free lower limb (disorder)
∧ ≡ Injury of lower leg (disorder)
∧ ≡ Injury of tibia (disorder)
∧ ≡ Fracture of tibia (disorder)

≡ **Closed fracture of shaft of tibia (disorder)**   ☆ ↗
SCTID: 28012007

---

## Including Attribute Relationships when Normalizing an Expression

ihtsdo | Delivering SNOMED CT

**What is the concept definition status?**

Primitive → **Do not replace the concept**

Fully-defined → **Replace with proximal primitive supertype(s)**

The concept was not replaced so no need to replicate its defining relationships

**Normal Form**

**Do not add attribute relationships – as these are part of the concept definition they are redundant**

Attributes of the fully-defined concept that differ from those of the proximal primitive(s) must be added to retain the meaning of the replaced concept

**Normal Form**

**Add any attribute relationships that are more specific than those of the primitive supertypes**

## Examples of the Results of Normalization

ihtsdo Delivering SNOMED CT

28012007 |Closed fracture of shaft of tibia|

6990005 |Fracture of shaft of tibia| :
{363698007 |Finding site| =
  52687003 |Bone structure of shaft of tibia|,
116676008 |Associated morphology| =
  20946005 |Fracture, closed|}

6990005 |Fracture of shaft of tibia|
 + 447139008 |Closed fracture of tibia| :
{363698007 |Finding site| =
  52687003 |Bone structure of shaft of tibia|,
116676008 |Associated morphology| =
  20946005 |Fracture, closed|}

447139008 |Closed fracture of tibia| :
363698007 |Finding site| =
  52687003 |Bone structure of shaft of tibia|

**Key**
Blue id - fully-defined concepts
Green id - primitive concepts

Normal Form

64572001 |Disease| :
{363698007 |Finding site| =
  52687003 |Bone structure of shaft of tibia|,
116676008 |Associated morphology| =
  20946005 |Fracture, closed|}

---

## Creating Alternative Short and Long Forms of a Normalized Expression

ihtsdo Delivering SNOMED CT

**What is the concept definition status?**

Primitive → Do not replace the concept

Fully-defined → Replace with proximal primitive supertype(s)

Short or long normal form?

Short Normal Form → Do not add attribute relationships – as these are part of the concept definition they are redundant

Long Normal Form → Add all attribute relationships as refinements

Short or long normal form?

Short Normal Form → Add attribute relationships that are more specific than those of the primitive supertypes

## Short and Long Normal Forms

- Short normal form
  - No redundancy or duplication of meaning
  - Includes only attributes that cannot be inferred from primitive supertype concepts in the expression
- Long normal form
  - Includes all attributes – even those that can be inferred from primitive supertypes in the expression
- Subsumption testing is most efficient if …
  - The test conditions are represented in short normal form
    - More efficient testing as there are fewer conditions to test
  - The expressions being tested for subsumption are represented in long normal form
    - All attributes and values are available for testing

# Normal Forms and Canonical Forms

Rainbow colors

RED  GREEN  BLUE
INDIGO  VIOLET
ORANGE  YELLOW

## Alternative Representations of the Same Normal Form Expression

125605004 |Fracture of bone| :
  { 363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|,
   116676008 |Associated morphology| = 20946005 |Fracture, closed | }

125605004 |Fracture of bone| :
  {116676008 |Associated morphology| = 20946005 |Fracture, closed |,
   363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|}

125605004 |Fracture of bone (disorder)| :
  { 363698007 |Finding site (attribute)| = 52687003 |Bone structure of shaft of tibia (body structure)|,
   116676008 |Associated morphology (attribute)| = 20946005 |Fracture, closed (morphological abnormality) | }

| | |
|---|---|
| 125605004:<br>{116676008 = 20946005,<br>363698007 = 52687003} | 125605004\|broken bone\|:{116676008=20946005,363698007= 52687003} |

## Canonical Forms
## Remove Terms and Space then Sort Attributes

125605004 |Fracture of bone| :
  { 363698007 |Finding site| = 52687003 |Bone structure of shaft of tibia|,
   116676008 |Associated morphology| = 20946005 |Fracture, closed | }

**Remove terms and remove spaces, tabs and line breaks**

125605004:{363698007=52687003,116676008=20946005}

**Order attribute value pairs character based sort**
**1**16676008 **before** **3**63698007

125605004:{116676008=20946005,363698007=52687003}

# Testing for Equivalence and Subsumption

---

## Testing if Two Expressions are Equivalent

- Convert both expressions to the normal form
  - This can be either the long normal form or the short normal form
  - The same normal form must be used for both expressions
- Convert both normal form expressions to the canonical representation
- Compare the two canonical expressions
  - If they are identical the original expressions have the same logical meaning

## Testing if Expression-A Subsumes Expression-B

- Transform both expressions to normal forms
    - To make the comparison simpler
        - Transform *expression-A* to short normal form
        - Transform *expression-B* to long normal form
- Compare the primitive focus concepts
    - Do the primitive focus concepts of *expression-A* subsume all the primitive focus concepts of *expression-B*?
        - If no – *expression-A* does not subsume *expression-B*
        - If yes – continue by comparing the defining attributes
- Compare the defining attributes
    - Are all the defining attributes of *expression-A* present in the *expression-B*?
    - If so are the values of those attributes in the *expression-B* subtypes of the values of those attributes in *expression-A*
- If so then *expression-B* is subsumed by *expression-A*

*Note:* The above outline is a simplification. In practice, attribute hierarchies, attribute groups and nested expressions also affect these tests

---

# Taking Account of Context

ihtsdo SNOMED CT
Delivering

## Equivalence and Subsumption in Context

- Clinical ideas represented by expressions may be modified by context
  - For example: Family history, past history, planned procedure
- Context may be represented by
  - User-interface designs in which information is captured
  - Information model structures that represent particular contexts
  - Additional data in related record or message elements
  - Applying the SNOMED CT |Situation with explicit context| model
- Subsumption testing must take account of context
  - Representing context using the |Situation with explicit context| model enables subsumption tests to take account of context
  - Terminology binding can allow context represented in other ways to be added to expressions prior to subsumption testing

---

ihtsdo SNOMED CT
Delivering
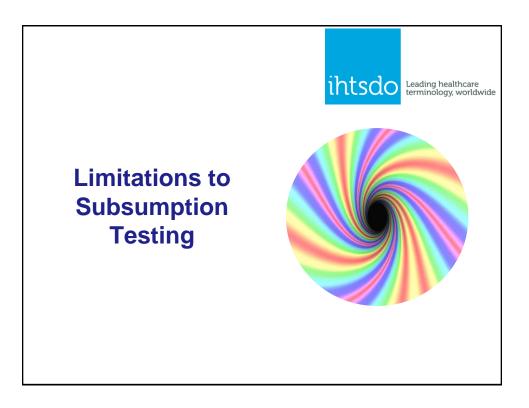
## Subsumption of Concepts with Negative Contexts

- The context model includes context values that indicate negative information about findings or procedures
  For example
  - |finding context| = |known absent|
  - |procedure context| = |not done|
- When working with these type of context the logical direction of subsumption is reversed
  For example
  - Subject does <u>not</u> have a |fracture of tibia|
    - Does <u>not</u> imply that they have no |fracture of bone|
    - Does imply they have <u>no</u> |fracture of shaft of tibia|

> ***Note***: There are other complexities associated with the way negative statements like these are used in practice and detailed consideration of this is beyond the scope of this presentation.

**Limitations to Subsumption Testing**

---

## Subsumption Tests and Primitive Concepts

- Description Logic classifiers can infer whether …
  - A fully defined concept or expression subsumes another fully defined concept or expression
  - A fully defined concept or expression subsumes a primitive concept
- Description Logic cannot infer whether a primitive concept subsumes another concept unless either …
  - There is a stated |is a| relationship between the potential subtype concept and the primitive concept; or
  - The potential subtype concept is subsumed by a concept that is itself subsumed by the primitive concept
- Description Logic cannot infer whether a primitive concept subsumes an expression unless
  - One of the focus concepts in the expression is either the primitive concept or a concept subsumed by the primitive concept

# Optimizing Expression Subsumption Tests

ihtsdo — Leading healthcare terminology, worldwide

---

## Using a Description Logic Classifier to Test Expression Subsumption

ihtsdo Delivering SNOMED CT

- Description Logic classifiers
  - DL classifiers are used to classify SNOMED CT concepts
  - They can also be used to check subsumption between expressions
    - Logically an expression is a definition of an anonymous concept (i.e. a concept without a SNOMED CT identifier)

> **Note**: Implementation Course Module F includes a presentation on practical use of a Description Logic classifier.

## Using an Expression Repository to Optimize Expression Subsumption Tests

- An expression repository is a database containing all the expressions used in an institution
  - Expressions are added to the repository when first used
  - Each expression is allocated a unique internal identifier
  - The internal identifier is used to represent the expression in health records
- An expression repository can be classified using a Description Logic classifier
  - In the same way as SNOMED CT concepts are classified
- A transitive closure table including expression identifiers allows expression subsumption to be tested rapidly for retrieval and analytics

## Summary

- Different expressions can represent the same clinical meaning
- Expressions can be transformed to a common normal form to facilitate equivalence and subsumption tests
- All concept identifiers in a normal form expression refer to primitive concepts
  - During normalization, each reference to a fully defined concept is replaced by its proximal primitive supertypes and it defining attributes
  - A canonical form orders expression elements, excludes terms and removes spaces
- Context can be taken into account during expression normalization
- Methods for rapid expression subsumption tests include
  - Use of Description Logic classifiers and/or an expression repository

## Links to Further Information

- Optional materials available in Implementation Module C
  - A more detailed presentation on normalization and subsumption processes
  - A short handout on using the SNOMED CT Brower to identify proximal primitive supertype concepts
- SNOMED CT Technical Implementation Guide
  - 7.8.2.4 - Expression retrieval and normal forms
    - http://snomed.org/tig