Semantic Interoperability and SNOMED CT: A Case Study in Clinical Problem Lists

James R. Campbell M.D.[a],

Alejandro Lopez Osornio M.D.[b]

Fernan de Quiros[b], Daniel Luna[b], Guillermo Reynoso[c]

[a]University of Nebraska Medical Center, Omaha, NE

[b]Departamento de Información Hospitalaria, Hospital Italiano de Buenos Aires, Argentina

[c]Conceptum, Buenos Aires, Argentina

University of Nebraska Medical Center

HOSPITAL ITALIANO de Buenos Aires

Abstract

Sharing clinical data between institutions requires a shared semantic model and reproducible metrics for comparing post-coordinated clinical record data. We evaluated the interoperable features of SNOMED CT by merging and comparing SNOMED problem list dictionaries between primary care sites in Nebraska and Argentina.

The two problem subsets had modest differences in semantic content, but differed substantially in frequency of post-coordinated content. 64.3% of the final merged problem vocabulary was post-coordinated. 0.9% of the merged concept set was found to have semantic duplication. Classification employing SNOMED Normal Forms was found to effectively identify semantic equivalency in 65.2% of cases studied.

The most frequent reasons for failure of the classification procedures included: post-coordination not compliant with SNOMED guidelines (28.8% of cases), conflicting guidance presented in SNOMED pre-coordinated content (3.6%), unclear definition of source clinical concepts (1.2%), and limited expressiveness of the SNOMED definitional model (1.2%).

We conclude that interoperable post-coordination employing SNOMED CT is feasible, but that better dissemination of SNOMED editorial guidelines, refinement of the concept model and training of the user community will be required for reproducible and reliable implementation.

Introduction

- Semantic interoperability requires that advanced compositional vocabularies such as SNOMED CT[1] support computation of equivalence between post-coordinated concepts authored at disparate sites
- Advancing to multi-national status with formation of a new international standards development organization[2], SNOMED CT must reliably support interoperability world-wide
- We analyzed the validity of algorithms for testing of semantic equivalence while comparing SNOMED problem lists at two institutions

Methods: Study sites

- University of Nebraska Medical Center; Omaha, NE, USA

  Clinics of the University of Nebraska Medical Center employ the GE Centricity Enterprise record which includes problem list and assessment functions with SNOMED CT as coding reference. Employing "just-in-time" post-coordination[3], UNMC has developed an extension to SNOMED CT which includes 6% post-coordinated concepts within a problem data dictionary of 9800 concepts and 14500 interface terms.

- Hospital Italiano; Buenos Aires, Argentina

  Hospital Italiano is a non-profit university hospital, associated with a large network of ambulatory care services, located in Buenos Aires, Argentina. Hospital Italiano employs a health information system developed in-house including an institution wide electronic medical record supporting all levels of care. Clinical data entry uses a local interface terminology for diagnosis and procedures. This terminology was created from 2 million terms of free text recorded by local users in legacy systems. Every concept from the terminology has been coded with SNOMED CT. Current Hospital Italiano terminology has 24,800 concepts and more than 100,000 interface terms, with approximately 70% post-coordinated content.

University of Nebraska Medical Center

HOSPITAL ITALIANO de Buenos Aires

Methods

- SNOMED CT problem lists were merged and compared in order to develop a shared problem list vocabulary for primary care
- The merged concept inventory was processed with a classifier compliant with SNOMED guidance[5] in order to compute Long Normal Forms[6] for each concept

Methods

A sample of the classifier output was systematically evaluated by the authors:

1) verifying the intended meaning in the original concept term
2) verifying SNOMED editorial guidance for modeling of the meaning
3) validating the post-coordinated concept against published SNOMED editorial principles[4]
4) verifying true synonymy of content where asserted by the classifier
5) assessing for “hidden synonymy” against concepts classified as unique by intensive review of related concepts and alternative post-coordination strategies.

Classification and comparison of post-coordinated concepts

SOURCE

PRE-COORDINATED

CLASSIFICATION RESULTS

CAP / SNOMED

85225000 NASAL SINUS CYST
IS A==>MASS OF RESPIRATORY STRUCTURE
IS A==>DISORDER OF NASAL SINUS
ASSOCIATED MORPHOLOGY==> CYST

85225000 NASAL SINUS CYST
IS A==>DISEASE (DISORDER)
ASSOCIATED MORPHOLOGY==> CYST

In this exemplar, Argentina and Nebraska sites have Independently authored concepts for “Quiste sino Maxilar” and “Maxillary sinus cyst”. Even though the modelers have chosen different parent concepts, the classification software identifies that they are equivalent concepts, and clearly different than the pre-coordinated concept “Nasal sinus cyst” in the standard SNOMED CT release.

LENCE

ARGENTINA

LENCE

NEBRASKA

2570001000004101 MAXILLARY SINUS CYST
IS A==>CYST OF NASAL SINUS (DISORDER)
ASSOCIATED MORPHOLOGY==>CYST
FINDING SITE==>MAXILLARY SINUS STRUCTURE

2570001000004101 MAXILLARY SINUS CYST
IS A==>DISEASE (DISORDER)
ASSOCIATED MORPHOLOGY==>CYST
FINDING SITE==>MAXILLARY SINUS STRUCTURE

Results

	Nebraska	H Italiano	Merged
Pre-coordinated	9734 (94.3%)	7666 (23.7%)	14069 (35.7%)
Post-coordinated	585 (5.7%)	24727 (76.3%)	25312 (64.3%)
Total concepts	10319	32393	39381

Problem list subsets at the two institutions were notably different in concept inventory, frequency of post-coordination and semantics. This table summarizes the number of concepts at each institution that were pre- and post-coordinated.

University of Nebraska Medical Center

Hospital Italiano de Buenos Aires

Results

	Nebraska	H Italiano	Merged
Disorders	7467 (72%)	15891 (65%)	20633 (66%)
Findings	1372 (13%)	3516 (27%)	7328 (23%)
Events	30 (0.3%)	82 (0.3%)	104 (0.3%)
Procedures	1098 (10.6%)	112 (0.5%)	1201 (3.8%)
Situation w/ explicit context	339 (3.2%)	864 (3.5%)	1178 (3.7%)
All other	(0.9%)	(3.7%)	3.2%)

This table summarizes the differences in SNOMED semantic inventories between the two institutions and the merged problem list subset. Note that Nebraska employed more procedure history and Hospital Italiano more findings.

Results

Failure of post-coordination	
Vague source concept utterance	3 (1.2%)
Limited expressiveness of SNOMED model	3 (1.2%)
Conflict of SNOMED guidance	9 (3.6%)
Non-compliance with SNOMED editorial guidelines	72 (28.85)
Subtotal of classification errors	87 (34.8%)
Masked synonymy	3 (1.2%)

We sampled 250 concepts classified as unique from the merged subset and systematically evaluated the post-coordination and classification output for accuracy. Our review supported reasonable and accurate performance of semantic equivalency testing in 64.8% of cases. This table summarizes the problems we identified during detailed study of the classifier analysis of post-coordinated concepts.

Discussion / Conclusion

- Studies have demonstrated the limitations of conventional classifications and confirmed the requirements of compositional forms to deal with an ever expanding clinical knowledge base[7]
- Our research supports the utility of SNOMED Normal Forms for comparing disparate data sets, however…
- The greatest deficiency we identified in post-coordination was inconsistent behavior in the data modeling staff at the two institutions
- Furthermore, the SNOMED standards board must extend support for post-coordination by:
  - Addressing consistency within the SNOMED CT data model
  - Document and disseminate SNOMED data model standards
  - Promoting open discussion and training in SNOMED use
  - Providing clear guidance supporting reproducible and understandable post-coordination algorithms

Bibliography

[1] Spackman, K.A. (2001) SNOMED CT: The Next Generation of Terminology for Global Health Care. World Markets Research Centre's "Next Generation Healthcare 2001 Business Briefing." 2001.

[2]http://www.snomed.org/news/documents/snomed_meeting_statement_nov2006_update.pdf

[3] Warren JJ, Collins J, Sorrentino C, Campbell JR. Just-in-time coding of the problem list in a clinical environment. Proc AMIA Symp. 1998;:280-4.

[4] SNOMED Clinical Terms: Concept Modeling Style Authoring Guide. College of American Pathologists. October 2004.

[5] SNOMED Clinical Terms®Guide: Transforming Expressions to Normal Forms. http://www.snomed.org/snomedct/documents/transformations_to_normal_forms.pdf, 2005.

[6] Spackman, K.A. (2001) Normal Forms for Description Logic Expressions of Clinical Concepts in SNOMED RT. Proceedings of the 2001 AMIA Fall Symposium. November 3, Washington, D.C., pp. 627-631.

[7] Campbell JR, Carpenter P, Sneiderman C, Chute CG, Warren JJ. Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity, for the CPRI Workgroup on Codes and Structures. JAMIA 1997 May-June; 4(3):238-251.

Address for correspondence:
James R. Campbell MD
Department of Internal Medicine
University of Nebraska Medical Center
983331 Nebraska Medical Center
Omaha NE 68198-3331

University of Nebraska Medical Center

HOSPITAL ITALIANO de Buenos Aires