MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

A Practical Approach to Advanced Terminology Services in Health Information Systems

Maria Laura Gambarte [a], Alejandro Lopez Osornio [a], Marcela Martinez [a], Guillermo Reynoso [b], Daniel Luna [a], Fernan Gonzlez Bernaldo de Quiros [a]

[a] *Medical Informatics Department, Hospital Italiano of Buenos Aires, Argentina*
[b] *Conceptum, Buenos Aires, Argentina*

Abstract

As the medical informatics field evolves, new functions appear as the focus of interest; a more advanced management of terminology is one of them. Using comprehensive and detailed terminology to represent clinical rules in computer systems, associated with patient information, would allow clinical software to provide patient specific recommendations or alerts. In order to uniform data collection through our HIS, and lay the foundations for future clinical decision support systems, we decided to move from our previous classification- based medical record into new terminology services built around Snomed CT, Spanish Language Version. This paper describes the characteristics of our Terminology Server.

The most important achievements of our new terminology system are the centralization of knowledge representation, using a much more detailed terminology system. Clinical data entered at any place of the institution and level of care, is represented uniformly through the whole health information system.

Keywords:

information storage and retrieval, classifications, terminology services

Introduction

Until recently, implementing vocabulary control in medical informatics implied selecting the most appropriate classification for the specific clinical scenario. This was a reliable solution for the intended objectives, like epidemiological description of reality, billing or simple chronic diseases detection on large populations.

As the medical informatics field evolves new functions appear as the focus of interest; a more advanced management of terminology is one of them. New Health Information Systems (HIS) need to store information in a way it can not only be interpreted by human users, but also by computer powered knowledge management rules (1). Encoding clinical rules in computer systems, associated with patient information, would allow clinical software to provide patient specific recommendations or to focus health professional attention to situations that may otherwise be overlooked (2).

The purpose of this paper is to describe our approach to this advanced knowledge representation system in the electronic medical records at the Hospital Italiano of Buenos Aires, with emphasis on data entry, storing and extraction processes.

Background

The most important new functions of complex knowledge representation systems are interoperability and clinical decision support.

Current HIS are designed to link all levels of care, like the primary care office, nursing stations, intensive care units, rehabilitation, etc. Vocabulary control in this new setting gets more complicated, and as each part of the system uses a different classification or list of terms we run the risk of getting into a Tower of Babel situation. Vocabulary control tools must have the ability to serve all settings, and provide an adequate level of detail to each one, as general concepts for ambulatory reasons for encounter or nursing diagnoses, as well as very specific concepts such as specialist registries.

Clinical information systems are expected to improve the quality of patient care. This can be achieved by linking clinical information stored in their databases with current scientific recommendations to bring decision support to physicians, or by selecting high-risk populations where population level interventions proved to bring clear benefits. These features require a strict vocabulary control of patient data and clinical rules data.

Old Classification-based coding systems, common in current Electronic Health Records (EHR) software, are not up to these new challenges. They fail to provide inter-setting validity, they have not the adequate granularity for different levels of care and they are also unsuitable for linking knowledge rules for clinical decision support systems.

To provide solutions to these new functions of the medical record, an advanced knowledge representation system is needed. Implementing this tool in institution-wide health information systems represents a great challenge.

There are some examples of the state of the art terminology systems with knowledge-based representations of medical concepts, like Snomed CT (3), Unified Medical Language System (UMLS) (4), the Medical Entities Dictionary (MED) (1), Nebraska Medical Center (5) or Kaiser Permanente Convergent Medical Terminology (6). Nebraska and Kaiser are based on Snomed CT.

Snomed CT is a nomenclature originally created by the College of American Pathologists (CAP), and is evolving into an international Standards Development Organization (SDO). This is currently regarded as the most advanced initiative in knowledge-based representations with clinical application. Each of the 300,000 terms included are defined using relationships with other terms, creating a powerful semantic network. Simple algorithms allow easy finding of sub types, super types and attributes of any given term. Snomed CT data model allows the continuous extension of the nomenclature, adding new terms, always following the same compositional concept representation model, called Description Logics (7, 8).

These characteristics give Snomed CT a great role as the first building block of a terminology implementation, as has been proposed in previous experiences, either private as Kaiser Permanente or Cerner (9), and governmental as the cases of health departments of the United Kingdom, Australia, and the United States.

Chute et al. outlined three functions required in a terminology system (10), as it was previously described by the U.S. Computer-based Patient Record Institute (CPRI) during the National Conference on Terminology in November 1997. These three functions or layers are:

Entry Terminology: this is the user interface, the terms used to interact with users during data entry process

Reference Terminology: this is the format used to store data, knowledge information should be stored in this layer

Aggregate Terminology: different data formats outputs for user analysis

Snomed CT provides functionalities in these three layers. As indicated before, it has every important feature of a *reference terminology*. Snomed CT also includes several descriptions that can be used as an *entry terminology*. Finally, Snomed CT has a standard cross mapping model; the official distribution includes data for mapping to ICD-9CM (ICD - International Classification of Diseases). Additional ICD-10 cross map data has also been developed in the UK, but is not as widely available, and more new cross maps are under development, like ICPC-2 or NANDA. Snomed CT is included in UMLS (11), which provides some degree of relationship with many other terminologies. These mappings provide the *aggregate terminology* features to Snomed CT.

In order to uniform data collection all through our HIS, and lay the foundations for future clinical decision support systems, we decided to move from our previous classification-based medical record into new terminology services built around Snomed CT.

HL7 (Health Level 7) organization has published a Common Terminology Services (CTS) standard specification. At this stage of the project we decided to create terminology services to provide a solution to the local terminology needs, in a second stage a CTS compatible API (Application Programming Interface) will be created as a new layer that will allow standard access to our services.

Setting

The Hospital Italiano of Buenos Aires is a 650-bed non-profit university hospital located in Buenos Aires, Argentina. More than 150,000 ambulatory visits and 3,000 hospitalizations are registered every month. It is affiliated with a Health Maintenance Organization (Plan de Salud) that takes care of a population of 140,000 patients.

Since 1998 a full scale HIS has been gradually implemented, including ambulatory Electronic Medical Record (EMR), inpatient discharge summary, administrative systems, scheduling systems, inpatient tracking systems, pharmacy systems and complementary studies report and visualization. Several health informatics standards had been implemented, including HL7, CDA Version 2, ICD-9CM, DRG, ICD10, and ICPC.

Previous vocabulary control system consisted in professionals entering free text descriptions, which were later assigned classification codes by a group of coders (12).

Design objectives

Our main objective was to design a new terminology system, whose objectives can be related to each of the functions of the terminology system previously described (entry, reference and aggregate terminology).

Entry terminology functions:

- Provide an institutional vocabulary for all user interfaces, so our professionals interact with familiar terms, including local jargon and preferences.
- Provide concept lookup functions, with loose lexical matches and options, to be used during data entry process of new items in a problems list or similar user interfaces.
- Provide short pick-lists definitions for more structured data entry in specific use templates, with a short list of valid entries, like occupation, level of consciousness, etc.
- Provide navigational tree interfaces for structured data entry in more complex templates, like cause of liver failure, etc.
- Provide different preferred terms for the same concept in different settings.
- It should include the ability to accept new terms from the user, in case a concept or description is not represented.
- It should detect inappropriate terms for being too general or not valid in a subset, as in the case of problem lists entries Emergency Consultation or Cardiology, etc.

Reference terminology functions:

- The entry terminology should be represented in the reference terminology (Snomed CT Spanish Language Version).
- New concepts should be created for institutional terms that cannot be represented with a standard Snomed CT code.
- The system should provide tools to take advantage of the knowledge stored in Snomed CT relationships, like obtaining more refined or more general terms.
- The system should provide means of updating to new versions of Snomed CT without loosing information, nor requiring large amounts of manual updates.

Aggregate terminology functions:

- Provide output to several standard classifications:
- ICD-9CM (diag & procedures)
- ICD-10 (diagnosis)
- ICPC-2 (diagnosis) (International Classification of Primary Care)
- ATC (drugs) (Anatomical Therapeutic Chemical Classification)
- Local billing nomenclatures
- Aggregate data according to Snomed CT hierarchies

All desired functions should run on a centralized software and data structure, referred to as a Terminology Server. The Terminology Server should provide these functions to all existing applications in the Health Information System in the form of Web Services.

A terminology maintenance software application should also be developed to administrate the institutional terminology, its relationship with Snomed and the mappings.

System description

Entry terminology

The institutional entry terminology is composed of concepts and descriptions. We use Snomed definition of these terms (13), where concepts represent distinct clinical meanings and descriptions are a phrase used to name a concept.

The development of the entry terminology is described with more detail in other paper in process.

Our institutional entry terminology can be divided in several subsets; examples of these are:

- Problems list terminology
- Procedures terminology
- Findings in chest radiography
- Administration routes for drugs
- State of consciousness description
- Physical examination subset
- Liver Failure Diagnosis

Some subsets are very large, including thousands of concepts (i.e. the problems list subset). Others are short lists (i.e. the liver failure subset). Each subset was designed in order to be used as the entry terminology in a specific scenario.

Concepts are defined only once, regardless of its inclusion in more than one subset; therefore, accessing liver cirrhosis from the problems list or from the Liver failure subset brings the user to the same concept.

The process of adding concepts to the entry terminology and organizing them in subsets is manual. This is done by trained coders that were previously working with the same information in secondary coding using classifications (12).

Construction of the problems list subset was one of our biggest challenges. We decided to base our work on the historic database of our EHR with more than 2 million free text inputs since 1998. All problems list entries and discharge notes were processed to extract all different textual descriptions, this strings were originally entered with a 50 character limit. We considered that these texts, entered by our own professionals in a completely unconstrained way, would be representative of the local natural language, including abbreviations and jargon. A manual depuration process, assisted by string normalization functions, led to the creation of the Problems List subset with 24,800 different concepts, with 110,000 descriptions in total.

Other subsets were created using arbitrary lists of concepts selected by the clinical terminology team with user input, like the case of Liver failure or Routes of administration subsets.

New concepts or descriptions were accepted from user interfaces and stored for manual evaluation.

The data model for the entry terminology was the standard Snomed CT data model for concepts, descriptions and subsets (14). New concepts and descriptions were added to the standard Snomed CT distribution following official Snomed rules for creating institutional extensions.

Reference terminology

Implementing the reference terminology involves establishing a very detailed terminology as the format in which clinical information will be ultimately stored. Reference terminology is invisible to the users: therefore, it is necessary to define the relationships of each concept included in the entry terminology with the concepts included in the reference terminology.

We used Snomed CT Spanish Language Version as the reference terminology, but is important to note that all different language versions of Snomed CT share the same concepts and relationships. During the translation process only new descriptions are added.

Both entry and reference terminologies were stored following the Snomed data model, and using Snomed tools to represent the concepts of the entry terminology. Snomed CT defines concepts by its relationships with others, so we created new relationships as part of our Snomed CT extension.

Snomed CT has around 300,000 concepts, but in a clinical setting, health professionals usually use very detailed expressions, adding modifiers to general concepts, like mild ankle sprain. To prevent the exponential growth of the nomenclature, Snomed CT avoids including such level of combination with modifiers, providing the general concepts (ankle sprain), the possible modifiers (mild) and the rules to correctly relate them (using the has severity relationship). Any new concept can be represented using this post-coordination technique, creating more detailed subtypes of existing Snomed CT concepts.

Around 33% of the concepts included in the Problems List subset could be directly mapped with existing Snomed CT concepts; the other 77% needed the addition of one or more modifiers (post-coordination) in order to fully represent the meaning of the entry terminology concept. This rate of post-coordination was dictated by a very permissive policy allowing the use of any term requested by the users, often very specific or personalized. The total of 24,800 concepts was represented with 45,000 new relationships.

Aggregate terminology

The official Snomed CT cross maps model was implemented, a multi-classification interface was created as part of the Terminology Maintenance Software to visualize, test and modify mappings from Snomed to different classifications.

An SQL algorithm was designed (Oracle SQL) to aggregate concepts according to knowledge stored in Snomed CT relationships, like all kinds of diabetes, including diabetes complications and excluding maternal and neonatal diseases. These queries are maintained from a module in the Terminology Maintenance Software.

Terminology server

We developed a set of software services to provide the accesses to the new terminology system to other applications:

Lexical Concept Lookup: introducing a keyword retrieves exact matches or a set of similar concepts from a given subset.

Short List Retrieval: using the subset identifier, retrieves a list of concepts to include in a pick list in user interfaces.

Navigational Tree Retrieval: using the subset identifier, retrieves a list of concepts and its relationships to create tree selector in user interfaces.

New Concept Proposal: this service accepts new concepts and adds them to the pending list in the Terminology Maintenance Software.

Map to Classification: given any entry terminology concept identifier and any available classification provides the resulting code in such classification.

Terminology maintenance software

The Terminology Maintenance Software includes the following modules:

- Entry Terminology Administration: allows the creation of new concepts, description assignment and modeling of each concept with Snomed CT.
- Subset Administration: creation of new subsets, addition and removal of concepts from the subsets, defining hierarchies for tree interfaces.
- Pending Concepts or descriptions: all proposed new concepts or descriptions are stored in a list, waiting to be evaluated and modeled, ordered by the number of proposals
- Cross Maps Administration: existing cross maps can be visualized, edited and tested using this module
- Data Extraction Rules Administration: A software interface to visualize and update Snomed based data extraction queries.

Invalid terms administration

In each subset, professionals usually try to enter terms that are not valid for later use, like trying to write Cardiology in the problems list. We would like the doctor to record the proper diagnosis or reason for encounter instead.

In order to reject these terms we tag them and add an information text so the professional understands the coding guidelines of the institution. This module provides the tools for tagging these terms and editing the information.

Status report

Four trained modelers are maintaining the interface terminology, modeling pending concepts or descriptions, running routine quality control checks, and maintaining subsets.

We created an ad-hoc automatic process to recode all historic data in our clinical repository, using string matching algorithms; more than 2,200,000 entries were processed. Around 85% of the original texts received a concept code of the new entry terminology, 10% of them were recognized as invalid entries: therefore, 75% were finally mapped to Snomed CT.

The coding services are used on-line by our ambulatory and inpatient medical record, receiving around 55,000 requests each month.

Mapping to ICD-9CM from Snomed is not fully functional. It is currently running with an approximate effectiveness of 50%, and is under revision by the modelers. Previous manual codification processes are still in use until cross maps system performance improves. The use of the ICD-9CM cross-map will be addressed in a future paper.

Data extraction from the clinical information repository, using Snomed-based aggregation functions, runs daily to feed lists of target patients to a Chronic Disease Patient Registry (15). Program managers referred a great improvement in patient detection rates as compared to the previous classification-based procedure.

Discussion

Snomed CT provided an excellent structure for the initial organization of our terminology system, even when documentation not always is fully comprehensive and some proposed models will be subjected to changes in the near future due to Snomeds technical board revisions. Snomed CT has been regarded as too complex to use in real world applications; we think that properly defining the entry terminology and cross maps are essential steps in its utilization.

The task of creating an institutional entry terminology demands a lot of work, but provides an excellent service to the users, and also isolates the terminology system from Snomed CT changes in newer versions. Local concepts will always be valid, and in the worst case a correction of modeling against Snomed CT would be required.

We found that Snomed CT cross maps data to ICD-9CM is still not adequate for clinical use in our setting, requiring additional manual work on the maps. This may be caused by a different use of the classification in Argentina and the United States.

Our clinical data extraction process, using rules based in Snomed CT knowledge data, is very effective; however, these rules should be revised for each new Snomed CT version, as changes in hierarchies and models may affect its effectiveness.

Further reduction of manual classification coding will require adjustments of mapping specifications and user interface changes, aimed to reduce the number of new concepts proposals, and enforcing the selection of existing terms. Due to acceptability issues, we have always tried to minimize user interface constrains, thus implementation of these changes will be a slow process.

The most important achievements of our new terminology system are the centralization of knowledge representation, using a much more detailed terminology system. Clinical data entered at any place of the institution, and level of care, is represented uniformly through all the health information system.

References

[1] Cimino JJ. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. J Am Med Inform Assoc. 2000 May-Jun;7(3):288-97.

[2] Munsen M SY, Shortliffe E. Clinical Decision-Support Systems. In: Shortliffe E CJ, editor. Biomedical Informatics: Springer; 2006.

[3] Wang AY, Sable JH, Spackman KA. The SNOMED clinical terms development process: refinement and analysis of content. Proc AMIA Symp. 2002:845-9.

[4] Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc. 1998 Jan-Feb;5(1):1-11.

[5] 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.

[6] Dolin RH, Mattison JE, Cohn S, Campbell KE, Wiesenthal AM, Hochhalter B, et al. Kaiser Permanente's Convergent Medical Terminology. Medinfo. 2004;11(Pt 1):346-50.

[7] Spackman KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. Proc AMIA Symp. 2001:627-31.

[8] Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Proc AMIA Symp. 1998:740-4.

[9] Cerner. Cerner and SNOMED Enhance the Sharing of Global Electronic Health Information 2004. (http://www.cerner.com/public/AboutCerner_2a.asp?id=3476)

[10] Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. Proc AMIA Symp. 1999:42-6.

[11] Fung KW, Hole WT, Nelson SJ, Srinivasan S, Powell T, Roth L. Integrating SNOMED CT into the UMLS: an exploration of different views of synonymy and quality of editing. J Am Med Inform Assoc. 2005 Jul-Aug; 12(4):486-94.

[12] Luna D, Bernaldo de Quiros FG, Garfi L, Soriano E, OFlaherty M. Reliability of secondary central coding of medical problems in primary care by non medical coders, using the International Classification of Primary Care (ICPC). Medinfo. 2001;10(Pt 2):300.

[13] Snomed. SNOMED Clinical Terms User Guide: College of American Pathologists; 2005.

[14] Snomed. SNOMED Clinical Terms Technical Reference Guide: College of American Pathologists; 2005.

[15] Dawidowski A, Figar S, Langlois E, Marchetti M, Baum A, Montenegro S, et al. Sensibility of a chronic disease management program software to identify diabetic patients. 8th SIS; 2005 September 2005; Santa F, Argentina: Sociedad Argentina de Informtica e Investigacin Operativa (SADIO); 2005.