

Clinical Terminology and Clinical Applications: Mind the gap

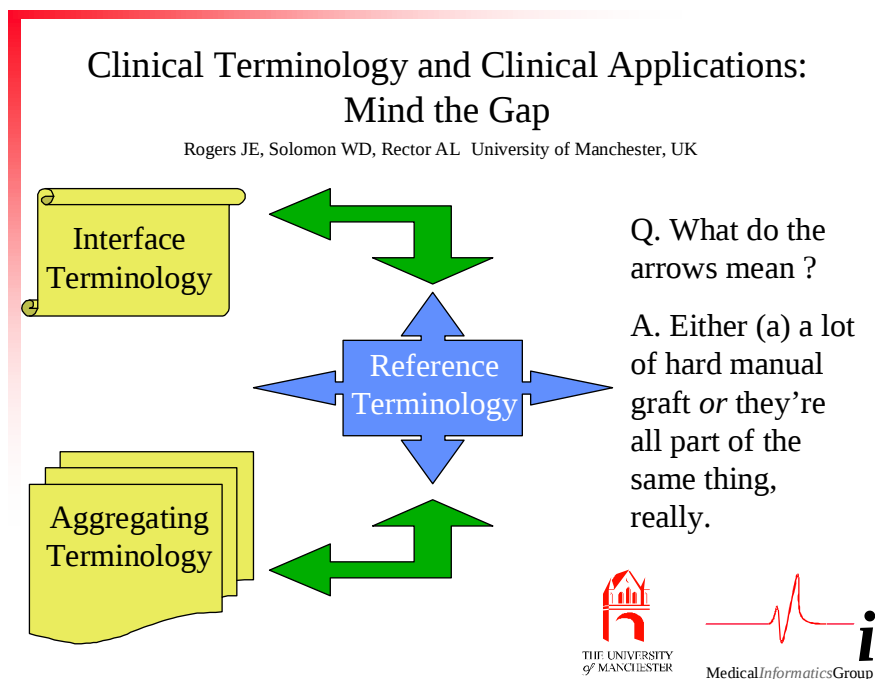
Jeremy Rogers, Danny Solomon, Alan Rector
Medical Informatics Group, Department of Computer Science, University of
Manchester, UK

<http://www.cs.man.ac.uk/mig/> jeremy@cs.man.ac.uk

Speaker Biography

Dr Jeremy Rogers is a Primary Care Physician who escaped to join the Medical Informatics Group at the University of Manchester in 1994, where he has been developing the terminology of the GALEN Common Reference Model of medicine in tandem with the tools to maintain, deliver and use that terminology.

PowerPoint summary slide



Clinical Terminology and Clinical Applications: Mind the gap

Jeremy Rogers, Danny Solomon, Alan Rector
Medical Informatics Group, Department of Computer Science, University of
Manchester, UK

<http://www.cs.man.ac.uk/mig/> jeremy@cs.man.ac.uk

Abstract: *'Reference terminologies' are being developed as entities distinct from those optimised for data capture ('interface terminologies') or for secondary data analysis ('reporting terminologies'). Interchange of data between these alternative representations is necessary. As medical terminology representation adopts compositional techniques, new mechanisms become available and necessary to achieve this data transformation. These suggest a formal relationship can exist between reference and other kinds of terminologies, such that they should no longer be considered independent entities but formal transformations of each other.*

Keywords: terminology galen workstation interface coding mapping

Introduction

An integrated environment to support the clinician across all aspects of the clinical task – the clinical workstation – has been postulated but has proved empirically difficult to build or even to specify. Hospital information systems support management functions, but use of computers by clinicians for clinical functions remains the exception rather than the rule.

A major component of the clinical workstation is expected to be the electronic patient record, whose functionality aims to at least duplicate and preferably to improve upon that offered by the existing paper-based record. However, many of the anticipated improvements over paper appear to require more sophisticated manipulation of the computer-based record than is currently possible. To a great extent we believe these difficulties arise because of limitations inherent in the medical terminology representations used to obtain, store and analyse the content of the medical record.

A recent development, in discussions relating to the broader requirements of clinical workstations and terminology, is the notion of different kinds of clinical terminologies - usually designated by reference to the functionality they aim to support. This paper examines specifically the notions of 'reference', 'reporting' and 'interface' terminologies, and describes our understanding of the nature of the gaps between them.

Limitations of traditional terminologies

The particular hierarchical structures and content of traditional clinical terminologies are optimised primarily to support post-processing of the record for secondary statistical purposes: for reporting and/or retrieval. Correspondingly they are less well suited to supporting the interface functionality required by the primary user of the clinical workstation and the data it captures: the clinician.

Schemes such as ICD, READ, ICPC and SNOMED International [Cot93,Pri96,WHO77] are used, substantially unaltered, in many clinical systems as the means by which the primary user enters the clinical record. The clinician's overall expressivity in this task is determined both by the scope and granularity of the available fixed term list and by the extent to which the navigational structure of the terms supports the process of recording more than one relevant term.

In use, however, this approach increasingly fails to satisfy the intended users. The stresses become most evident when the scope of a scheme is extended to support multiple different kinds of users: as the term set grows it quickly loses focus for any one user, and the original technical conflation of functionality makes further local configuration to restore end-user focus very hard.

This paper suggests that terminology systems can be characterised according to three different functional criteria:

- *Reference* - representation of the concepts underlying the terms
- *Interface* – how the terms are used and expressed in the process of care
- *Reporting* – how the terms are grouped for statistical, management and other secondary information purposes including decision support

Understanding the differences between these and making them explicit is a prerequisite for developing effective terminologies. Transforming between the three is a central task of any overall clinical system.

Interface Terminologies

Clinical System builders have responded to user complaints regarding limited expressivity by developing new styles of data entry interface that are variably independent of traditional clinical terminology. At one extreme, free text entry has always been available and is now increasingly coupled with inexpensive voice recognition software.

Other suppliers have developed *proprietary* term sets for use in structured data entry interfaces. These are tailored and structured to specific user groups and clinical tasks, as opposed to being another superset of all user requirements.

Typical examples in the academic field are ORCA [Van97] and PEN&PAD [Kir96] and in the commercial field [MEDCIN](#) and the terminologies behind clinical systems such as those of [Purkinje](#), [MedicaLogic](#), HBOC and [Clinergy](#).

The special knowledge required in an interface terminology concerns how clinicians usually organise information in different clinical situations and settings – for example, what constitutes a 'cardiac examination' for a young healthy patient in general practice as opposed to an elderly patient on admission to a geriatric ward? Such clinical priorities clearly influence the final documentation result, but in a computer-based system ease of access may be an additional influence – what is easy to enter quickly and routinely?

Data entry, of course, is not the whole story: review of data already entered is also an issue. Clinicians prefer not to see the entire electronic medical record and expect instead either a summary presentation or, ideally, a display of only the information relevant to the current clinical context. In this respect the interface terminology must also support the interface in determining what information from the record the clinician will most likely want to see.

Proprietary interface terminologies therefore set out to improve ease of access for the clinician – both when entering data and when reviewing the record later. However, different clinicians have different ideas of what should be easy to enter or review. Suppliers who have adopted the proprietary interface approach informally report being under pressure from their users to support a continuum of slightly different interface terminologies.

Reporting Terminologies

Improvements in interfaces for primary data capture - whether through proprietary controlled interface terminologies or by completely uncontrolled ones (voice or free text input) - are made at the expense of the secondary data uses. For example, automatic exchange of structured data between different supplier systems is often achieved only by printing the record onto paper at the source and then manually re-coding it at the destination.

Equally, the analysis of recorded data (e.g. for statutory central data returns or local decision support) can not usually be achieved within the interface terminology itself. This is because interface terminologies typically do not contain the kinds of abstraction used by statisticians (e.g. ICD-9-CM Diii Endocrine, Nutritional and Metabolic Diseases and Immunity Disorders) or by decision support algorithms (e.g. anti-anginal preparation; gastrointestinal disturbance). Consequently it is generally necessary to convert or otherwise map all of the data captured in an interface terminology to alternative representations where such analysis is possible.

Reference Nomenclatures, Reference Models, and Language

Mapping between terminologies is hard. As a solution to the growing mapping burden, the notion of a single universal reference terminology has been postulated: an all-encompassing superset representation that either links to every reporting and interface terminology or that supports most of the useful interface or analytic functions within its own structure. If such an object could be constructed, it might no longer be necessary to map interface terminologies to each other for purposes of data exchange or to map them to multiple reporting or analytic representations. Each interface terminology would instead maintain one set of mappings only – to the universal reference terminology – and via this gain access to any other representation.

Three techniques to construct a reference terminology have been tried, and we characterise their results as a reference nomenclature, a reference classification, and a reference model.

UMLS [McC90] sought to provide a universal catalogue and at least approximate cross mappings between the major reporting terminologies. It is really most easily understood as a 'reference nomenclature', a universal registry of 'concept unique identifiers' (CUIs) and corresponding 'lexical unique identifiers' (LUIs). Its strength is its comprehensiveness; its weaknesses are the weakness of the individual terminologies of which it is constructed. Details required for clinical care are more often than not missing, although as UMLS moves to incorporate the results of HL7 and more recent developments in SNOMED and Read this is becoming less true.

SNOMED-International and the UK Clinical Terms (Read Codes) Versions 2 and 3, represent attempts to develop terminologies which are the union of all existing codes – what we describe as 'reference classifications'. Both have internal structure, but both are fundamentally enumerated terminologies – long lists of terms organised manually into structures according to principles that are neither explicitly stated nor represented. Neither scheme has been as successful as their originators hoped.

More recent quests for a reference terminology have employed compositional terminologies underpinned by a formal logical model: SNOMED-RT [Spa98] and GALEN [Rec99]. Compositional techniques fundamentally aim to represent an infinity of arbitrarily complex terms by means of structured collections of more basic terms, of which there is a finite set. They use formal rules to classify the composed terms into a multi-hierarchical structure. The Clinical Terms (Read Codes) V3 represent an enumerated/compositional hybrid.

The implied term-space for a fully compositional scheme - the set of all possible combinations - is typically enormous, and this is what makes the scheme much more expressive than traditional enumerative schemes. However, because the term space is so large, it is not possible to pre-enumerate all possible term combinations in a compositional scheme. Such schemes must therefore choose whether to be pre-coordinated or post-coordinated. In pre-coordinated schemes, only a selected subset of all possible compositions are constructed in a fixed list. In this respect the result is essentially a very well constructed reference classification. In post-coordinated systems, concepts may be composed 'on the fly' and then classified subsequently and dynamically. Such post-coordinated schemes retain the full expressivity of the compositional approach, and in so doing have the potential to be what we characterise as a 'reference model': data transformation is achieved not by reference to pre-coordinated static look-up tables – a giant traveller's phrasebook – but by dynamic reference to a shared set of much more basic concepts and a model of how they interact.

Mind the Gap

A universal reference terminology is an attractive proposition. But a consequence of its existence - in fact, the purpose of its existence - is that it must allow data to be freely translated between it and the other kinds of terminology. The remainder of this paper concentrates on how that translation might be achieved. We begin by characterising two possible mechanisms:

Asserted – in which static mapping tables are drawn up manually according to arbitrary and implicit rules, exhaustively listing 1:1 mappings between each term in the source and reference terminologies.

Algorithmic – in which one or more formal translation rules are specified, such that a mapping or translation for any term may be determined dynamically by computation, with reference to the various semantic and lexical properties of the source and destination terminologies. A prerequisite for this approach is that the necessary semantic properties are represented explicitly and in a form usable by computers. The computer can't process what isn't there.

Manual enumeration of a static declarative mapping table has traditionally been the method of mapping between static, monoaxial terminologies (e.g. between 4-byte READ and ICD). It is inherently a very labour intensive exercise, and beset by difficulties such as terms in the source scheme having no directly equivalent mapping in the destination scheme. Data mapping is often an imprecise 'best fit', and the nature of any detail lost in translation is frequently not explicitly retrievable.

The introduction of post-coordinated ('generative') compositional terminologies introduces new configurations and new mapping challenges. Post-coordination is particularly attractive in a data entry interface, because of its increased expressivity. One example is the PEN&PAD/Clinergy™ interface [Kir96]. It is less important in an analytic terminology, and may even be positively discouraged for statistical reasons. A likely configuration, therefore, is where a generative interface terminology must be mapped to a static reporting or reference terminology. This poses a special problem for the static mapping table approach: how do you construct exhaustive static mapping tables from a dynamic resource?

Dynamic algorithmic mapping between terminologies is clearly an attractive alternative to traditional manual mapping technology. It offers the promise of:

- automated management of updates to source or destination terminologies
- reductions in inconsistency, error and costs inevitable in the manual enumerative approach
- dynamic and exhaustive bi-directional mapping between generative schemes

However, if an algorithmic approach can determine such mappings then this implies (for example) that interface terminologies are no longer entirely independent of reference terminologies but are instead a systematic transformation of them **and vice versa**. The corollary of this is that such a systematic relationship is a prerequisite to making the algorithmic approach work.

There is an additional benefit to be gained in a world where terminologies are maintained as systematic views on something else: the demands of users for a continuum of different interface terminologies can be delivered by adjusting extraction algorithms rather than by less principled alterations of term lists. Further, with the nature of the differences between the resulting variants thus made explicit, the relationship between them can be computed.

Conclusion

Interface and reporting terminologies are currently viewed as entirely distinct from each other and from reference terminologies. However, this may be an illusion born of the fact that no mechanism has historically been available to explore or instantiate a formal relationship between them. This situation can be changed with the introduction of compositional schemes, and in particular with post-coordinated generative schemes. Continued development of terminologies in isolation from each other and without special reference to mechanisms to bridge the gap between them is a missed opportunity.

References

- [Cot93] Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L (eds), The Systematised Nomenclature of Human and Veterinary Medicine: SNOMED International, College of American Pathologists, Northfield, IL., 1993, 3rd edition
- [Kir96] Kirby J, Cope N, Souza A d, Fowler H and Gain R (1996). The PEN&PAD Data Entry System. *Medical Informatics Europe (MIE-96)*, Copenhagen, IOS Press: 430-434.
- [McC90] McCray, A. and W. Hole (1990). The Scope and Structure of the First Version of the UMLS Semantic Net. *Fourteenth Symposium on Computer Applications in Primary Care (SCAMC-90)*, Washington DC, IEEE Computer Society. 126-130.
- [Pri96] Price C, et al. Anatomical Characterisation of Surgical Procedures in the Read Thesaurus. JAMIA 1996; symp. Suppl.;110-114
- [Rec99] Rector, A.L., Zanstra, P.E., Solomon W.D., Rogers J.E., Baud, R., et al. (1999). Reconciling Users' Needs and Formal Requirements: Issues in developing a Re-Usable Ontology for Medicine. *IEEE Transactions on Information Technology in BioMedicine* 2(4): 229-242.
- [Spa98] Spackman K.A., Campbell K.E. (1998). Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. *Journal of the American Medical Informatics Association* (1998 Fall Symposium special issue): 740-744.
- [Van97] Van Ginneken AM, De Wilde M, Van Mulligen EM, Stam H. (1997) Can data representation and interface demands be reconciled? Approach in ORCA. *Journal of the American Medical Informatics Association* 1997;4 (suppl.):779-83.
- [WHO77] World Health Organisation. International Classification of Diseases, 9th Revision. Geneva: WHO, 1977