EXTENDING SNOMED TO INCLUDE EXPLANATORY REASONING

Kurt L Zimmerman

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Veterinary Medical Sciences

Jeff Wilcke
Bernard Feldman
Taranjit Kaur
Loren Rees
John Robertson

December 1, 2003
Blacksburg, VA

Keywords: SNOMED, Bayesian Network, Expert Systems, NORMEX, Clinical Pathology

EXTENDING SNOMED TO INCLUDE EXPLANATORY REASONING

By
Kurt L Zimmerman

Jeff R Wilcke, Chairman
Pathobiology and Biomedical Sciences

(ABSTRACT)

The field of medical informatics comprises many subdisciplines, united by a common interest in the establishment of standards to facilitate the sharing, reuse, and understanding of information. This work depends in large part on the ability of controlled medical terminologies to represent relevant concepts. This work augments a controlled terminology to provide not only standardized content, but also standardized explanatory knowledge for use in expert systems.

This experiment consisted of four phases centered on the use of the controlled terminology-- Systemized Nomenclature of Medicine (SNOMED). The first phase evaluated SNOMED's ability to express explanatory knowledge for clinical pathology. The second developed the Normalized Medical Explanation (NORMEX) syntax for expressing and storing pathways of causal reasoning in the domain of clinical pathology. The third segment examined SNOMED's capacity to represent concepts used in the NORMEX model of clinical pathology. The final phase incorporated NORMEX-based

pathways of influence in a Bayesian network to assess ability to predict causal mechanisms as implied by serum analyte results.

Findings from this work suggest that SNOMED's capacity to represent explanatory information parallels its coverage of clinical pathology findings. However, SNOMED currently lacks much of the content necessary for both of these purposes. Additional explanatory content was created with an ontology-modeling tool. The NORMEX syntax was defined by SNOMED hierarchy names. Complex sequences of explanations were created using the NORMEX syntax. In addition, medical explanatory knowledge represented in the NORMEX format could be stored in an architectural framework consistent with that used by a controlled terminology such as SNOMED. Once stored, such knowledge could be retrieved from storage without loss of meaning or introduction of errors. Lastly, a Bayesian network constructed from the retrieved NORMEX knowledge produced a network whose prediction performance equaled or exceeded that of a network produced by more traditional means.

Grant information

This work has been supported by the National Institutes Of Health, National Library Of Medicine Informatics Research Fellowship Award, grant number 1 F37 LM07699.

Additional support for this work has been received from Lilly Research Laboratories and the Virginia Maryland Regional College Of Veterinary Medicine.

Dedication

To I Am Whom I Am

The beginning and the end of all things

To Leslie Allen

My friend, my love and my wife. We have grown in understanding and love as our lives have intertwined in support of one another. I love you.

To The Memory Of Earl Zimmerman

My father and instructor in this life. You taught me how to do what was right and how to work hard every day. Thank you for giving me these gifts and your love.

To Sondra Zimmerman-Dickson

My loving and caring mother. You taught me by example the meaning of unconditional love. You have given me the greatest gift of all.

To Jerry Masty

My best friend in the world. This would not have been possible without you.

To Phil Sponenberg

There is a proverb that recommends we surround ourselves with wise counsel. You have been both a wise counselor and a good friend to me.

To Jeff Wilcke

This work and much of my education have been made possible by your help. Thank you for your patience, instruction and support.

Acknowledgments

This work was not produced by my efforts alone, but represents the many contributions from a wide circle of people. Each in someway has given freely of themselves in the form of support, instruction and encouragement. This acknowledgment can only project a pale shadow of the degree of gratitude and appreciation I feel towards these individuals. I have not repeated the names of my PhD committee members here, but I sincerely wish each of them to know how important they have been to me and how much I have valued their guidance and support throughout this work. To each of you, my appreciation and thanks.

- Bob Duncan
- Christine Holtz
- Clare Fisher-Davies
- Dan Boon
- Elizabeth Bestemen
- Geoff Saunders
- Holly Bender
- John Sullivan
- John Vahle
- Kevin Scheibe
- Larissa Bowman
- Lud Eng
- Penny Livesay
- Renee Prater
- Rick Mills
- Vince Meador

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Glossary

Table of terms / abbreviations and definitions	
Accuracy	Accuracy is the nearness of a measurement to the accepted or true value. (# of true positive results + # of true negative results) / total number of results
Acyclic	See DAG
Algorithm	A sequence of steps for performing a task; often a computer program of code designed to be executed one step at a time[i]
ALP	Alkaline phosphatase; an enzyme in the body associated with the liver
ALT	Alanine aminotransferase; an enzyme in the body associated with the liver
Ancestor	**Bayesian network**: all nodes along a direct trail leading to a focus node. **SNOMED *IsA* hierarchy**: all concepts that are above the focus concept
Arc	Also known as a "link" or an "edge," is a connection between two nodes indicating dependence, and is usually drawn as a line with an arrow at one end (a causes b)[ii]
Arc reversal	A process of changing the direction of the arcs in a Bayesian network. Parent nodes become children and children become parents[iii]
Attribute	A relationship between two concepts. An attribute usually expresses the nature of the relationship. See OAV
Bayesian network	A Bayesian network (also known as a belief network or probabilistic causal network) captures believed relations (which may be uncertain, stochastic, or imprecise) between a set of variables, which are relevant to some problem. They might be relevant because they will be observable, because their value is needed to take some action or report some result, or because they are intermediate or internal variables that help express the relationships between the rest of the variables[iv]
Belief	The belief of a node is the set of probabilities, taking into account the currently entered findings by using the knowledge encoded in the Bayesian network[v]
Child	**Bayesian network**: if there is a link going from node A to node B (tail of link at A), then B is said to be a child node of A. Some refer to B as a "direct descendent" of A. **SNOMED *IsA* hierarchy**: if there is an *IsA* link going from concept A to node B (tail of link at A), then A is said to be a child concept of B

[i] Cawsey A. The essence of artificial intelligence. London ; New York: Prentice Hall; 1997.
[ii] Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.
[iii] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.
[iv] Ibid.
[v] Ibid.

Table of terms / abbreviations and definitions	
Concept	A self-sufficient and singular unit of thought that represents an individual understandable thing or event in the world[vi]
Conditional probability	See Bayesian
Coverage	The degree to which a terminology is capable of representing concepts in a domain
DAG	Directed acyclic graph; a directed graph where there are no pathways along the direction of arc arrowheads which form a loop
Data flow diagram	A graph showing the movement of information within a systems and the process that transform this information
D-connected	In a Bayesian network when, there is an undirected explicit or implicit pathway linking two nodes. Evidence of one will directly influence the probability of the other
Dec_blood_hemo_measure	Altered blood hemostatic measurement, node name
Decision support	A computer-based information system that combines models and data in an attempt to solve nonstructured problems with extensive user involvement[vii]
Decreased result	A relationship in the NORMEX model
Depth first search	A search method that explores a branch of a hierarchy to its end before exploring other branches[viii]
Descendent	**Bayesian network**: all nodes along a direct trail leading away from a focus node. **SNOMED *IsA* hierarchy**: all concepts that are below the focus concept
Directed	Attention is given to the direction of the arrowhead of arcs in a pathway of a Bayesian network
Directed cycle	A looping trail along nodes in a network, formed when link directions are followed[ix]
Disorder	A type of concept in the NORMEX model
Distal	Beyond a point of focus
D-separation	A node in the network that has no influence on the probability of the state of another node is said to be d-separated from the other node[x]
Edge	See arc
EMR	Electronic medical record
Expert system	A computer system that applies reasoning methodologies to knowledge in a specific domain to render advice or recommendations[xi]

[vi] Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
[vii] Turban E, Aronson J. Decision support systems and intelligent systems. 5 ed. Upper Saddle River, NJ: Prentice-Hall; 1995.
[viii] Cawsey A. The essence of artificial intelligence. London ; New York: Prentice Hall; 1997.
[ix] Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.
[x] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.

Table of terms / abbreviations and definitions	
Favors	A relationship in the NORMEX model
Favors decrease	A relationship in the NORMEX model
Favors increase	A relationship in the NORMEX model
Fidelity	The degree of accuracy by which a system can reproduce or emulate a target object
Grandchildren	**Bayesian network**: the collection of nodes along a direct trail one node level away below the child of a focus node. **SNOMED *IsA* hierarchy**: the concepts that are one tree level below the child of the focus concept
Grandparent	**Bayesian network**: the collection of nodes along a direct trail one node level above the parent of a focus node. **SNOMED *IsA* hierarchy**: the concepts that are one tree level above the parent of the focus concept
Granularity	Refers to the relative position of a concept in an *IsA* hierarchy. More granular concepts are more specific concepts found lower in the tree.
Has change	A relationship in the NORMEX model
Has core function	A relationship in the NORMEX model
Has location	A relationship in the NORMEX model
Has object	A relationship in the NORMEX model
Has state	A relationship in the NORMEX model
Heuristic	Informal use of knowledge about a topic area to solve a problem (rules of thumb)
Hierarchy	A tree structure with more general objects coming before objects they dominate
Inc_hep_cellular_inflam	Increased hepatic inflammation; node name
Inc_hep_cellular_injury	Increased hepatic injury; node name
Inc_hep_cellular_necrosis	Increased hepatic necrosis; node name
Inc_skm_cellular_necrosis	Increased skeletal muscle necrosis; node name
Increased result	A relationship in the NORMEX model
Inference engine	The part of an expert system that performs the reasoning
Influences core function	A relationship in the NORMEX model
Interpretive finding	A type of concept in the NORMEX model
Interprets	A relationship in the NORMEX model
IsA	A relationship in the NORMEX model and in SNOMED. Used to construct hierarchies with more general concepts at the top and more specific versions of these concepts found lower in the tree
Iterative	Repeated passes through a loop or repeated steps in a process

[xi] Turban E, Aronson J. Decision support systems and intelligent systems. 5 ed. Upper Saddle River, NJ: Prentice-Hall; 1995.

Table of terms / abbreviations and definitions	
Knowledge base	A collection of facts, rules, and procedures organized into schema[xii]
Laboratory finding	A type of concept in the NORMEX model
LNS	See location nonspecific
Location nonspecific	Location non-specific; used in reference to a type of function concept that has not been defined by other attributes
Location specific	Location specific; used in reference to a function concept that have been defined by other attributes such as *has location*
LOIT	A format for expressing the NORMEX chain were the disorder concept is considered the parent concept in terms of Bayesian network trail
LS	See location specific
Mapping	The process of establishing a connection between a concept in one terminology and another
MTD	Maximum tolerated dose. Dose of a compound that has measurable adverse affect on body physiology short of causing death in most cases l[xiii]
Multiply connected network	A Bayesian network that has multiple directed paths between one node and another
Naïve Bayesian network	Assumption is made that there is independence between parent nodes of the network
Netica	A computer program for working with belief networks and decision networks[xiv]
Network transformation	Modification of a network model which modifies its representation without modifying its meaning. Nodes are absorbed and of links are reversed in such a way that any inference done with the resulting network yields precisely the same results as the original[xv]
NOAEL	No observable adverse effect level; greatest concentration or amount of a substance, found by experiment or observation, which causes no detectable adverse alteration of morphology, functional capacity, growth, development, or life span of the target organism under defined conditions of exposure[xvi]
Node	A node is a component of a belief network or decision network, used to represent a variable (i.e. Scalar quantity) of interest[xvii xviii]
Node absorption	See summing out a variable

[xii] Ibid.
[xiii] Klaassen C, editor. Casarett & Doull's Toxicology, the basic science of poisons. 5 ed. New York, NY: McGrall-Hill Health Professions Division; 1996.
[xiv] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.
[xv] Ibid.
[xvi] Klaassen C, editor. Casarett & Doull's Toxicology, the basic science of poisons. 5 ed. New York, NY: McGrall-Hill Health Professions Division; 1996.
[xvii] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.
[xviii] Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.

Table of terms / abbreviations and definitions	
Nomenclature	A system of terms used in a scientific discipline to denote classifications and relationships among objects and processes[xix]
Normal form	Refers to a way of organizing a database to minimize data redundancy and to avoid updating errors
OAV	Object attribute value triple; one of several methods of representing knowledge; see object, attribute, value
Object	Any part of the perceivable or conceivable world[xx]
Observable analyte	A concept type in the NORMEX model
Observable function	A concept type in the NORMEX model
Ontology	A description of the concepts and relationships that can exist for an agent or a community of agents[xxi]
Parent	**Bayesian network**: if there is a link going from node A to node B (tail of link at A), then A is said to be a parent node of B. Some people refer to it as a "direct predecessor". **SNOMED *IsA* hierarchy**: if there is an *IsA* link going from concept A to node B (tail of link at A), then B is said to be a parent concept of A
Polytree	A singularly connected network[xxii]
Posterior probability	The updated probability that the condition of interest is present after additional information has been acquired[xxiii]
Prevalence	The frequency of the condition under consideration in the population[xxiv]
Prior probability	The probability that the condition of interest is present before additional information has been acquired. In a population this is the same as the prevalence[xxv]
Probabilistic inference	The process of calculating new beliefs for a set of variables, given some findings[xxvi]
Proximal	Before a point of focus
Region	One of three segments of the NORMEX syntax [TOIL, $(FL)_n$, FD]
Relationship	See attribute
ROC curve	Receiver operating characteristic curve; a summary statistics graph showing the relationship between the sensitivity results (y axis) versus 1-specificity (x-axes) results for a test

[xix] Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
[xx] Ibid.
[xxi] Ibid.
[xxii] Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.
[xxiii] Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
[xxiv] Ibid.
[xxv] Ibid.
[xxvi] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.

Table of terms / abbreviations and definitions	
Schema	The organization map of a database's content[xxvii]
Semantic	The meaning of a thing[xxviii]
Sensitivity (test)	The probability of a test detecting a true positive case
Singularly connected	A node in a Bayesian network that has a single parent[xxix]
SKMN	Skeletal muscle necrosis
SNOMED-CT®	Systemized Nomenclature of Medicine - Clinical Terminology; a controlled medical terminology
Specificity (test)	The probability of a test detecting a true negative case
Summing out a variable	The process of removing a node from a belief or decision network, and adjusting the remaining links and node relationships so that subsequent inference done on the remaining nodes will yield the same results[xxx]
Syntax	The legal organization of constituents in a language
TBil	Total bilirubin; derived from the catabolism of hemoglobin in the mammals
Term	Designation of a defined concept in a special language by a linguistic expression
Terminology	A set of terms representing the system of concepts of a particular subject field[xxxi]
TLC	See top level concept
TOIL	A format for expressing the NORMEX chain where the top level concept is considered the parent concept in terms of Bayesian network trail
Top level concept	SNOMED concepts that are immediate children concepts of the root SNOMED concept[xxxii]
Trail	An undirected path following links between nodes in a network
UEP	Unstructured explanation pathway; a chain of explanations that is constructed without any syntax requirements
Undirected cycle	A trail between nodes that is not directed and which forms a loop in the network
Undirected path	A trail between nodes that does not depend upon the link direction
Value	A concept that expresses the state of an attribute; see OAV
VB	Visual Basic; a programming language

[xxvii] Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
[xxviii] Ibid.
[xxix] Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.
[xxx] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.
[xxxi] Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
[xxxii] College of American Pathologists. SNOMED clinical terms consultation document: core structure: SNOMED International; 2000.

Table of terms / abbreviations and definitions	
Vertices	See node

Introduction

The proverb "necessity is the mother of all invention"[xxxiii] certainly describes the work presented in this dissertation. Perhaps the applied nature of work in the field of medical informatics dictates this relevancy. The nidus for the research outlined in the following four manuscripts grew out of an informatics need in the field of clinical pathology. The contained manuscripts present the research layers upon which an applied informatics solution is offered. Considerable time has been invested in answering these research questions, and numerous pages produced in explaining their results. However, the relevancy of the research is best understood in the context of the originally perceived problem. Therefore, a few introductory words are offered to the reader to anchor this research prior to a lengthy discussion of its components.

Acute toxicity studies for the evaluation of novel pharmaceutical compounds generate both anatomical and clinical pathology data. Most pathologists evaluating acute toxicity studies for pharmaceutical companies are veterinarians trained in anatomic pathology but with limited training in clinical pathology. This mismatch between a job requirement and available personnel with a particular expertise is a classic environment for using expert system tools. An expert system tool for this particular problem requires a knowledge base that links the explanations for specific serum chemistry changes to specific disorder concepts. Such an expert system has two fundamental components: a knowledge base and an inference engine.

Early on, the controlled terminology Systematized Nomenclature of Medicine (SNOMED) was identified as a potential knowledge-base source for this project.[1] The most current version, SNOMED-CT, contains over 325,000 concepts located in polyhierarchies.[2, 3] SNOMED is also the only actively maintained nomenclature that includes veterinary concepts. Concepts are defined by more than 1,450,000 relationships referred to as concept-attribute-value triples.[2] These triples, when assembled graphically, produce a semantic network, one of several ways to organize information within a knowledge base.[4] The *IsA* relationship conveys subtype inheritance of supertype properties in the design (e.g., serum alkaline phosphatase measurement *IsA* biochemical test) and is used to establish explicit concept hierarchies.[2, 5] Other relationships such as *Associated_morphology*, *Finding_site*, and *Has_interpretation* are used to build definitions and provide the primary means to link concepts in unrelated hierarchies.[2]

Description logic is a subset of first-order logic that uses a restricted set of operators and inference mechanisms, and is used to formally describe the semantic network of SNOMED.[6-8] Description-logic statements can be evaluated algorithmically for equivalency and truths.[8] The description logic formalization assures consistency and validity of SNOMED's semantic network.[6, 9] Much of the existing content in SNOMED expresses concepts dealing with Disorders, Findings, and Laboratory tests, and attributes

[xxxiii] Plato

that provide explicit relationships between them.[5] At this point in its development, however, SNOMED is unlikely to contain all of the concepts and interconnecting attributes necessary for an expert system's knowledge base in the domain of clinical pathology. New attributes and concepts created to address any content deficiencies should adhere to the consistency constraints imposed by SNOMED's description-logic framework.

Creating an expert system also demands an inference engine.[10] The inference engine in an expert system drives the production of new information by comparing supplied raw data with domain knowledge in the knowledge base. With probabilistic designs, conclusions are produced with their associated degree of certainty, providing a means of relative ranking and utility assessment.[11-13] Probabilistic methods work well in larger domains when complex interactions occur between the input data.[14] A Bayesian inference engine is one such probabilistic method, and was chosen for this project because the sequential probabilities used in the decision-making process are visible and the probability-based designs are able to deal with the uncertainty inherent in medicine.[11, 12, 15-19]

Bayesian networks are an extension of Bayes' theorem, creating, much like a semantic network, a graphical model of variables and associated relationships pertaining to a hypothesis or hypotheses, and are therefore considered intuitive to create and use.[20] Supplied with the given state of some variable within the network (e.g., results of particular laboratory tests), the Bayesian network is capable of inferring the posterior probabilities of the other nodes (e.g., the likelihood that a disorder exists).[21] The network nodes themselves represent the variables or concepts that are important in the decision-making process and may be scalar, categorical, or discrete in nature.[22] The relationships between connected nodes, depicted by connecting arcs, imply conditional dependency, or if absent, conditional independency.[18] The arcs can be viewed as representing a causality relationship between nodes, resulting in strength-of-belief probability about a conditionally dependent node's state. The root nodes of the network generally represent the conclusions or hypotheses that the model is capable of predicting. Prior probabilities are necessary and can be derived from the literature, supplied by experts in the field, or, as in this experiment, by the use of software that can extract the probabilities from retrospective data.[23] While Bayesian networks deal nicely with uncertainty and offer understandable visual representations of causal relationships, it can be challenging to capture and represent this "causal" domain knowledge.

The research of this dissertation examined a plausible use of SNOMED in the creation of an expert system. The project evaluated the ability of SNOMED's structure to convey concepts, attributes, and values in a knowledge base of an expert system. In addition, since the semantic network structure of SNOMED shares some graphical and relational attributes with Bayesian networks, the possibility of using SNOMED-like knowledge to express causal relationships for designing the network itself was explored.

A question arises following the proposed focus on using SNOMED in the development of an expert system – why go to all this effort? Using the knowledge contained in a sophisticated medical terminology system offers certain immediate and long-term benefits, leveraging the time, money, and effort already spent on the medical terminology. In addition, an expert system using such a knowledge base will likely use standard medical terminology in communicating conclusions to its users and to the electronic medical record system, making retrieval, aggregation and analysis of output information computationally easier. Finally, the breadth of the topic coverage within a large medical terminology system suggests that a variety of domain-specific expert systems could be developed from a single source, using a common methodology.

Bibliography

1. College of American Pathologists. SNOMED CT product profile: SNOMED International; 2003.
2. College of American Pathologists. SNOMED clinical terms consultation document: core structure: SNOMED International; 2000.
3. Southwick K. SNOMED, Read codes Unit in Blockbuster alliance, Combine sets up unrivaled leadership. Paper presented at: CAP Today, 1999; Northfield, IL.
4. Cawsey A. The essence of artificial intelligence. London ; New York: Prentice Hall; 1997.
5. College of American Pathologists. Concept modeling style/authoring guide: SNOMED International; 2002.
6. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. J Am Med Inf Assoc 1994;1(3):218-32.
7. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp 1997:640-4.
8. Sowa JF. Knowledge representation: logical, philosophical, and computational foundations. Pacific Grove: Brooks/Cole; 2000.
9. Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scalable methodologies for distributed development of logic- based convergent medical terminology. Methods Inf Med 1998;37(4-5):426-39.
10. Munsen MA, Shahar Y, Shortliffe EH. Clinical Decision support systems. In: Shortliffe EH, editor. Medical informatics. 2nd ed. New York: Sringer-Verlag; 2001. p. 573-609.
11. Andreassen S, Riekehr C, Kristensen B, Schonheyder HC, Leibovici L. Using probabilistic and decision-theoretic methods in treatment and prognosis modeling. Artif Intell Med 1999;15(2):121-34.
12. Sox HC. Medical decision making. Boston: Butterworths; 1988.
13. Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
14. Ward CD. Meditations on uncertainty. In: Kohout L, Bandler W, editors. Knowledge representation in medicine and clinical behavioral science. Cambridge, Mass.: Abacus Press; 1986. p. 19-25.

15. Baxt WG. Complexity, chaos and human physiology: the justification for non-linear neural computational analysis. Caner Lett 1994;77:85-93.
16. Berner ES. Clinical decision support systems: theory and practice. New York: Springer; 1999.
17. Hamilton PW, Montironi R, Abmayr W, Bibbo M, Anderson N, Thompson D, et al. Clinical applications of Bayesian belief networks in pathology. Pathologica 1995;87(3):237-45.
18. Heckerman D. Bayesian networks for data mining. Data Mining and Knowledge Discovery 1997;1(1):79-119.
19. Parmigiani G. Modeling in medical decision making - a Bayesian approach. West Sussex, England: John Wiley & Sons, Ltd; 2002.
20. Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.
21. Bartels PH, Thompson D, Bibbo M, Weber JE. Bayesian belief networks in quantitative histopathology. Anal Quant Cytol Histol 1992;14(6):459-73.
22. Witten I, Frank E. Data mining - practical machine learning tools and techniques with java implementations. San Francisco, CA: Morgan Kaufman; 2000.
23. Geiger D, Verma T, Pearl J. Identifying independence in Bayesian networks. Networks 1990;20:507-34.

SNOMED Representation of Explanatory Knowledge In Veterinary Clinical Pathology

by
Kurt Lee Zimmerman

Virginia Polytechnic Institute & State University
Department of Biomedical Sciences and Pathobiology
540.231.3692
kzimmerm@vt.edu

Keywords: SNOMED, Expert Systems, Controlled Terminology, Knowledge Representation, Clinical Pathology

Abstract

Work in the area of medical informatics has established standards for the expression of medical concepts, with numerous benefits to medicine. Increasing capacity for the sharing of medical information, establishing a common reference point of understanding, improving aggregation and analysis of medical data, and allowing for the orderly expansion of the terminology content are just a few such benefits to electronic medical records (EMR). SNOMED is an example of these standardization efforts, being recognized as a highly expressive terminology in its capacity for representing clinical concepts within EMR. While medical expert systems often store medical concepts as facts within their knowledge base, there has been little examination of the capacity for controlled terminologies to represent such concepts. Benefits similar to those associated with use of controlled terminology in EMR may occur with similar efforts to standardize the types of explanatory concepts used in the knowledge bases of medical expert systems.

This experiment looked at concepts relevant to veterinary clinical pathology, isolated from three different types of documents, each containing concepts of a clinical type and those used in explaining laboratory results. No significant difference was found in the ability of SNOMED to represent concepts from the three different documents (coverage: Finding document 45.4%, Discussion document 47.5%, and Textbook document 48.3%). SNOMED's overall coverage was low (<50%) for all three text sources with a majority of the problem mappings located in the Findings, Qualifier value, Observable entity, and Attribute hierarchies. While a portion (31-38%) of the poorly represented concepts was found to be due to missing content, a larger portion (48-54%) of the problematic concepts was due to a lack of granularity within an appropriate *IsA* hierarchy.

This work highlights limitations in SNOMED's ability to provide content to represent laboratory findings and associated explanations. This missing information is confined to a relative small area of the terminology. Deficiencies in this area must be resolved in order to develop a standard for the representation of clinical explanations based upon SNOMED's content and existing architecture.

Table of Contents

List of Figures

List of Tables

Introduction

The development of standards within medicine for the exchange, analysis, representation, abstraction, aggregation, and retrieval of medical information has been a very active area within the field of medical informatics.[1-13] The most recent release of the representation terminology, Systemized Nomenclature of Medicine – Clinical Terminology (SNOMED-CT) represents one important facet of these standardization efforts.[14-17] As with other representational terminologies, SNOMED-CT is designed to express the full range of clinical information found within an electronic medical record.

SNOMED-CT is an amalgamation of two of the larger representational terminologies, Read Codes Clinical Terms Version 3 and SNOMED-RT (SNOMED-Reference Terminology).[15, 18-20] Arranged in a polyhierarchy structure, using interrelationships among concepts to provide formal concept definitions, SNOMED meets many of the desiderata for a controlled medical terminology.[17, 21-23] SNOMED-CT currently contains approximately 350,000 concepts that are located in eighteen hierarchies headed by the following top-level concepts: Attribute, Body structure, Context-dependent category, Disease, Environment, Events, Findings, Observable entity, Organism, Pharmaceutical / biologic product, Physical force, Physical object, Procedure, Qualifier value, Social context, Specimen, Staging and scales, or Substance.[17, 23, 24] Concepts (objects) are defined by their associated collection of roles (relationship/attribute) and associated set of value concepts (values), which in and of themselves must be true. With approximately 1,500,000 sets of these object-attribute-value triples (OAV), SNOMED is very expressive in its ability to relate concepts to one another. The *IsA* attribute is a part of the definition for all concepts and plays a key role in establishing the hierarchy structures associated with the eighteen SNOMED top-level concepts.[17, 23, 24] Many of the other SNOMED attribute concepts, when used in OAV triples, create interhierarchy relationships.

The knowledge of SNOMED is represented by its collection of OAV triples stored within its database Tables.[23, 25-31] The creation and organization of this knowledge is guided by the use of description logic.[26, 32-35] With existing OAV triples representing expressions of logic statements, ontology modeling applications that are capable of performing logic calculations can compute concept equivalency for new content.[32, 35, 36] This approach assures a logically sound and self-consistent representation of the terminology by automating the subsumption of new children concepts within appropriate *IsA* hierarchies. This approach provides a powerful mechanism for the orderly expansion of the terminology's content.

Representation terminologies are not the only types of medical applications that utilize facts stored as OAV triples within a knowledge base. This type of knowledge representation is also commonly used within the knowledge bases of medical expert systems.[8, 30, 31, 37-45] Expert systems use this knowledge to assist their users in a variety of ways, ranging from helping to organize large volumes of medical information, to

bringing professional attention to significant findings, to generating specific recommendations for actions.[42, 43, 46] All of these tasks are made possible by the interaction of these systems with the facts stored within their knowledge bases. These facts provide the explanation and justification necessary for the alerts or suggested courses of action.[43] These facts are in and of themselves medical concepts with associated relationships (implying some influential or causal relationship) and values that provide explanatory medical knowledge.

While researchers in medical informatics have devoted great energies to creating standards for the representation of information within medical records (SNOMED), little has been done to address the standardization needs of the medical facts used within the knowledge bases of medical expert systems. This experiment investigates the capacity of an existing representation terminology, SNOMED-CT, to express in a standardized manner the type of explanatory knowledge that may be used within the knowledge base of an expert system. To gauge SNOMED's capacity to express explanatory, as opposed to clinical knowledge, the experiment compared SNOMED's coverage of explanatory concepts found in a text document rich in such knowledge to its coverage of clinical concepts in another document source abounding in those types of concepts, which are traditionally well represented by SNOMED. Other authors have documented SNOMED's ability to express concepts of the clinical type.[47-57] This comparison will help shed light on the question-- can SNOMED adequately express such knowledge--and if not, what additional content might be required?

Methods

Preparation of Three Types of Text Documents

Medical concepts designated as Finding concepts for this experiment are those used to convey medical results, measurements, or observations pertaining to a target population. Explanatory concepts convey influence and causality for Findings. Three text sources were selected on the basis of presumed differences in proportional use of these two general categories of concepts.

One electronic text source was prepared from a liver enzymology chapter of a well-respected veterinary clinical pathology textbook[xxxiv]. Two additional text sources were produced by experts in the line of work (application-derived). Application-derived text material consisted of pathology reports prepared from fourteen-day investigational studies of pharmaceutical compounds. These reports were prepared by a group of industry pathologists using clinical observations, serum chemistry, hematology, and necropsy results from Fisher 344 female rodents, 30-60 days of age. Nine pathology reports were selected in which at least one study animal had a clinically significant excursion of measured alanine aminotransferase activity. The pathology reports

[xxxiv] Duncan JR, Prasse KW, Mahaffey EA. Veterinary laboratory medicine. 3rd ed. Ames: Iowa State University Press; 1994

contained sections labeled respectively as Findings and Discussions. These sections were collated into single Findings and single Discussion documents which provided the remaining two text sources.

All formatting and all special characters were removed from the body of each document. One page of textual content was selected from each document type. In the case of the Findings and Discussion text, this consisted of the first full page of material (the length of the complied document was less than two pages for both of these text sources). For the textbook document, the total number of pages was determined and one page was selected at random for use in the experiment.

Concept Isolation

Sentences from the documents were parsed into individual words and short phrases representing single strings via use of a barrier words algorithm and a purpose-built application[xxxv].[58-60] The application parsed each sentence into individual lexical strings that were presented to the user for final review prior to approval. Isolated strings, individual sentences and source document were stored in a database[xxxvi] that maintained links from isolated concepts to originating sentences and from sentences to originating text sources. A list of lexically unique strings from each document was created. Number and percentage of lexically similar strings within the three document sources were determined.

Concept Mapping to SNOMED

The isolated strings were manually compared to SNOMED-CT[xxxvii] and were reviewed independently for equivalency of semantic meaning by the author and another SNOMED terminology expert. Using two different SNOMED browsers, searches were conducted for concepts within SNOMED.[61-63] Mapping results included the following information: text document concept, quality and bias scores, and the fully specified name of the mapped SNOMED concept. In cases where no SNOMED match was found, a SNOMED top-level concept (Attribute, Body structure, Context-dependent category, Disease, Environment, Events, Findings, Observable, Organism, Pharmaceutical / biologic product, Physical force, Physical object, Procedure, Qualifier value, Social context, Specimen, Staging and scales, or Substance was assigned.[17, 23]

A quality score of 1 was used when no similar SNOMED concept was found. A quality score of 2 was used when a similar concept was found, but the "similar" concept could not be considered to mean the same thing as the text concept. In some cases, lexically similar SNOMED concepts were located in hierarchies that implied a difference in semantic meaning (horizontally different). In other cases, the SNOMED concept was

[xxxv] Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000
[xxxvi] Microsoft. Microsoft access 2000. 9.0.2720 ed: Microsoft Corporation; 1999
[xxxvii] College of American Pathologists. SNOMED clinical terms. 2002-01Q ed: SNOMED International; 2002

in a semantically correct hierarchy, but it was either less or more granular in meaning. A quality score of 3 was assigned when the SNOMED concept meant the same thing as the text concept. Mapping scores for each quality score category for each of three documents were tallied for each reviewer. Percentages of mapping scores categories for each document were calculated for both reviewers. Mapping scores category percentages were calculated by dividing the individual reviewer's tallied number of scores in a quality score category by the total number of unique strings in the document source.

A bias category score was assigned to those concepts with a quality score of 2. A bias score of 1 was assigned when the SNOMED concept was lexically similar but semantically distinct. The bias score 2 was assigned when the mapped SNOMED concept was in the appropriate *IsA* hierarchy but was less granular than the document concept. A bias score of 3 was assigned when the SNOMED concept was in the appropriate *IsA* hierarchy but was more granular than the document concept. Bias scores for each bias score category and for each of three documents were tallied from each reviewer. Percentages of Bias scores categories for each document were calculated for both reviewers. Bias scores category percentages were calculated by dividing each individual reviewer's tallied number of scores in a bias category by the number of unique strings classified as having a mapping quality score of 2.

Comparison of Mapping Results

A commercial statistical software package[xxxviii] on a personal computer was used for the statistical comparison of scores between the various document types. Reviewer agreement was assessed by tabulating instances of agreement in assignment of mapping quality scores for each of the three document sources. Agreement percentage on quality scores was determined for each document source by summing the instances of agreement then dividing by the total number of unique strings in the document. In addition, a Kappa value was calculated for each of the quality score categories in each of the documents by using the individual reviewer-assigned quality scores for each unique string of the document. Lastly, agreement on the selection of SNOMED concepts was assessed. Instances where both reviewers assigned the SNOMED concept and then gave the match the same quality score were tallied and summarized for quality score categories 2 and 3. Percentages were determined by dividing the tallied number of identical concept mappings by the total number of unique strings where reviewers agreed on the quality score.

A minimally acceptable Kappa value of 0.60 was established for each quality score category in each document.[64-67] An alternative quality scoring system was created for use in the event of Kappa failure. If Kappa failure occurred in quality score categories 1 and/or 2, mapping of these categories were to be rescored as a quality score of 1 (classified as a Problem mapping). Mapping with a quality score of 3 were to be rescored as a quality score of 2 (classified as a Good mapping). Bias scores for partially

[xxxviii] Minitab. Minitab statistical software. 13.32 ed: Minitab Inc; 2002

mapped concepts would be retained. If Kappa failure occurred in quality score categories 3, mappings of categories 2 and 3 were to be rescored as a quality score of 2 (classified as a Possible mapping). Mappings with a quality score of 1 were to retain their prior quality score of 1 (Bad mapping). Following remapping of quality scores, Kappa values were to be reassessed. If the minimal acceptable Kappa value still had not been reached in the remaining two quality score categories, a third reviewer was to be recruited for arbitration on quality score assignments.

The average number of instances and percentages for each quality score category and for each document source was determined following establishment of acceptable Kappa agreement between reviewers. The tallied quality score instances and prior calculated percentages from both reviewers were individually summed and divided by 2 for this determination. The averaged bias score and percentage were similarly calculated. A test of Independence was conducted at an alpha value of 0.05.[67] The Chi Square Test used the averaged number of instances of quality scores to test for an association between the document source and the frequency distribution of quality score assignments.

Following the agreement test, the averaged frequency distribution for the SNOMED top-level concepts was calculated. A frequency distribution for mapped top-level concept was prepared for each reviewer. The distributions for both reviewers were summed and divided by 2 to produce an averaged distribution for Good and Problem mappings.

Results

Preparation of Text Documents

The findings and discussion documents each consisted of slightly more than one page of material. The textbook document was thirteen pages in length and page eight was randomly selected for use. Examples of strings isolated from each document are shown in Figure 1**.** Word processing indices associated with each of the one-page documents are shown in Table 1. The lexical similarity of isolated strings (range from 4-22%) in the three documents is shown in Table 2**.**

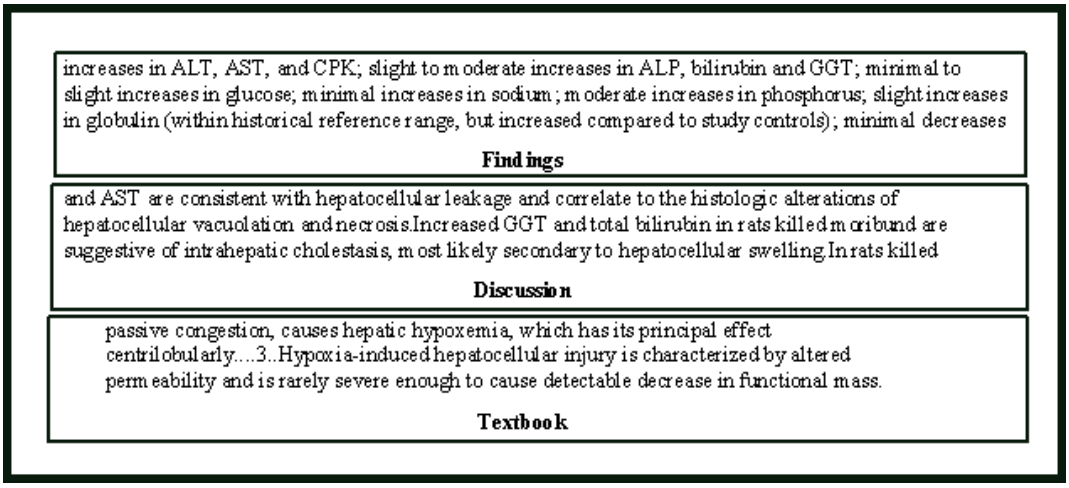


Figure 1 Samples from the three documents after formatting was removed

Table 1 Indices associated with the isolated strings from the three document sources

Document	Number of Sentences	Number of Words	Number of Strings	Number of Unique Strings (Num/Sen, % Strings)
Pathology Report Finding	37	664	292	133 (3.6, 45%)
Pathology Report Discussion	48	813	289	183 (3.8, 63.3%)
Textbook: Liver Enzymology	49	694	240	176 (3.6, 73.3%)

Table 2 Comparison of lexically similar strings across the three text sources

Matrix of Lexically Matching Strings		
	Pathology Report Discussion (183)	Textbook (176)
Pathology Report Finding (133)	41 (22%)[xxxix]	14 (8%)[xl]
Pathology Report Discussion (183)		7 (4%)[xli]

[xxxix] % Based on number of discussion strings
[xl] % Based on number of textbook strings
[xli] % Based on number of discussion strings

Mapping to SNOMED

Tallied summary of instances for each quality and bias category in each document and for each reviewer is shown in Table 3. Agreements[xlii] between reviewers on instances of similarly assigned quality scores were 72.7% for Discussion, 78.2% for Findings, and 79.5% for the Textbook document. Agreement percentage on the selected SNOMED concepts was highest for quality category 3 (average 95.2%), followed by category 2 quality scores (average 84.4%). The acceptable cutoff Kappa value for reviewer agreement (0.60) was not reached for quality score category 1 in the Findings document or for category 1 and 2 in the Discussion document (Table 3). Quality scores for categories 1 and 2 were combined as outlined in the methods section. Quality and bias score tallies along with averages following the restructuring of quality score categories are shown in Table 3.

[xlii] Using the 3 category quality schema

Table 3 Tallied quality and bias scores from both reviewers with agreement and averaged statistical values

Document	Num Unique Strings	Quality Score Schema	Quality Score Cat	Bias Score Cat	Reviewer 1				Reviewer 2				Agreement					Averaged			
					Quality Score	Quality Score %	Bias Score	Bias Score %	Quality Score	Quality Score %	Bias Score	Bias Score %	Ag On Quality Score	Ag % On Quality Score	Ag Kap On Quality Score	Ag On Quality Score & SNOMED	Ag % On Quality Score & SNOMED	Quality Score	Quality Score %	Bias Score	Bias Score %
Pathology Report Finding	133	3 Cat	1 (Bad)		21	15.8			19	14.3			13		0.5881						
			2 (Problem)	1	53	39.9	17	32.1	52	39.1	16	30.8	40	78.2	0.6066	33	82.5			17	31.4
				2			28	52.8			29	55.8								29	54.3
				3			8	15.1			7	13.5								8	14.3
			3 (Good)		59	44.4			62	46.6			51		0.7119	48	94.1				
		2 Cat	1 (Problem)		74	55.6			71	53.4			63	85.7	0.7119			73	54.5		
			2 (Good)		59	44.4			62	46.6			51					61	45.5		
Pathology Report Discussion	183	3 Cat	1 (Bad)		30	16.4			15	8.2			12		0.4679						
			2 (Problem)	1	64	35.0	26	40.6	83	45.4	29	34.9	51	72.7	0.4884	44	86.3			28	37.8
				2			32	50.0			44	53.0								38	51.5
				3			6	9.4			10	12.1								8	10.7
			3 (Good)		89	48.6			85	46.5			70		0.6275	67	95.7				
		2 Cat	1 (Problem)		94	51.4			98	53.6			79	81.4	0.6275			96	52.5		
			2 (Good)		89	48.6			85	46.5			70					87	47.5		
Textbook: Liver Enzymology	176	3 Cat	1 (Bad)		31	17.6			20	11.4			18		0.656						
			2 (Problem)	1	54	30.7	22	40.7	77	43.8	28	36.4	51	79.5	0.6474	43	84.3			25	38.6
				2			27	50.0			37	48.1								32	49.0
				3			5	9.3			12	15.6								9	12.4
			3 (Good)		91	51.7			79	44.9			71		0.6814	68	95.8				
		2 Cat	1 (Problem)		85	48.3			97	55.1			77	84.1	0.6814			91	51.7		
			2 (Good)		91	51.7			79	44.9			71					85	48.3		

Using the two category quality scoring schema, agreement between reviewers was greater than 80% and Kappa value was greater than 0.60 for assigned quality scores in each of the three documents (Table 3). The averaged mapping quality[xliii] and bias[xliv] scores are shown in Table 2. The percentage of mappings judged as being Good ranged from a low of 45% with Findings to the high of 48% for the textbook document (

Figure 2, Table 3). The most common bias score for all three document types was 2 (SNOMED concepts less granular than necessary), followed in frequency by bias score 1 (SNOMED concept horizontal semantically difference), and least, bias score 3 (SNOMED concept too granular) (Figure 3, Table 3). Using the averaged results, the Chi Square test for non-association (alpha 0.05) showed that there was not enough evidence to reject the null hypothesis suggesting that there was no association between the document source and frequency distribution of quality scores (Table 4).

[xliii] Using the 2 category quality schema
[xliv] Using the 3 category quality schema

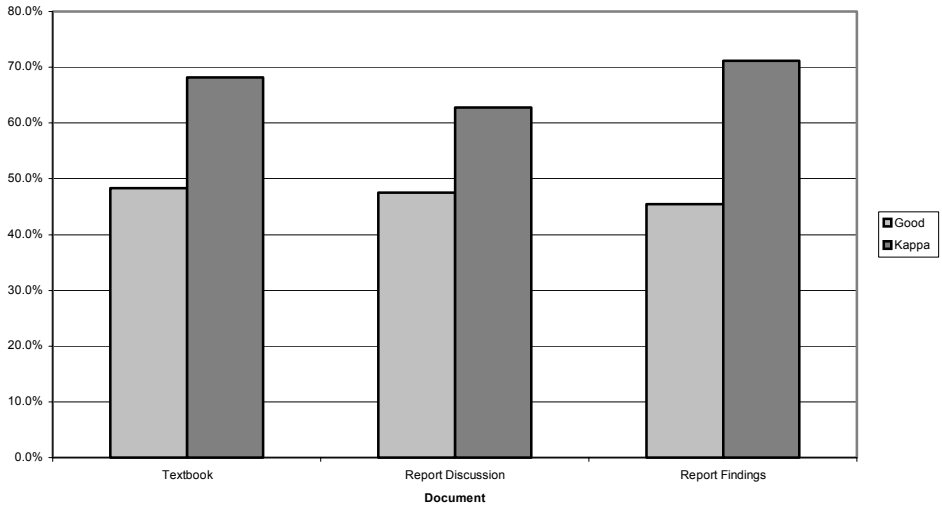


Figure 2 Good quality mapping scores percentages and Kappa by document type[xlv]

[xlv] Using the 2 category quality schema

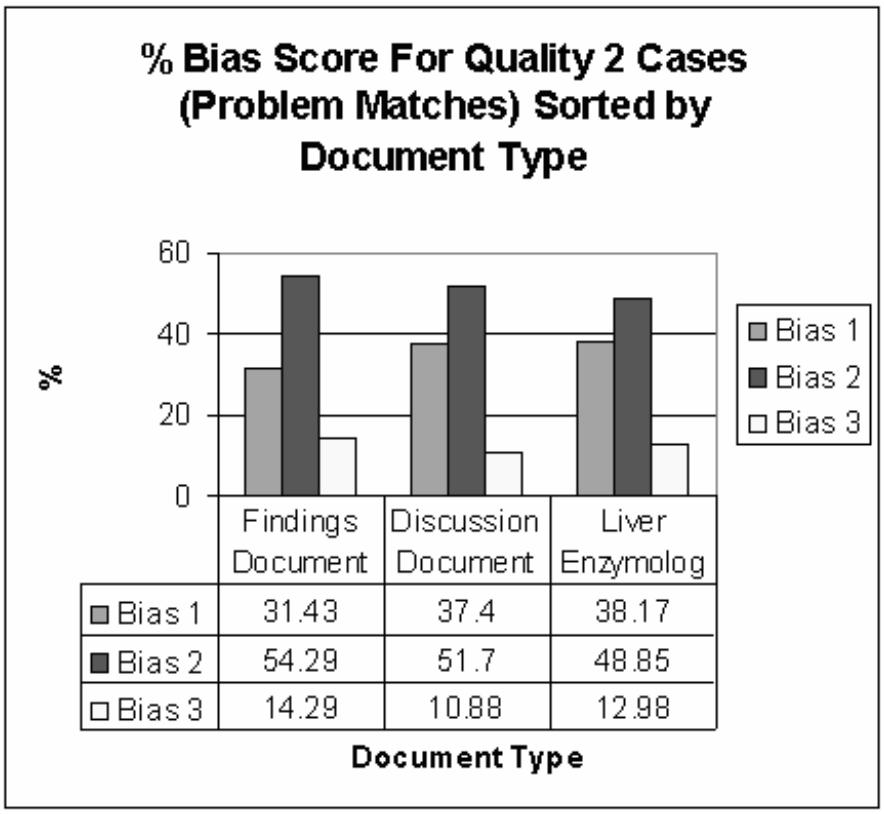


Figure 3 Bias score for quality Problem SNOMED matches by document type[xlvi]

[xlvi] Using the 3 category quality schema

Table 4 Chi Square test of association between quality score frequency distributions and document source[xlvii]

Quality Score		Finding	Discussion	Textbook	Total
1	Actual	72	96	91	260
	Expected	70.23	96.71	93.01	
2	Actual	61	87	85	232
	Expected	62.72	86.29	82.99	
Total		133	183	176	492
Chi-Sq	0.325				
DF	2				
P-Value	0.85				

[xlvii] Using averaged quality score data and the 2 category quality score schema

The names and frequency of assignment of top-level SNOMED concepts for quality score categories of "Good" and "Problem" and for document sources are shown in Figure 4 and Figure 5. In both quality score categories, Finding (concepts that result from an assessment or judgment) and Qualifier value (concepts that serve as values for attributes) represented over 50% of the assigned top-level concepts. For all three document types in the quality score category of "Good", Finding, Qualifier value, Body structure (normal anatomical concepts), and Substance (concepts that represent biologic substances, body substances, diagnostic substances, and dietary substances) represented over 75% of top-level SNOMED concepts. The top-level SNOMED concepts - Finding, Qualifier value, Observable entity (concepts that represent potential findings), and Attribute (concepts that are used to relate two different concepts) comprised 75% of quality score "Problem."

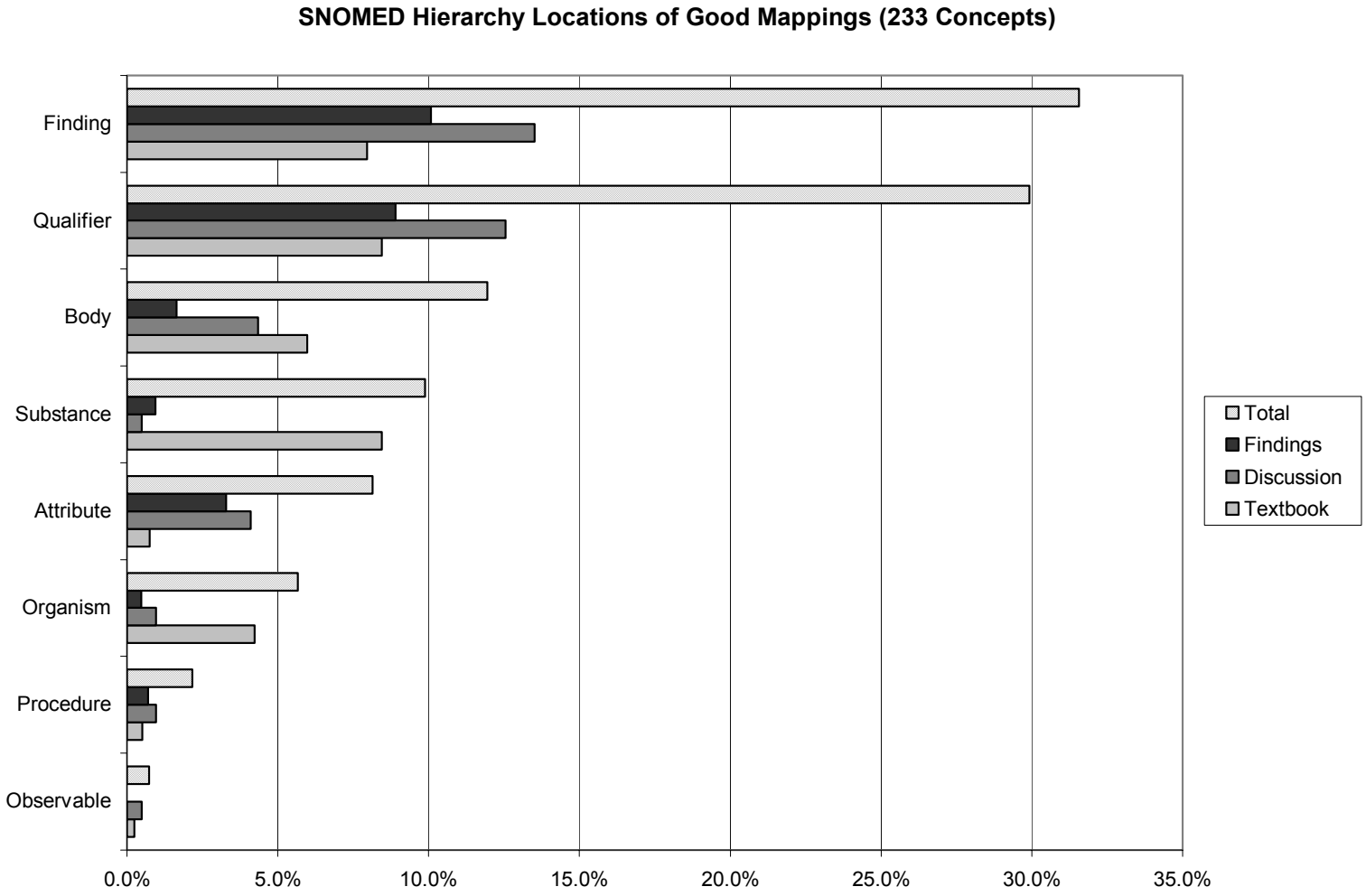


Figure 4 Frequency distribution of Good mapping to SNOMED hierarchies and document source

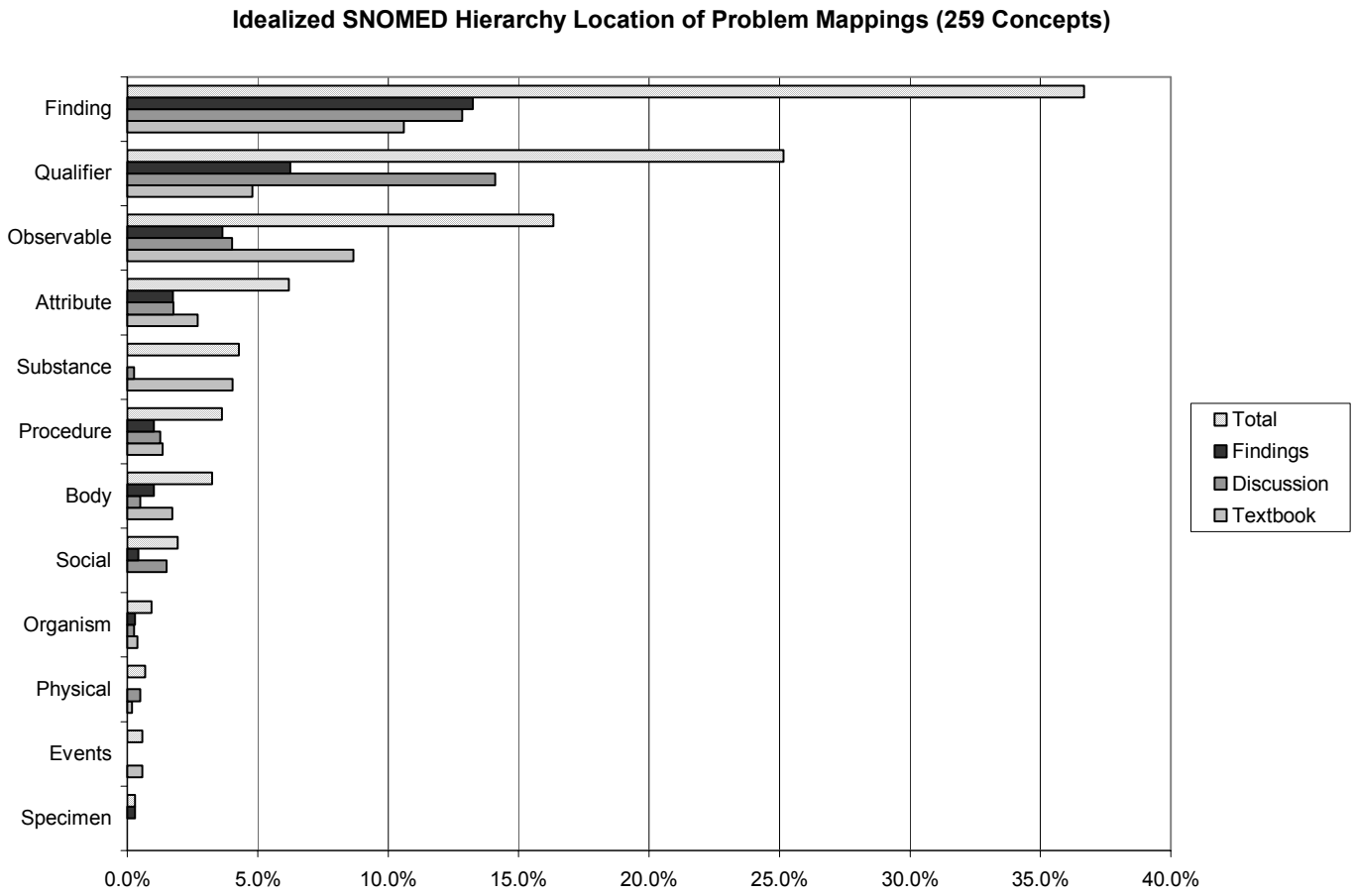


Figure 5 Frequency distribution of Problem mapping to SNOMED hierarchies and document source

Discussion

SNOMED has been shown to be quite capable of representing clinical concepts in numerous evaluations.[48-50, 55, 68-72] However, similar evaluations of the ability of SNOMED to represent concepts of a less clinical nature dealing with clinical pathology and explanations of laboratory findings have been preformed. There was no significant difference in the ability of SNOMED to represent concepts from text sources selected for this study (coverage: Finding document 45.4%, Discussion document 47.5%, and Textbook document 48.3%). SNOMED coverage was considered low overall (<50% for all three text sources), with a deficiency noted most heavily within the area of Observable Entity. While a portion of the problematic concepts was simply due to poor representation (31-38%[xlviii]), a larger portion resulted from a lack of granularity within an appropriate *IsA* hierarchy (48-54%[xlix]). As a whole, this work suggests that SNOMED's ability to meet the vocabulary needs of explanatory knowledge, as related to clinical pathology, is similar to its coverage of clinical finding concepts in that domain. However, SNOMED still lacks much of the necessary content for both of these purposes.

It is interesting that the coverage of content from these three text sources was found to be as low as it was (45.5-48.3%). These results are in contrast to the coverage found by other reviewers, which ranged from 40-71%.[49, 54, 69, 72, 73] Several reasons may account for the lower results obtained in this study. First, the content area of interest was veterinary clinical pathology. SNOMED does explicitly include veterinary concepts but its content is largely driven by human medicine. Concepts gleaned from the selected documents may over-represent concepts that are exclusive to veterinary medicine. Concepts were matched lexically and semantically by each reviewer to individual concepts within the terminology; there was no allowance for concept phrase construction (combining existing concepts to provide greater specificity) which could have dramatically improved the coverage results. The parsing and concept selection methodology used in this study may not be directly comparable to the other studies. This could be affected during both concept isolation and mapping to SNOMED and could reflect on either available tools or the investigators' abilities and biases. Both reviewers found similar low overall scores of coverage, but this could reflect the training of one reviewer by the other. Regardless of the cause for the observed low coverage score, the results still serve as an internal control by which the three text sources can be compared to one another.

[xlviii] Using the 3 category quality schema; % represents the proportion of Quality 2 concepts with a Bias score 1; quality score 1 combined with quality score 2 bias score 1 concepts represent ~ 28% of all concepts

[xlix] Using the 3 category quality schema; % represents the proportion of Quality 2 concepts with a Bias score 2

This experiment considered all concepts isolated from three text sources which were assumed to contain different types of content. This supposition was supported by the variation in percentages of unique concepts in the three (unique concepts: Finding document 45%, Discussion document 63%, and Textbook document 73%). On a lexical basis, at least, the documents also appeared to be quite different, as shown in Table 2. While this alone does not prove that these explanatory sources contain concepts that differ in their intended usage, it does suggest increased complexity of relationships between the expressed concepts. Such increased complexity would be expected in a more analytic type of text that expresses both the interpretation of simple observable findings and the justification for such interpretations.[74] Also, it was speculated that the pathology report discussion document would lie somewhere between the report findings document and the textbook document in terms of the proportion of interpretive concepts in its content, and the concept density results are consistent with this presumption.

The original reason for assuming that the documents were different was based on a perception of the intent of the documents. In a typical pathology findings report, the author's intent is to report a list of facts (increased ALT, died, decreased liver weight, etc) recorded for the study. These concepts similar to those for which previous assessments of SNOMED exist.[49] These assessments have recognized SNOMED's proficiency in representing such clinical findings. Because SNOMED's coverage of clinical concepts has traditionally been considered good, concepts isolated from the pathology report findings were expected to provide a standard against which SNOMED's coverage of the more explanatory concepts from the other two text sources could be judged. The authors of the pathology discussion reports and the liver chapter wrote with a different intent than the authors of the pathology report findings text. These two text sources (referred to as explanatory text for the remainder of this discussion) aim not simply to report facts, but to explain or interpret them; not only to document the "what" of observable facts but to answer the "why" of their causality and significance. This answer to *why* in the realm of clinical pathology is often due to a cascading series of physiologic events initiated by some trigger.[75-77] Such events are commonly expressed as measured alterations of physiologic functions. However, based on the percentage of concepts well and poorly represented by SNOMED for each document as a whole (Table 3), and the fact that there seemed to be considerable overlap in the SNOMED's hierarchy location of concepts classified as good and bad, it cannot conclude that the document types are, in fact, different (Figure 4, and Figure 5).

The distribution of quality scores was similar for the three text sources. In the second analysis (two category system), agreement between the reviewers' assigned quality scores for all concepts from the three sources was generally good (72-79), in agreement with assessment variability previously noted between experts.[78-80] However, there was a noticeable problem with the kappa statistic between the reviewers concerning quality score categories 1 and 2 for the pathology report discussion document. Review of these two quality score categories for the other two text documents showed that the kappa statistic was lower for these text sources also.

To explore the source of this disagreement, a comparison of SNOMED concept selection within each quality category was made (Table 3). From this work it can be seen that both reviewers selected the same SNOMED concept in a high percentage of the cases (>80%). This high agreement in SNOMED concept selection suggests that the disagreement between reviewers more likely arises from differing quality scores. The similarity in mapping between the reviewers was not surprising in that both approached the mapping challenge with a common background of knowledge as develops in a mentor-graduate-student relationship. The disagreement on the selected quality score for the similarly mapped SNOMED concepts reflects differing interpretations as to what constitutes a problem concept versus a bad concept.

In the initial analysis employing the three-tiered quality scoring scheme, assignment of quality scores to certain concepts was difficult to implement consistently. For example, the locating concept "alanine aminotransferase" (in association with hepatocellular necrosis) in the substance hierarchy led one reviewer to assign a quality score of 2 with a bias of 1 (indicating a mapping concept located but with the bias of being in the wrong hierarchy) while the other assigned a quality of 1 (indicating no potential mapping concept located). Therefore, the distinction between a bad and problem concept was simply removed. With this simplification of the quality-scoring scheme, overall agreement and Kappa statistic exceeded 80% and 0.60 respectively for all quality score categories with all three text sources (Table 3).

Review of the SNOMED concepts which scored as being good matches shows that over 85% resided within the following five hierarchies: Finding, Qualifier Value, Body Structure, Substance, and Attribute. This suggests that most of the concept content necessary to represent laboratory findings and associated explanations (for clinical pathology) resides in a relatively narrow slice of the terminology.

Over 85% of the SNOMED mappings labeled as Problem in this experiment reside within the following hierarchies: Finding, Qualifier Value, Observable Entity, Attribute, and Substance. These are the same hierarchies identified as having most of the good matches, with the exception of Observable Entity. This similarity in the hierarchy locations for good and bad matches further strengthens the supposition that necessary content for explanatory knowledge occupies a comparatively narrow domain in the terminology. The exception, Observable Entity, deserves special mention. The absence of this hierarchy from the top 5 good mapping hierarchies suggests that this is the weakest content area of the hierarchies mentioned. Content missing from this hierarchy comprises many physiologic function concepts.

Examining the bias score associated with the problem quality score concepts helps to delimit the nature of the deficiencies in question. The majority of these problem concepts are simply not refined enough (that is, are insufficiently granular) within the *IsA* hierarchies (bias score 2). This suggests that SNOMED has logical locations for the

"missing" content. As expressed by Cimino, content is everything, and SNOMED has sought to provide necessary content to the whole of medicine.[22] This is an incredible challenge and one at which SNOMED has been judged successful by previous reviewers.[48-50, 55, 68-72] Refined content within the *IsA* hierarchies is likely added as users expert in a particular domain notice deficiencies and bring them to the attention of SNOMED. SNOMED's existing review and modeling processes can be expected to address some of the identified content deficiencies noted from this work. Figure 3 also shows that closely following lack of concept granularity, lexically similar but semantically distinct concepts compose a large portion of the problem concepts (bias score 1).

Conclusion

This work demonstrates SNOMED's capacity to represent concepts in veterinary clinical pathology. This work also highlights limitations in SNOMED's ability to provide content to represent clinical pathology laboratory findings and associated explanations. This missing information is confined to a relative small area of the terminology, indicating a fruitful area on which to focus future efforts to develop and refine SNOMED. The Observable entity hierarchy is particularly deficient in concepts used by veterinary clinical pathologists to describe their field or conduct their research. This class of concepts was frequently used in the selected text sources to explain why clinical finding events occurred. Deficiencies in this area must be resolved in order to develop a standard for the representation of clinical explanations based upon SNOMED's content and existing architectural.

Acknowledgment

This work has been funded by the National Institutes of Health, National Library of Medicine Informatics Research Fellowship Award, grant number 1 F37 LM07699.

Bibliography

1. Zanstra PE, Rector AL, Ceusters W, de Vries Robbe PF. Coding systems and classifications in healthcare: the link to the record. Int J Med Inf 1998;48(1-3):103-9.
2. Smith RD, Williams M. Applications of informatics in veterinary medicine. Bull Med Libr Assoc 2000;88(1):49-55.
3. Lussier YA, Côté RA. Are we closer to standard medical coding? Paper presented at: Canadian Medical Informatics, 1995.
4. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. J Am Med Inf Assoc 1997;4(3):213-21.
5. Chute CG. The Copernican era of healthcare terminology: a re-centering of health information systems. Proc AMIA Symp 1998:68-73.
6. Côté RA. The development of a systemized nomenclature of medicine. Proceedings of Journees d'Informatique Medicale. Toulouse, France; 1973.

7. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. Methods Inf Med 1991;30(3):179-86.
8. Bemmel JHv, Musen MA, editors. Handbook of medical Informatics. Houten, the Netherlands: Springer-Verlag; 1997.
9. Gantner GE. History of coding. Pathologist 1977;XXXI(7, July).
10. Cimino JJ. Review paper: Coding systems in health care. Methods Inf Med 1996;35(4-5):273-84.
11. Hammond EW, Cimino JJ. Standards in medical informatics. In: Shortliffe EH, editor. Medical informatics. 2nd ed. New York: Sringer-Verlag; 2001. p. 212-56.
12. Ackerman MJ, Ball MJ, Clayton PD, Frisse ME, Gardner RM, Greenes RA, et al. Standards for medical identifiers, codes, and messages needed to create an efficient computer-stored medical record. J Am Med Inf Assoc 1994;1(1):1-7.
13. Blair JS. An overview of health care information. Proc AMIA Symp 1995:201-12.
14. College of American Pathologists. SNOMED clinical terms. 2002-01Q ed: SNOMED International; 2002.
15. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp 2001:662-6.
16. College of American Pathologists. SNOMED RT: SNOMED International; 2001.
17. College of American Pathologists. SNOMED CT product profile: SNOMED International; 2003.
18. Southwick K. SNOMED, Read codes Unit in Blockbuster alliance, Combine sets up unrivaled leadership. Paper presented at: CAP Today, 1999; Northfield, IL.
19. NHS. SNOMED clinical terms - a global standard for health care terminology. Paper presented at: NHS Information Authority Coding and Classification Help Desk, 1999; Loughborough, Leicestershire, UK.
20. Nash S, Sable JH, Stearns MQ, Robinson D. Merging domain hierarchies in the development of SNOMED CT. Proc AMIA Symp 2001:984.
21. Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. Proc AMIA Symp 1999:42-6.
22. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(Nov(4-5)):394-403.
23. College of American Pathologists. SNOMED clinical terms consultation document: core structure: SNOMED International; 2000.
24. College of American Pathologists. Concept modeling style/authoring guide: SNOMED International; 2002.
25. Spackman KA, Connelly DP. Knowledge-based systems in laboratory medicine and pathology - a review and survey of the field. Arch Pathol Lab Med 1987;111(2):116-19.
26. Sowa JF. Knowledge representation: logical, philosophical, and computational foundations. Pacific Grove: Brooks/Cole; 2000.
27. Rothwell DJ. SNOMED-based knowledge representation. Methods Inf Med 1995;34(1-2):209-13.

28. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Inf Assoc 1994;1(1):35-50.
29. Cimino JJ. Data-storage and knowledge representation for clinical workstations. Int J Bio-Med Comput 1994;34(1-4):185-94.
30. Bartels PH, Thompson D. Expert systems in histopathology. III. Representation of knowledge as "structured objects". Anal Quant Cytol Histol 1989;11(6):367-74.
31. Bartels PH, Hiessl H. Expert systems in histopathology. II. Knowledge representation and rule- based systems. Anal Quant Cytol Histol 1989;11(3):147-53.
32. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. J Am Med Inf Assoc 1994;1(3):218-32.
33. Campbell KE, Tuttle MS, Spackman KA. A lexically-suggested logical; metric for medical terminology maturity. Proc AMIA Symp 1998:785-9.
34. Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. J Am Med Inf Assoc 1998;5(2):203-13.
35. Spackman KA. Normal forms for description logic expressions of clinical concepts in SNOMED RT. Proc AMIA Symp 2001:627-31.
36. Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scalable methodologies for distributed development of logic- based convergent medical terminology. Methods Inf Med 1998;37(4-5):426-39.
37. Roach J, Lee S, Wilcke J, Ehrich M. An expert system for information on pharmacology and drug interactions. Comput Biol Med 1985;15(1):11-23.
38. Oberholzer M, Feichter G, Dalquen P, Ettlin R, Christen H, Buser M. A simple "expert system" for morphometric evaluation of cells in pleural effusions. Pathol Res Pract 1989;185(5):647-51.
39. Nathwani BN, Clarke K, Lincoln T, Berard C, Taylor C, Ng KC, et al. Evaluation of an expert system on lymph node pathology. Hum Pathol 1997;28(9):1097-110.
40. McNeely MDD. The use of expert systems for improving test use and enhancing the accuracy of diagnosis. Clin Lab Med 2002;22:515-28.
41. Baak JP, Kurver PH. Development and use of a rule-based pathology expert consultation system. Anal Quant Cytol Histol 1988;10(3):214-8.
42. Bartels PH, Weber JE. Expert systems in histopathology. I. Introduction and overview. Anal Quant Cytol Histol 1989;11(1):1-7.
43. Berner ES. Clinical decision support systems: theory and practice. New York: Springer; 1999.
44. Connelly DP, Bennett ST. Expert systems and the clinical laboratory information-system. Clin Lab Med 1991;11(1):135-51.
45. Diamond LW, Mishka VG, Seal AH, Nguyen DT. Are normative expert systems appropriate for diagnostic pathology? J Am Med Inform Assoc 1995;2(2):85-93.
46. Munsen MA, Shahar Y, Shortliffe EH. Clinical Decision support systems. In: Shortliffe EH, editor. Medical informatics. 2nd ed. New York: Sringer-Verlag; 2001. p. 573-609.

47. Berman JJ, Moore GW, Donnelly WH, Massey JK, Craig B. A SNOMED analysis of three years' accessioned cases (40,124) of a surgical pathology department: implications for pathology-based demographic studies. Proc Annu Symp Comput Appl Med Care 1994:188-92.
48. Campbell JR, Payne TH. A comparison of four schemes for codification of problem lists. Proc Annu Symp Comput Appl Med Care 1994:201-5.
49. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. J Am Med Inf Assoc 1997;4(3):238-51.
50. de Bruijn LM, Hasman A, Arends JW. Supporting the classification of pathology reports: comparing two information retrieval methods. Comput Methods Programs Biomed 2000;62(2):109-13.
51. Henry SB, Campbell KE, Holzemer WL. Representation of nursing terms for the description of patient problems using SNOMED III. Proc Annu Symp Comput Appl Med Care 1993:700-4.
52. Henry SB, Holzemer WL. Can SNOMED International represent patients' perceptions of health-related problems for the computer-based patient record? Proc Annu Symp Comput Appl Med Care 1994:184-7.
53. Henry SB, Holzemer WL, Reilly CA, Campbell KE. Terms used by nurses to describe patient problems: can SNOMED III represent nursing concepts in the patient record? J Am Med Inf Assoc 1994;1(1):61-74.
54. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inf Assoc 1997;4(6):484-500.
55. Lussier YA, Bourque M. Comparing SNOMED and ICPC retrieval accuracies using relational database models. Proc AMIA Annu Fall Symp 1997:514-8.
56. Sinha U, Yaghmai A, Dai B, Thompson L, Taira R, Dionisio J, et al. Evaluation of SNOMED3.5 in representing concepts in chest radiology reports: integration of a SNOMED mapper with a radiology reporting workstation. Proc AMIA Symp 2000;20(Suppl):799-803.
57. Mullins HC, Scanland PM, Collins D, Treece L, Petruzzi P, Jr., Goodson A, et al. The efficacy of SNOMED, Read Codes, and UMLS in coding ambulatory family practice clinical records. Proc AMIA Annu Fall Symp 1996:135-9.
58. Moore GW, Berman JJ. Automatic SNOMED coding. Proc Annu Symp Comput Appl Med Care 1994:225-9.
59. Moore GW, Miller RE, Hutchins GM. Indexing by mesh titles of natural-language pathology phrases identified on 1st encounter using the barrier word method. Mod Pathol 1988;1(1):A110-A10.
60. Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000.
61. Clinical Information Consultancy. CIC Look Up Engine (CLUE). 5.0.0047 ed: Clinical Information Consultancy; 2002.
62. Wilcke JR. SNOMED CT Browser. 2.0 ed. Blacksburg: Drug Information Laboratory, Virginia Maryland Regional College of Veterinary Medicine; 2003.

63. Wilcke J. SNOMED browser latest in informatics. J Am Vet Med Assoc 2000;216(7):1049.
64. Skov BG, Broholm H, Engel U, Franzmann MB, Nielsen AL, Lauritzen AF, et al. Comparison of the reproducibility of the WHO classifications of 1975 and 1994 of endometrial hyperplasia. Int J Gynecol Pathol 1997;16(1):33-7.
65. Firriolo FJ, Levy BA. Computer expert system for the histopathologic diagnosis of salivary gland neoplasms. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1996;82(2):179-86.
66. Baumann RP, Moret J. Weighted kappa statistics for measuring the divergence of rapid section and definitive diagnosis. Pathologe 1997;18(5):412-6.
67. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford; Boston: Blackwell Scientific Publications; 1994.
68. Ruggieri AP, Elkin P, Chute CG. Representation by standard terminologies of health status concepts contained in two health status assessment instruments used in rheumatic disease management. Proc AMIA Symp 2000;(20(Suppl)):734-8.
69. Klimczak JC, Hahn AW, Sievert ME, Petroski G, Hewett J. Comparing clinical vocabularies using coding system fidelity. Proc Annu Symp Comput Appl Med Care 1995:883-7.
70. Hall PA, Lemoine NR. Comparison of manual data coding errors in two hospitals. J Clin Pathol 1986;39(6):622-6.
71. Franz P, Zaiss A, Schulz S, Hahn U, Clar R. Automated coding of diagnoses - three methods compared. Proc AMIA Symp 2000;(20(Suppl)):250-4.
72. Chute CC, Cohn SP, Campbell JR, Campbell KE, Oliver DE. The content coverage of clinical classifications. Am J Epidemiol 1995;141(11):S82-S82.
73. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. J Am Med Inf Assoc 1996;3(3):224-33.
74. Bender H, Lockee B, Danielson J, Mills E, Boon G, Burton J, et al. Mechanism-based diagnostic reasoning: thoughts on teaching introductory clinical pathology. Vet Clin Pathol 2000;29:77-83.
75. Duncan JR, Prasse KW, Mahaffey EA. Veterinary laboratory medicine. 3rd ed. Ames: Iowa State University Press; 1994.
76. Cotran RS, Kumar V, Collins T, Robbins SL. Robbins pathologic basis of disease. 6th ed. Philadelphia: Saunders; 1999.
77. Stockham SL, Scott MA. Fundamentals of veterinary clinical pathology. 1st ed. Ames, Iowa: Iowa State Press; 2002.
78. Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
79. Sox HC. Medical decision making. Boston: Butterworths; 1988.
80. Morgan MG, Henrion M, Small M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge; New York: Cambridge University Press; 1990.

Normalized Medical Explanation (NORMEX) Syntax To Integrate Causal Reasoning With SNOMED

by
Kurt Lee Zimmerman

Virginia Polytechnic Institute & State University
Department of Biomedical Sciences and Pathobiology
540.231.3692
kzimmerm@vt.edu

Keywords: SNOMED, NORMEX, Clinical Pathology, Expert Systems, Controlled Terminology

Abstract

Controlled terminologies are part of ongoing endeavors to establish standards for the transmission, abstraction, and representation of information within electronic medical record (EMR) systems. SNOMED is a controlled medical terminology organized in a standard knowledge representation format. Medical expert systems can use a similar format to organize requisite knowledge. The fundamental difference between the two realms is the organizing principle that determines relationships between interconnected facts. Medical terminologies categorize concepts; expert system knowledge bases create causal connections between them. Medical expert systems must not only record facts, but also interpret them, explain raw data, and justify recommendations. Expert system construction and EMR integration could be facilitated IF controlled vocabularies could be made to provide both conceptual and causal knowledge. Expert systems that import these knowledge bases would have ready connections with existing EMR nomenclatures to facilitate information exchange.

This work establishes a Normalized Medical Explanation (NORMEX) syntax, and demonstrates the creation, storage and retrieval of explanatory pathways in a way that is compatible with a controlled terminology. A total of 2,601 explanation pathways were constructed to interpret alterations in serum chemistry values of three analytes: total serum bilirubin (TBil), alanine aminotransferase (ALT) and alkaline phosphatase (ALP). Syntactical classifications of concepts within the pathway were derived from SNOMED top-level hierarchy concepts and labeled as: Observable Analyte, Interpretive Finding, Laboratory Finding, Observable Function, and Disorder. An algorithm was created to convert the pathways into a collection of 1,081 OAV triples and 516 unique concepts. The pathways were then reassembled using only the knowledge of the NORMEX syntax and the stored OAV triples and concepts. Reassembly produced a total of 3,561 pathways, (100% recovery and 960 new pathways). Newly created pathways resulted from correction of ambiguous constructions present in the original Pathways. No significant difference in semantic quality was observed when new pathways were compared to the original by independent reviewers.

Table of Contents

List of Figures

List of Tables

Introduction

Medical expert systems and help systems integrated into electronic medical record systems (EMR) assist users by acting as simple alerting systems or suggesting courses of action.[1-9] Facts stored within the knowledge base of these systems, in conjunction with inferencing techniques, can derive the new knowledge necessary not only to create an alert or a recommendation, but also to justify and explain it.[10-16]

Explanations for the interplay of events in medicine are seldom as simple as asserting that one event caused another; rather, cascading sequences of events often combine to culminate in some clinically detectable phenomenon.[17-19] Explanations vary in complexity and involve diverse factors, ranging from external causative forces to internal perturbations of normal physiology. Additionally, the complexity of human language introduces enormous potential variability into the expression of concepts. [20] A large and active subdiscipline of medical informatics addresses the need to standardize the representation of medical concepts, efforts which may help to resolve the mentioned issues.[21]

Medical informatics standardization efforts have been focused on developing more complex representational terminologies, designed to express the full range of information within an electronic medical record (EMR).[21-29] These representational terminologies are often constructed as polyhierarchies, using interrelationships among concepts to provide formal concept definitions in accordance with the desiderata for the construction of controlled medical terminologies.[28]

One key example of such a terminology is Systematized Nomenclature of Medicine - Clinical Terminology (SNOMED-CT), an amalgamation of two of the larger representational terminologies, Read Codes Clinical Terms Version 3 and SNOMED-RT (Reference Terminology).[26, 30-32] SNOMED-CT currently contains approximately 350,000 concepts that are located in eighteen hierarchies headed by the following top-level concepts: Attribute, Body structure, Context-dependent category, Disease, Environment, Events, Findings, Observable entity, Organism, Pharmaceutical / biologic product, Physical force, Physical object, Procedure, Qualifier value, Social context, Specimen, Staging and scales, or Substance.[33] Concepts (objects) are defined by their associated collection of roles (relationship/attribute) and value concepts (values). These roles and values express facts and relationships which in and of themselves can always be considered true. With approximately 1,500,000 sets of these object attribute value triples (OAV), SNOMED can express complex relationships between concepts. [33] The *IsA* attribute plays a key role in establishing the hierarchy structures within SNOMED.[26, 32-34] Other SNOMED attributes create interhierarchy relationships.[33, 35]

While representational terminologies have provided the capacity for the rich expression of medical concepts within an EMR, their ontologies are focused on content categorization rather than content explanation. A “necessarily true” axiom prohibits

creation of relationships that may be true in some circumstances but not others (less then 98% of the time). As an example, SNOMED states mycobacterial pneumonia *IsA* pneumonia but does not provide the medically important fact that mycobacterial pneumonia is a more likely to occur in an AIDS patient. Such relationships represent much of the knowledge commonly used in medicine for explaining clinical and laboratory information.[20, 36, 37 16, 36, 38-40] This dissonance limits the terminology's utility as a knowledge base in medical expert systems, which rely on such causal relational facts.[1-3, 5, 10, 41, 42]

A syntax for the creation of explanation pathways in medicine would assist in the knowledge engineering necessary to represent such facts.[6, 11, 27, 29, 36] Appropriately constructed knowledge bases could be packaged for dissemination by terminology providers. Finally, importing these knowledge bases into expert systems would provide ready connections with existing EMR nomenclatures to facilitate information exchange.

A Normalized Medical Explanation (NORMEX) syntax was used to standardize the modeling of explanatory information. The syntax provides a format for explaining cause and effect events in medicine defined by using a specific syntactical sequence of concepts from specified SNOMED hierarchies. Representations of these pathways in the NORMEX format were modeled using an ontology tool that was purpose built for this task[1].

For SNOMED to be used as a knowledge base capable of representing explanatory knowledge in an expert system, a mechanism must be devised for the standard representation of that knowledge, allowing for its effective storage and retrieval. A method, based on the NORMEX Syntax, is demonstrated for storage and retrieval of medical explanations from a database structure similar to the one currently used by SNOMED. From this database information alone, the NORMEX-formatted explanation pathways will be reassembled and assessed for degree of information recall and semantic quality as compared to original pathways.

Methods

Creation of Unstructured Explanation Pathways

The pathways represent explanations for alterations of three serum chemistry analytes, Total Bilirubin (TBil), alanine aminotransferase (ALT) and alkaline phosphatase(ALP). Unstructured explanatory pathways (UEP) were plausible sequences of pathophysiologic perturbations that could account for alterations of analyte measurements. UEP were created using a purpose-built ontology-modeling tool designed to allow for concept addition, deletion and duplication. The modeling tool also prevented

[1] Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000

duplication of a concept within any one pathway. Finally, the tool stored the pathways in a database[li].

UEP were assembled as a nested hierarchy. Each analyte represented the trunk of a tree from which explanatory pathways branched. Explanation pathways were assembled sequentially with each new concept a child of a pre-existing concept. Following the analyte concept itself, concepts that represented an altered state of the analyte (increased or decreased measurement) were added as the next children of the UEP. Children concepts explained why the parent concept was observed. The pathway branched when two or more concepts existed as plausible explanations. UEP terminated when a clinically recognized initiator was reached (typically, a disease or disorder).

Evaluation of Concept Categories

The UEP were initially created as a simple one-attribute hierarchy (influences). No attention was paid to any syntactical classification of concepts or the true nature of the relationships that established the hierarchies. Detailed analysis of the UEP began by creating a semantic equivalence map to SNOMED,[lii] using both a desktop and web-based SNOMED browser[liii]. The SNOMED top-level concepts (TLC), also known as SNOMED hierarchy identifiers (i.e., Body structure, Disease, Findings, Observable entity, Organism, Qualifier value, Substance, etc.), were determined for the matched SNOMED concepts. If no SNOMED match was found, the unmatched concept was assigned a logical location within the SNOMED hierarchy. TLC values could then be displayed for UEP concepts. By displaying the TLCs, UEP pathways were examined for consistencies and differences in TLC sequence.

Formalization of the Syntax

A commonly identified TLC sequence was adopted to represent the syntactical normal form of the explanation pathways (NORMEX format). The common TLC sequence was divided into three distinct regions {from results section: [T.O.I.L.]-[(F.L.)$_n$]-[F.D.]}. In addition, functional concepts (SNOMED TLC -- Observable entity) were introduced into the central region of the pathways. The NORMEX syntax was altered to account for the addition of the functional TLC. All UEP were then remodeled, via the ontology tool, into the final NORMEX syntax format.

Relationships

Attributes were created by the author, or existed as intra-hierarchy relational attributes in SNOMED.[35] OAV triples stored in the *Relationships* tables were of two

[li] Microsoft. Microsoft access 2000. 9.0.2720 ed: Microsoft Corporation; 1999

[lii] College of American Pathologists. SNOMED clinical terms. 2002-01Q ed: SNOMED International; 2002.

[liii] Wilcke JR. SNOMED CT Browser. 2.0 ed. Blacksburg: Drug Information Laboratory, Virginia Maryland Regional College of Veterinary Medicine; 2003; Clinical Information Consultancy. CIC Look Up Engine (CLUE). 5.0.0047 ed: Clinical Information Consultancy; 2002

general kinds. One group of OAV contained attribute concepts that expressed relationships between child-parent NORMEX concepts of the reasoning pathway and were directly specified by the structured syntax. These relationships were identified and explicitly expressed as object-attribute-value (OAV) triples via the ontology tool.[2, 22] The other OAV group included attributes that were required for accurate storage and re-assembly of the pathways. These later concepts were either created manually and also served to clarify concept definitions, or were created as the result of processing algorithms.

For Laboratory Finding concepts, four attributes were added manually: *Has state, Has location, Has object*, and *Has focus*. The value for the *Has state* relationship conveys what is happening to the *Has object* value and uses concepts like Increased and Decreased from the SNOMED Qualifier hierarchy. *Has location* values, drawn from the SNOMED Anatomic-structure hierarchy, suggest the site where the Laboratory Finding is occurring. *Has object* is the core item of the NORMEX concept and can be defined as the element affected by the *Has focus* relationship; *Has object* takes as its values SNOMED concepts from the hierarchies of Substances and Anatomic body structure. Lastly, *Has focus* expresses the function or process acting on the *Has object* component and is altered by the *Has state* relationship (Table 5).

Table 5 Example of non-causal relationships created for use with NORMEX concepts

NORMEX Concept	NORMEX Semantic Type	New Non-Causality Relationships	SNOMED Value Range	Example of NORMEX Relationship value
Object		**Attribute**		**Value**
INCREASED MACROPHAGE ERYTHROPHAGIA	Laboratory Finding (L)	*Has state*	Qualifier	Increased
		Has location	Anatomic body structure	Macrophage
		Has object	Substance / Anatomic body structure	Red blood cell
		Has focus	Observable	Phagocytosis
		Has core function	Calculated	Macrophage RBC phagocytosis
		Influences core function	Calculated	Depends on position in hierarchy

Because disorders and findings influence functions in a location-specific (LS) manner and the syntax allows for application of generic location non-specific (LNS) functions, certain additional information must be derived. A new attribute, *Has core function,* was used for all Laboratory Finding concepts. Values assigned with the *Has core function* attribute were word strings concatenated from values for the Laboratory Finding concepts' *Has location*, *Has object* and *Has focus* attribute values. Once *Has core function* values have been assigned, a second new attribute, *Influences core function*, was invoked for all Laboratory Finding concepts that had an Observable Function parent and Disorder concepts (which have an Observable Function parent by design). For a given Laboratory Finding and based on the NORMEX syntax, the next Laboratory Finding concept ancestor (in the relevant pathway) was isolated. From the ancestor, the value associated with its *Has core function* attribute was assigned as the value for *Influences core function* for the new OAV triple.

Creation of Location Specific Function Concepts

Although relationships between findings and location non-specific (LNS) functions (e.g., erythrophagia) remain true in this ontology, strings representing the location-specific (LS) instantiation of the functions (e.g., macrophage erythrophagia) seemed more natural when browsing the assembled hierarchy. LS Observable Function concepts were created and added to the concepts table. OAV triples maintaining the same relationships as the LNS version were added to the relationships table. This allowed for the selection of either LS or LNS versions of the hierarchy when the functions were viewed in the ontology tool. LS Concept names were created by adding the *Has location* value of the parent Laboratory Finding name to the beginning of the original Observable Function name (string concatenation).

Database Structures

An algorithm[liv] was created for use in the ontology-modeling tool for storage of NORMEX tree hierarchies in a database.[lv] The NORMEX knowledge was stored in two tables, Concepts and Relationships. The Concepts table contained unique occurrences of NORMEX concepts and included 3 fields: concept identifier (numeric), concept name (text), concept syntactical classification (text). The Relationships table included 4 fields: relationship identifier (numeric), child concept identifier (numeric), attribute concept identifier (numeric), and parent concept identifier (numeric). The data tables were arranged according to the schema shown in Figure 6. Algorithms used to reassemble pathways created new versions of the Concepts and relationships tables. These tables shared the same structure as the original tables.

[liv] Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000
[lv] Microsoft. Microsoft access 2000. 9.0.2720 ed: Microsoft Corporation; 1999

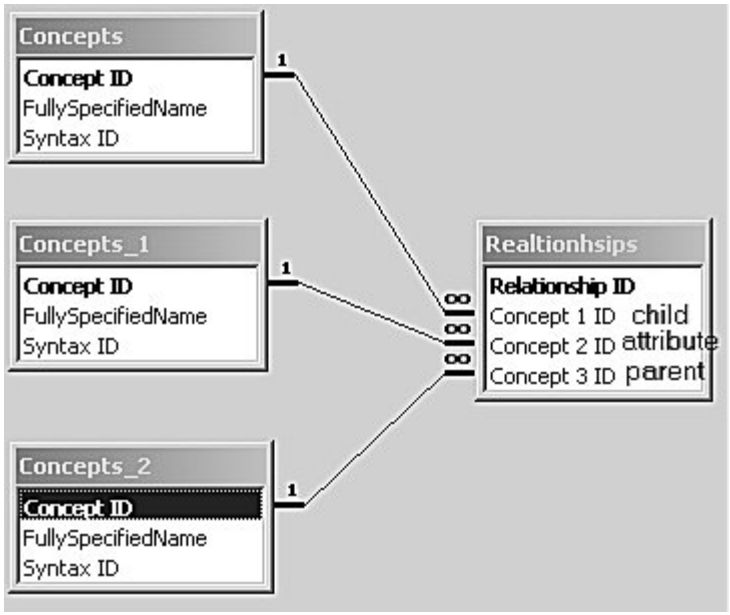


Figure 6 Database schema

Explanation Pathway Storage and Reassembly

NORMEX OAV triples were created from the NORMEX syntax starting with Disorder and concluding with the Top syntactical element. As each concept was processed, OAV triples represented hierarchical triples (the concept, the parent concept and the attribute linking them); non-hierachical triples created as user-supplied definitions; and non-hierachical triples were calculated from existing definitions. Relationships were determined by syntactical elements on either side and by modeler-supplied functional state information.

NORMEX pathways were assembled in the TOIL format by use of a depth-first search algorithm.[45] An additional algorithm was applied that used the *Has Core Function* and *Influences Core Functions* as a processing filter to reject formation of inappropriate pathways that would otherwise occur at Observable Function nodes. Once assembled, the NORMEX model for the three analytes was available for viewing or further modeling within the ontology tool. NORMEX pathways were presented visually as a nested hierarchical arrangement with an analyte concept appearing as a trunk of the tree, and its nested branches representing explanation pathways leading to its probable "Cause(s)".

Semantic Assessment of Original and Reassembled NORMEX Pathways

Individual explanation pathways from original and reassembled hierarchies were extracted as delimited text strings from the original node map and from the reassembled node map. During the extraction process, Observable Function concepts were removed and respective children promoted. Each unique pathway was assigned a random identifier. Each reassembled pathway was compared sequentially to all original pathways by comparing string lists. In this way, reassembled and original pathways were paired. Unmatched reassembled pathways were identified.

Thirty matched NORMEX pathways and thirty non-matching reassembled NORMEX pathways were randomly selected, blinded as to their type, randomly ordered in a single collection, and given to two pathologists for independent review. For easier appreciation of implied casual relationships between concepts in the pathway, the NORMEX concept sequences were placed in the LOIT format[lvi]. The pathologists were asked to score each of the sixty NORMEX pathways using a scale of 1 to 5, with 1 representing a completely non-plausible explanation and 5 representing a completely plausible explanation. Reviewers were instructed to focus on plausibility and not probability of occurrence in their score assignments.

A commercial statistical package[lvii] was used to calculate all statistical parameters. The quality of original (matched) pathways was compared to that of non-

[lvi] The NORMEX syntax expressed in the reverse sequence, stating with (D) and concluding with (T), is referred to as the LOIT format. See: Explanation Pathway Storage and Reassembly

[lvii] Minitab. Minitab statistical software. 13.32 ed: Minitab Inc; 2002

matched pathways (generated during the extraction process) for each pathologist. The nonparametric One-Sign test was used to determine median values, and a Mann-Whitney test was used to evaluate median value differences. Once it could be asserted that there was no statistical difference in quality between the two pathway types, data were pooled for each pathologist to determine whether he considered the explanations to be acceptable overall.[46] A median score of 3 was set as the target score representing a semantically acceptable NORMEX pathway. Again a nonparametric One-Sign test was used for this evaluation.[46]

Results

An example of a NORMEX explanatory pathway is given in Table 6. The NORMEX syntax was developed using 6 semantic types, namely (T)op Of Pathway, (O)bservable Analyte, (I)nterpretive Finding, (L)aboratory Finding, and (D)isorder. The syntax uses 7 attributes, the use of which is dictated by the semantic types involved in various relationships. NORMEX attributes allowed for particular semantic elements, as well as the specific attributes instantiated for an example, are listed in Table 6. Figure 7 demonstrates a typical explanatory pathway as viewed in the ontology modeling tool.

Table 6 Example of NORMEX pathway

NORMEX Region Name	NORMEX Concept	SNOMED Top Level Concept	NORMEX Semantic Types	Syntax Symbol	NORMEX Attribute(s) [lviii]	
					Allowed	Example
T.O.I.L.	ANALYTE	Substance	(T)op of pathway	T		
	TOTAL BILIRUBIN CONCENTRATION	Observable	(O)bservable analyte	O	*IsA*	*IsA*
	HYPERBILIRUBINEMIA	Finding	(I)nterpretative finding	I	*Interprets*	*Interprets*
	INCREASED BLOOD UNCONJUGATED BILIRUBIN CONCENTRATION	Finding	(L)aboratory finding	L	*Favors*	*Favors*
(F.L.)n	ANABOLISM OF BILIRUBIN	Observable	Observable (F)unction	F	*Increased result* *Decreased Result*	*Increased result*
	INCREASED MACROPHAGE BILIRUBIN PRECURSOR CONCENTRATION	Finding	(L)aboratory finding	L	*Favors increase* *Favors decrease*	*Favors increase*
	CELLULAR CATABOLISM OF HEMOGLOBIN	Observable	Observable (F)unction	F	*Increased result* *Decreased Result*	*Increased result*
	INCREASED MACROPHAGE RED BLOOD CELL ERYTHROPHAGIA	Finding	(L)aboratory finding	L	*Favors increase* *Favors decrease*	*Favors increase*
F.D.	CELLULAR ERYTHROPHAGIA	Observable	Observable (F)unction	F	*Increased result* *Decreased Result*	*Increased result*
	DECREASED BLOOD RED BLOOD CELL DEFORMABILITY	Finding	(L)aboratory finding	L	*Favors increase* *Favors decreased*	*Favors increased*
	CELLULAR DEFORMABILITY	Observable	Observable (F)unction	F	*Increased result* *Decreased Result*	*Decreased result*
	ANAPLASMOSIS	Disorder	Disorder	D	*Favors increase* *Favors decrease*	*Favors decrease*

[lviii] Relational concepts in dark grey exist in SNOMED, light grey relationships created by author

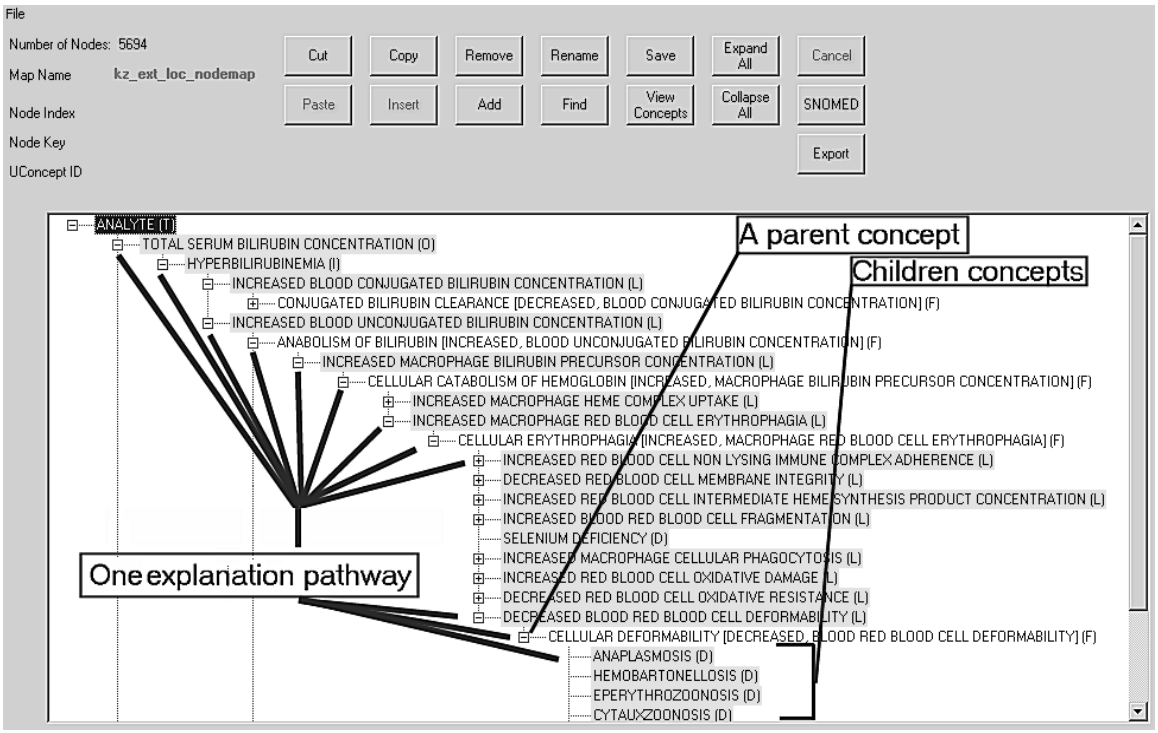


Figure 7 Screen shot of ontology modeling tool showing nested tree structure of pathways[lix]

[lix] Parent/child concepts are depicted, along with the root concept, and example of branching point is noted. The item in parentheses at the end of the concept name represents the syntactical classifier as implied by the final NORMEX format.

The top-down specific NORMEX syntax of the example in Table 6 is T.O.I.L.F.L.F.L.F.L.F.D. The repeating series of F.L. syntax symbols example can be simplified to the following general sequence, referred to as the TOIL format of NORMEX: T.O.I.L. - (F.L.)n - F.D. where n is equal to a collection, whose size is 0 or more, formed by an Observable Function concept followed by a Laboratory concept. Three distinct regions of the NORMEX syntax can be seen ([T.O.I.L.], [(F.L.)n], and [F.D.]). The NORMEX syntax can be also be expressed in the reverse sequence, starting with (D) and concluding with (T), referred to as the LOIT format.

Explanation Pathway Storage and Reassembly

There were 2601 NORMEX pathways created using the ontology-modeling tool. The pathways ranged from 12 to 47 concepts in length. The distribution of the pathways among the analytes is shown in Table 3. A total of 468 unique concepts were used. The number of concepts assigned to each NORMEX syntax category is shown in Table 7.

Table 7 Distribution of number of unique concepts by NORMEX syntax classification as originally modeled

Pathways	Analyte	# Pathways	Nodes	Unique Concepts	NORMEX Syntax Symbol	Number Unique Concept
2601	TBil	1,090	4,346	468	T	1
					O	3
	ALP	1,011			I	4
					L	130
	ALT	500			F	120
					D	210

There were 3561 NORMEX pathways reassembled from the concept and relationship table storage format. This resulted in a net increase of 960 new pathways not present in the original ontology tool model, with 100% of original pathways identified in the reassembled set (Table 8). Reassembled pathways were viewable using the modeling interface provided by the ontology-modeling tool (Figure 7).

Table 8 Comparison between number of original and reassembled NORMEX pathways

Analyte	Original Pathways (N)	Reassembled Pathways (N)	Recalled Original Pathways [N (%)]	New Reassembled Pathways (N)
TBil	1,090	1,486	1,090 (100%)	396
ALP	1,011	1,391	1,011 (100%)	380
ALT	500	684	500 (100%)	184
Total	**2,601**	**3,561**	**2601 (100%)**	**960**

Semantic Assessment of Original and Reassembled NORMEX Pathway

Median scores for the original and reassembled groups of NORMEX pathway samples each ranged from 3-4 and were statistically equivalent to each other (Table 9). Frequency distribution of assigned scores by NORMEX group for each pathologist is shown in Table 10. For both pathologists, the null hypothesis that the median score for the NORMEX sample was less than or equal to 2 was not supported (Table 10).

Table 9 Median score comparisons between original and reassembled NORMEX Pathways

Original and Reassembled Pathway Score Comparison		
	Pathologist 1	**Pathologist 2**
	N, (95% CI)	
Median Original NORMEX Score*	4.000 (4.000, 4.771)	3.000 (2.000, 3.000)
Median Reassembled NORMEX Score *	4.000 (3.000-4.000)	3.000 (2.000-3.000)
H_0: Original - Reassembled = 0#	0.000 (0.0001, 1.0002)	0.000 (-0.9998, 0.0000)
P for H_0	0.1405	0.2627
Conclusion	Cannot Reject Equality	Cannot Reject Equality
* 1-Sample Sign Test		
# Mann-Whitney Test		

Table 10 Frequency distribution of scoring by the two pathologists looking at both original and new NORMEX Pathways

Pathway Score Tally And Median Assessment Score (1-Sample Sign)		
	Pathologist 1	Pathologist 2
Score Category	N (%)	N (%)
1	0 (0)	4 (6.67)
2	3 (5)	21 (35.00)
3	17 (28.33)	28 (46.67)
4	23 (38.33)	7 (11.67)
5	17 (28.33)	0 (0)
Total	60 (100)	60 (100)
Median Score (95% CI)	4.000 (4.000, 4.000)	3.000 (2.000, 3.000)
P for H_0: Median <= 2	0.0000	0.0000
Conclusion	Reject H_0 (Median <= 2)	Reject H_0 (Median <= 2)

Discussion

This paper demonstrates a method for the standardization of medical explanations interpreting clinical pathology information. Five syntactical categories were created based on the relevant SNOMED hierarchy locations: Observable Analyte, Interpretive Finding, Laboratory Finding, Observable Function, and Disorder. A sixth category, Top, was a user-defined class created to represent the root of the NORMEX model.

Initially, a single attribute hierarchy was considered for the intended purpose of organizing knowledge about causation. This is consistent with SNOMED's hierarchical organization based on single attributes (i.e., *IsA* or *Part_of*). Causation could be modeled as a single attribute hierarchy (e.g., "Causes"). However, causation is linked to probability (alterations in laboratory values associated with a particular disorder have a frequency distribution). Infection with the hemoparasite *Anaplasma marginale* will reduce red cell mass under certain circumstances and not others. Integration of causation with strict inheritance-based hierarchies such as SNOMED encourages us to maintain fidelity with the "always and necessarily true rule". As a consequence, definitions were created which can be processed to assemble a hierarchy using multiple attributes. When assembled, this hierarchy can be viewed as a hierarchy with a single pseudo-attribute "Causes."

The first step in the approach was to build UEPs based on the pseudo-attribute "Causes." The purpose-built ontology tool was developed to facilitate integration of required defining attributes while viewing the hierarchy as if it were a single-attribute hierarchy. While other ontology modeling applications exist, the created ontology tool provided specific functionality not present in these other applications.[47-52] Using knowledge representation based upon frame structure or principles of logic, these other applications result in the construction of ontology hierarchies with a single (e.g., *IsA*) relationship between parent and child concepts. Construction of the single-attribute (explanation) hierarchies in the NORMEX model was based on the mix of the relationships or attributes shown in Table 6.

The relationship names used in the NORMEX model were selected with the knowledge that SNOMED only contains relationships that are true at least 98% of the time.[33, 35] All of the explanatory relationships used in the pathways meet this requirement. For example, in the F.D. section of the model (Table 6) the Disorder concept is related to the Observable Function concept by either the *Favors increased* or the *Favors decreased* relationship. The model is designed such that all other relationships exist in a true state.

Adoption of this approach allows the flexibility to express an explanation that is always true and the single pseudo-attribute hierarchy accommodates the inherent uncertainty that exist in much of medicine.[36, 39] This uncertainty is accommodated in the (F.L.)n. and F.D. sections of the model. Here the relationships used in a particular pathway instance identify a particular combination of functions and findings which point

to a single possible outcome. The generic nature of the Observable Function concepts serves as a possible point of attachment for competing influences and opposing outcomes. We do not claim that the syntax can carry actual probabilities, nor determine outcomes. However, the NORMEX model allows us to express influences between parts of the model within a framework of truthful relationships. In the T.O.I.L. section of the model, the employed relationships do not accommodate this level of uncertainty.

An alternate syntax for causal reasoning was considered as this project unfolded. It is possible to create causal reasoning chains without employing Observable Functions. In fact, pathways were presented to pathologists for evaluation after functions were eliminated without loss of understandability. Specifically, the use of Function state findings, (e.g., Increased macrophage erythrophagia) and generic causal relationships (e.g., favors without increase or decrease) would have simplified the pathways and processing. However, during the initial creation of the UEP, it was a constant challenge to stay focused on the underlying physiology responsible for the observed Laboratory Finding concept event within the pathway. Using a generic Observable Function concept, and envisioning the function as being perturbed in some way, made it easier to imagine what observable Laboratory Finding concept would result from the function perturbation. This became the basis for modeling all of the Laboratory Finding concepts and their children Observable Function concepts within the model. As a pathologist, the author found this a very intuitive approach to explaining the noted Laboratory Findings within the pathways. Diseases are manifestations of the loss of homeostasis in some physiologic function.[53, 54] Generic functions provide another important theoretical advantage in that it should be possible to express function concepts within partonomy hierarchies with associated inheritance and transitivity characteristics.[55, 56] Such a partonomy hierarchy would allow the modeling of subparts of functions, which could conceivably be carried to the cellular or molecular level. A partonomy of functional concepts will allow biomedical informatics data from proteomic and genomic work to be integrated with clinical observable events represented within the NORMEX model.

The use of General Function concepts does not impair processing efficiency in the conversion to the OAV storage format. During reassembly, however, a problem can arise when two or more pathways share a General Function concept. An example of two such pathways is shown in Figure 8.

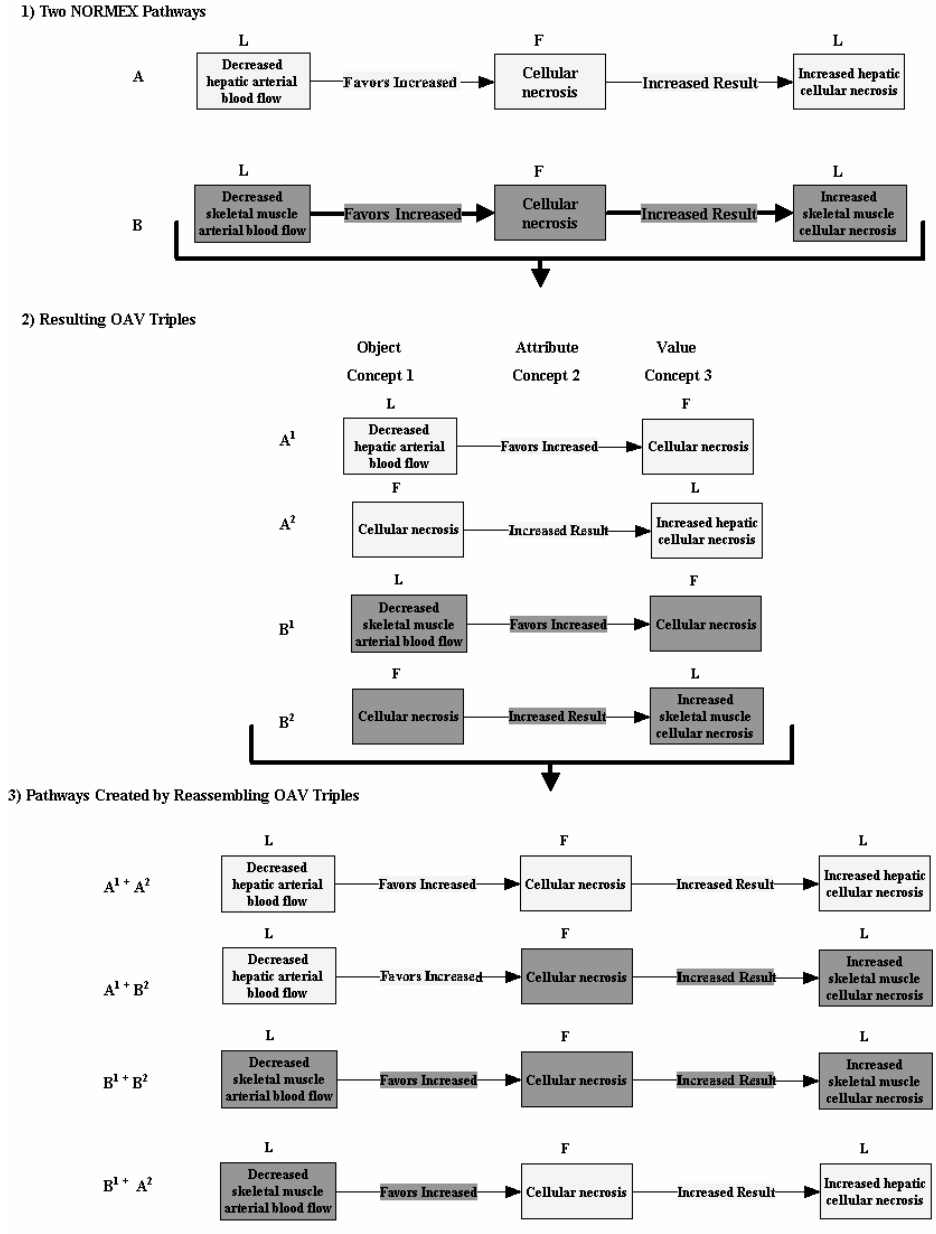


Figure 8 NORMEX pathways converted to OAV triples and reassembled

The pathway segments labeled A and B in Figure 8 were created during the initial modeling process. When stored, these two segments are represented by OAV triples A^1, A^2, B^1 and B^2. On reassembly and without additional intervention, the four pathways at the bottom of the figure emerge. Pathways ($A^1 + A^2$) and ($B^1 + B^2$) are the original and correct pathways. Pathways ($A^1 + B^2$) and ($B^1 + A^2$) represent physically non-valid, but logically valid, reassembly of OAV triples (physiologically inconsistent pathways --PIP). Pathways ($A^1 + B^2$) and ($B^1 + A^2$) clearly do not make sense. The explosion in the number of PIP was quite dramatic when all 1,861 OAV triples were involved in the reassembly process, resulting in approximately 23,000 new pathways not seen in the original model (Table 11).

Table 11 Variation in number of pathways, concepts, and relationships between location nonspecific (LNS) and location specific (LS) needed to represent ontology tree information

	Ontology Tree	LNS	LS
Pathways	**2601**	22,822	3,561
Unique Concepts	**468**	606	627
Relationships	**4346**	1,101	1,116

The processing steps that created *Influences core function* and H*as core function* attributes and values were invoked to deal with this reassembly ambiguity. The value concept for each of these attributes was determined at the conversion of the original pathways into the OAV format for storage, as outlined in the Methods section of this manuscript. These two new attribute relationships provided a common value for pathway concepts on either side of a Function concept and allowed inappropriate pathway combinations to be rejected. The cost of this the reduction in ambiguity was an increased number of OAV representing these new relationships and some number of processing steps. Rejection of inappropriate pathways is displayed graphically in Figure 9.

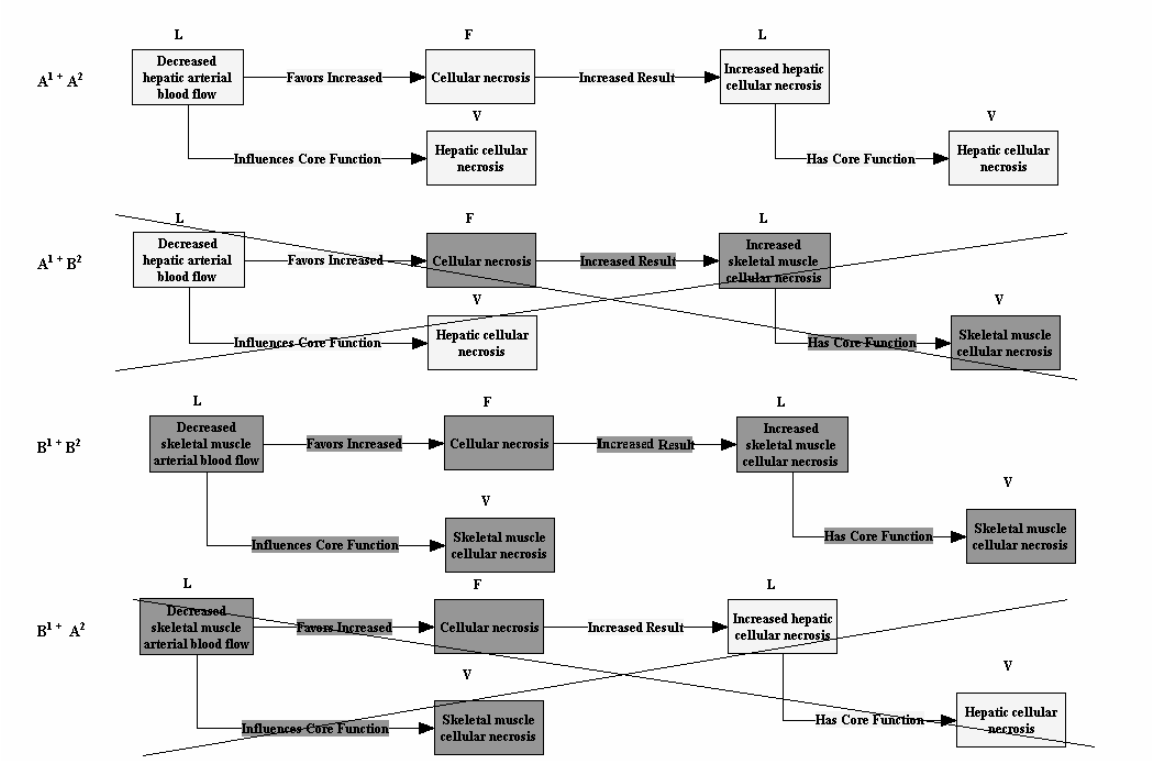


Figure 9 Identification of semantically incorrect pathway by use of *Has core function* and *Influences core function* attributes and values[lx]

[lx] Circled concepts indicate semantic misalignment of pathways (L = Laboratory Finding, F = Function, V = Value

There were still 960 new NORMEX pathways present in the reassembled pathways following adoption of the two new attributes. The original hierarchy was visually compared to the reassembled hierarchy in order to evaluate the source of these new pathways. To facilitate this process, the second of the two modifications of the original NORMEX model was made. Specifically, the semantic meaning of LNS instantiations of General Function concepts was improved by the use of more location-specific (LS) forms of the General Function concepts. This involved the derivation of a location-specific or LS version of the NORMEX Function - concept.

The algorithm for creating LS forms was amended to the process that converted the original pathways to OAV triples to create a second LS version of the OAV triple tables. The H*as location*-value relationship associated with the parent Laboratory concepts was prepended to the LNS Function concept to create the name for a new LS Function concept. While this modification of the original Function concept provided no additional discriminating power for the reassembly of semantically correct pathways, it did clarify the origins of these new pathways and the overall understandability of the reassembled explanation pathways (Figure 10).

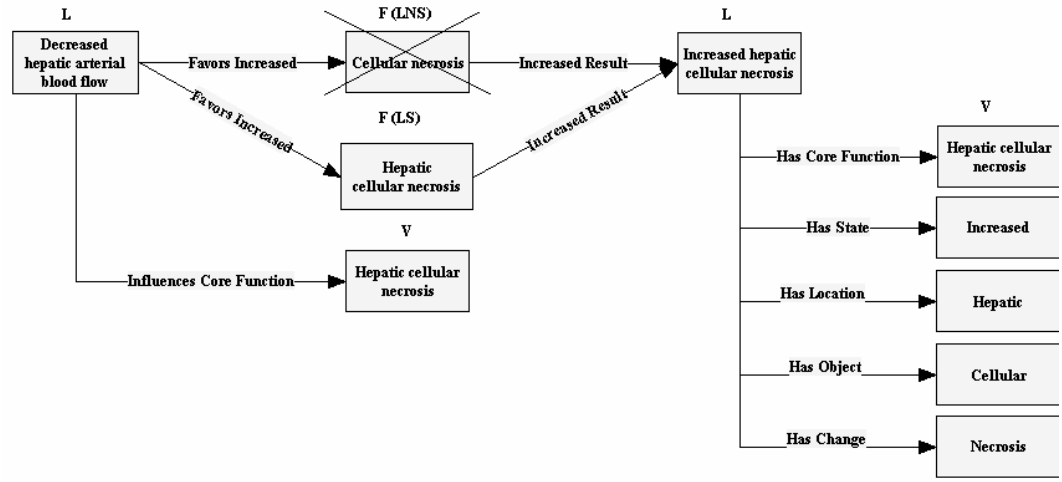


Figure 10 Example of location specific Function type concept values[lxi]

[lxi] L = Laboratory Finding, F = Function, V = Value, LS = Location Specific, LNS = Location Nonspecific

Display of LS Observable Function improved the understandability of the reassembled pathways. These new pathways were found to result from variability that could have been controlled during the construction of the original pathways. While modeling a pathway from a particular perspective, the modeler was likely to select well recognized influences for any individual step. Less well recognized Observable Function influences were occasionally ADDED during the creation of additional pathways. In the example shown in Figure 11, the Observable Function "cellular necrosis" has two *favors increased* children in the first pathway. In the second pathway, one of these two children is repeated and two new children are added. After deconstruction and reassembly, the reconciled pathways include four *favors increased* children appearing distal to the Observable Function "cellular necrosis."

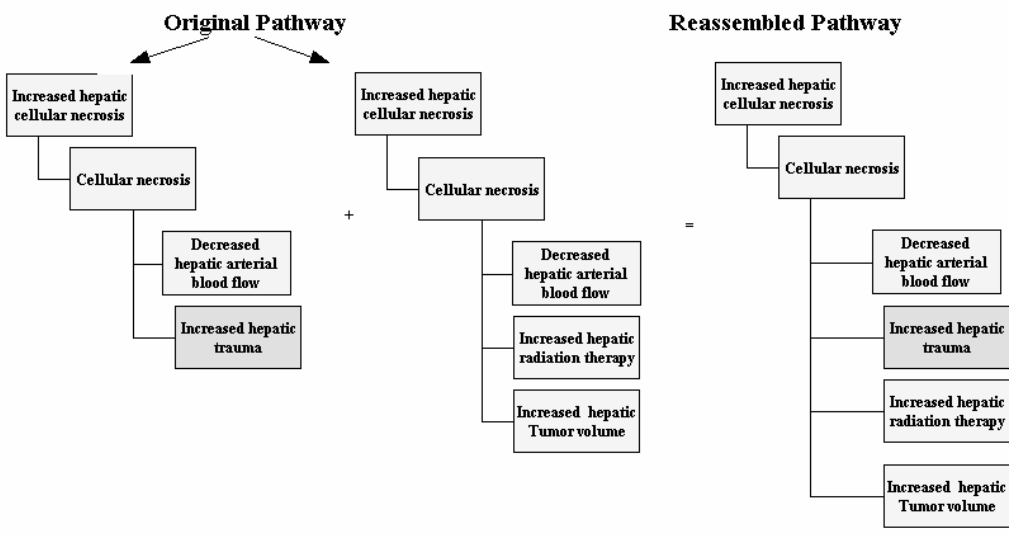


Figure 11 Example of variability in original NORMEX pathways and its resolution in the reassembled pathways[lxii]

[lxii] Case of ambiguous construction for parent concept "Cellular necrosis"

While the ontology tool prevented the creation of inconsistency in the relation order of concepts in any one pathway and the model as whole by resolution refutation, the tool did not prevent variability of children at similar branch points, as long as logical ordering inconsistency was not produced. [22, 45] The reassembly algorithms using the OAV triples fully integrated these variable representations of children concepts (Figure 11). To verify that that was in fact true for the reassembled model, a quality control algorithm was run on the pathways to check for the presence of ambiguity or redundancy in the pathways in regards to explanatory relationships. This ambiguity and redundancy was defined by examining the collection of children concepts associated with each parent concept in the reassembled LOIT-formatted pathways. As expected, there were no cases of ambiguity or redundancy in the reassembled NORMEX model. The ontology-modeling tool could be redesigned to prevent this ambiguity during modeling.

The lack of redundancy, ambiguity, or inconsistencies in the logical order of concepts in the reassembled NORMEX model does not address the question as to whether these new pathways made semantic sense.[28] Independent evaluation by two pathologists was used to assess the semantic quality of explanations provided by both original and new pathways. Using criteria established for this evaluation, each considered the median plausibility for all reassembled (original and new) pathways to be adequate. There was also no statistical difference in either pathologist's evaluation of new pathways as compared to originals.

It is not surprising that there is variability in the median score and distribution of scores between the two pathologists for the two types of pathways. Some of the difference seen between the two pathologists is also likely due to the difference in their individual training background and in experience level. Expert opinions can vary to a great extent on the precise nature of the cause for a particular laboratory finding.[40] More important than agreement between pathologists was their general assessment that both types of pathways made sense and that there was no significant difference in semantic quality between the types of paths.

While developing this model, two bases for explanation were ignored. First, explanations of spurious results were removed from consideration. Spurious results may be due to selection of the wrong test method for a particular species, to limits of the test range itself, to improper specimen handling, to interference by substances within the specimen, or simply to errors in quality control or in reporting. The sources of spurious results thus represent a particular group of explanations whose size was beyond the scope of the present work. No specific aspect of the existing NORMEX model prevents their representation, however, and certainly spurious influences are important to consider in explaining laboratory results.[53, 57] The explanation of spurious results was considered to be a special case that bears critical analysis in its own right. Explanations of such results were therefore omitted from the model for the three analytes at this stage of the project. Another simplification imposed on the model was the omission of explanations that might vary with species.[53, 57] All the pathways modeled represent classic explanations, as

might be found in a standard reference text on veterinary clinical pathology. Most of the modeled explanations were not species-dependent. In those rare cases where a clinical disorder was species-specific, the relevant species was indicated in the name of the disorder (*bovine* postparturient hemoglobinuria). A more robust solution considered for future version of the model is to add concept relationships that will explicitly associate particular clinical laboratory findings with the appropriate species.

The purpose-built ontology tool was enabled with certain features that distinguish it from other ontology modeling tools.[47, 50, 52] The purpose-built ontology tool allowed for assembly of explanation hierarchies without regards to inheritance constraints imposed by individual concept definitions. Traditional ontology tools assemble and evaluate hierarchies after concept definitions are created. The purpose-built tool simultaneously facilitated assembly of explanation hierarchies, guided the assignment of concept categories, and enforced the integrity of concept definitions created by the process. Concept-attribute-value relationships could still be modeled differently in different locations. When viewed as a hierarchy, concept definitions might appear to be ambiguous.[28, 58-60] Concept reusability proved very useful, and allowed the user to specifically express the particular explanations unconstrained by prior explanations. Resolution refutation was used to avoid logic inconsistency in the explanation pathways where one concept appeared before another in one location and then after it in another. [22, 61] Thus the tool insured the relative order of the appearance of all concepts within the model.

In this experiment, all concepts and relationships were modeled as if they existed outside the current content of SNOMED. However, steps were taken to align the syntax with SNOMED as it is currently organized. When SNOMED attributes could serve a useful purpose in the model (the T.O.I.L section), they were used. Many of the concepts employed in pathways are currently represented by SNOMED. Finally, the technique for OAV triple representation in relational tables mimics the table structures used in SNOMED. However, during the modeling process no attempt was made to identify SNOMED concepts that might have been applicable. Such concepts would be a subset of SNOMED. Necessary concepts that could not be located in SNOMED would exist as an extension of SNOMED.[35] Similarly, attributes that have no functional classification role in SNOMED might remain outside the nomenclature.

Conclusion

The creation of explanatory information stored as facts plays a vital part in the development of expert systems.[11] It is hoped that designing a standard syntax to represent an explanation, one aligned with SNOMED, will assist in the creation and reuse of such explanation knowledge. Assessment of the NORMEX Syntax capability of providing just such a mechanism will open opportunities to incorporate various types of explanatory information from multiple fields of medicine into SNOMED. It also important to remind the reader that the NORMEX model can be used to represent explanations of clinical

disorders and related clinical findings from areas of medicine other than clinical pathology. Clinical pathology (specifically serum chemistry) was selected as the domain for modeling purposes because of the author's expertise and interest in this topic area. This domain also is easily envisioned as potentially applicable for use in expert systems and decision-support applications. This paper is intended to demonstrate the feasibility of NORMEX's methodology to model, store, and retrieve knowledge, irrespective of the knowledge's particular domain. It is speculated that such NORMEX- represented domain knowledge in an architectural storage format compatible with SNOMED will enable the efficient dissemination and reuse of that knowledge.[42]

Explanation hierarchies within the NORMEX model were constructed via a specific set of relationships stipulated by the location of the various syntactical elements, and can be regarded as being true greater than 98% of the time. The modeling process enforced logical causality sequence of concepts, and used non-explanatory relationships and values to prevent the creation of redundant and ambiguous concepts. The purpose-built ontology tool provided a development environment that allowed the assemblage of explanatory pathways adhering to the syntactical pattern $TOIL(FL)_nFD$ created by using SNOMED-CT hierarchies.

The pathways can then be stored in a SNOMED- like database schema. NORMEX information can then be reassembled from the database storage format without loss of original information content, creation of redundant or ambiguous relationships, or loss of the semantic explanation quality. The tools and methodologies used for the creation of the original NORMEX pathways and their storage as OAV triples and reassembly may be applicable for use as a standardized method for representing various types of medical explanations and permitting their retrieval from an existing controlled medical terminology, such as SNOMED.

Acknowledgment

This work has been funded by the National Institutes of Health, National Library of Medicine Informatics Research Fellowship Award, grant number 1 F37 LM07699.

Special thanks are given to Dr. Elizabeth Besteman, Dr. Tripp Almay and Dr. Bernard Feldman for offering their clinical pathology expertise in review of explanation pathways.

Bibliography

1. Baak JP, Kurver PH. Development and use of a rule-based pathology expert consultation system. Anal Quant Cytol Histol 1988;10(3):214-8.
2. Bartels PH, Hiessl H. Expert systems in histopathology. II. Knowledge representation and rule- based systems. Anal Quant Cytol Histol 1989;11(3):147-53.

3. Bartels PH, Weber JE. Expert systems in histopathology. I. Introduction and overview. Anal Quant Cytol Histol 1989;11(1):1-7.
4. Baxt WG. Complexity, chaos and human physiology: the justification for non-linear neural computational analysis. Caner Lett 1994;77:85-93.
5. Berner ES. Clinical decision support systems: theory and practice. New York: Springer; 1999.
6. Connelly DP, Bennett ST. Expert systems and the clinical laboratory information-system. Clin Lab Med 1991;11(1):135-51.
7. Dolin RH. Advances in data exchange for the clinical laboratory. Clin Lab Med 1999;19(2):385-91, vii.
8. McNeely MDD. The use of expert systems for improving test use and enhancing the accuracy of diagnosis. Clin Lab Med 2002;22:515-28.
9. Sieben G, Praet M, Roels H, Otte G, Boullart L, Calliauw L. The development of a decision support system for the pathological diagnosis of human cerebral tumours based on a neural network classifier. Acta Neurochir (Wien) 1994;129(3-4):193-7.
10. Bartels PH, Thompson D. Expert systems in histopathology. III. Representation of knowledge as "structured objects". Anal Quant Cytol Histol 1989;11(6):367-74.
11. Warner HR, Sorenson DK, Bouhaddou O. Knowledge engineering in health informatics. New York: Springer; 1997.
12. Birndorf NI, Pentecost JO, Coakley JR, Spackman KA. An expert system to diagnose anemia and report results directly on hematology forms. Comput Biomed Res 1996;29(1):16-26.
13. Diamond LW, Mishka VG, Seal AH, Nguyen DT. Are normative expert systems appropriate for diagnostic pathology? J Am Med Inform Assoc 1995;2(2):85-93.
14. Edwards G, Compton P, Malor R, Srinivasan A, Lazarus L. PEIRS: a pathologist maintained expert system for the interpretation of chemical pathology reports. Pathology (Phila) 1993;25(1):27-34.
15. Roach J, Lee S, Wilcke J, Ehrich M. An expert system for information on pharmacology and drug interactions. Comput Biol Med 1985;15(1):11-23.
16. Heckerman DE, Horvitz EJ, Nathwani BN. Toward normative expert systems: Part I. The Pathfinder project. Methods Inf Med 1992;31(2):90-105.
17. Heckerman D, Shacte R. Decision-theoretic foundations for causal reasoning. Journal of Artificial Intelligence Research 1995;3:405-30.
18. Pearl J. Causal diagrams for empirical research. Biometrika 1995;82(4):669-710.
19. Nease RF, Jr, Owens DK. Influence diagrams in structuring decisions. Med Decis Making 1997;17(3):236-75.
20. Rector AL. Clinical terminology: Why is it so hard? Methods Inf Med 1999;38(4-5):239-52.
21. Zanstra PE, Rector AL, Ceusters W, de Vries Robbe PF. Coding systems and classifications in healthcare: the link to the record. Int J Med Inf 1998;48(1-3):103-9.
22. Sowa JF. Knowledge representation: logical, philosophical, and computational foundations. Pacific Grove: Brooks/Cole; 2000.

23. Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
24. Cimino JJ. Review paper: Coding systems in health care. Methods Inf Med 1996;35(4-5):273-84.
25. Rector AL. Terminology and concept representation languages: where are we? Artif Intell Med 1999;15(1):1-4.
26. Rothwell DJ. SNOMED-based knowledge representation. Methods Inf Med 1995;34(1-2):209-13.
27. Musen MA, Wieckert KE, Miller ET, Campbell KE, Fagan LM. Development of a controlled medical terminology - knowledge acquisition and knowledge representation. Methods Inf Med 1995;34(1-2):85-95.
28. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(Nov(4-5)):394-403.
29. Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. J Am Med Inform Assoc 1998;5(6):503-10.
30. Nash S, Sable JH, Stearns MQ, Robinson D. Merging domain hierarchies in the development of SNOMED CT. Proc AMIA Symp 2001:984.
31. Southwick K. SNOMED, Read codes Unit in Blockbuster alliance, Combine sets up unrivaled leadership. Paper presented at: CAP Today, 1999; Northfield, IL.
32. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp 1997:640-4.
33. College of American Pathologists. Concept modeling style/authoring guide: SNOMED International; 2002.
34. College of American Pathologists. SNOMED CT product profile: SNOMED International; 2003.
35. College of American Pathologists. SNOMED clinical terms consultation document: core structure: SNOMED International; 2000.
36. Sox HC. Medical decision making. Boston: Butterworths; 1988.
37. Owens DK, Sox HC. Medical decision making: probabilistic medical reasoning. In: Shortliffe EH, editor. Medical informatics. 2nd ed. New York: Sringer-Verlag; 2001. p. 76-131.
38. Parmigiani G. Modeling in medical decision making - a Bayesian approach. West Sussex, England: John Wiley & Sons, Ltd; 2002.
39. Ward CD. Meditations on uncertainty. In: Kohout L, Bandler W, editors. Knowledge representation in medicine and clinical behavioral science. Cambridge, Mass.: Abacus Press; 1986. p. 19-25.
40. Morgan MG, Henrion M, Small M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge; New York: Cambridge University Press; 1990.
41. Bartels PH, Thompson D, Weber JE. Expert systems in histopathology. IV. The management of uncertainty. Anal Quant Cytol Histol 1992;14(1):1-13.

42. Cimino JJ. Distributed cognition and knowledge-based controlled medical terminologies. Tecklenburg, Federal Republic of Germany: Burgverlag; 1989.
43. Microsoft. Microsoft access 2000. 9.0.2720 ed: Microsoft Corporation; 1999.
44. Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000.
45. Cawsey A. The essence of artificial intelligence. London ; New York: Prentice Hall; 1997.
46. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford; Boston: Blackwell Scientific Publications; 1994.
47. Grosso WE, Eriksson H, Fergerson RW, Gennari JH, Tu SW, Musen MA. Knowledge modeling at the millennium (the design and evolution of protégé-2000): Stanford Medical Informatics; 1999. p. 1-36.
48. Stevens R, Goble C, Horrocks I, Bechhofer S. Building a bioinformatics ontology using oil. IEEE Trans Inf Technol Biomed 2002;6(2):135-41.
49. Rothenfluh TE, Gennari JH, Eriksson H, Puerta AR, Tu SW, Musen MA. Reusable ontologies, knowledge-acquisition tools, and performance systems: PROTEGE-II solutions to Sisyphus-2. Int J Hum Comput Stud 1996;44(3/4):303-32.
50. Rector AA. Oiled normalized ontology tutorial-biomedical version: Medical Informatics Group / Bio & Health Informatics Forum; 2002.
51. Noy NF, Fergerson, R. W., Musen MA. The knowledge model of Protégé-2000: combining interoperability and flexibility. Juan-les-Pins, France: Stanford Medical Informatics; 2000. p. 1-20.
52. Apelon. Terminology development environment. Ridgefield: Apelon; 2003.
53. Stockham SL, Scott MA. Fundamentals of veterinary clinical pathology. 1st ed. Ames, Iowa: Iowa State Press; 2002.
54. Cotran RS, Kumar V, Collins T, Robbins SL. Robbins pathologic basis of disease. 6th ed. Philadelphia: Saunders; 1999.
55. Artale A, Franconi E, Guarino N, Pazzi L. Part whole relations in object centered systems: an overview. Data and Knowledge Engineering 1996;20:347-83.
56. Schultz S, Romacker M. Part whole reasoning in medical ontologies revisited - introducing SEP triplets into classification - based description logics. Proc AMIA Symp 2001.
57. Duncan JR, Prasse KW, Mahaffey EA. Veterinary laboratory medicine. 3rd ed. Ames: Iowa State University Press; 1994.
58. Rector AL. Defaults, exceptions and ambiguity in a medical knowledge representation system. Med Inform 1986;11(4):295-306.
59. Cimino JJ. Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. Methods Inf Med 1996;35(3):202-10.
60. Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scalable methodologies for distributed development of logic- based convergent medical terminology. Methods Inf Med 1998;37(4-5):426-39.
61. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. J Am Med Inf Assoc 1994;1(3):218-32.

SNOMED Coverage of NORMEX Concepts

by
Kurt Lee Zimmerman

Virginia Polytechnic Institute & State University
Department of Biomedical Sciences and Pathobiology
540.231.3692
kzimmerm@vt.edu

Keywords: SNOMED, NORMEX, Clinical Pathology, Coverage

Abstract

Normalized Medical Explanation (NORMEX) is a syntax and storage format based on a controlled terminology- Systematized Nomenclature of Medicine (SNOMED). The NORMEX format was used to create explanation pathways for the interpretation of clinical pathology laboratory data. This experiment examined the capacity of SNOMED to represent concepts used in the NORMEX model of the topic area.

Wide variability was found in SNOMED coverage of NORMEX concepts ranging from 5-62% depending on the NORMEX-syntactical classifications being examined. The overall average SNOMED coverage of NORMEX concepts was 31%. NORMEX-syntax Disorder concepts were the best-represented syntax elements with a coverage of 62% followed by Top, Observable Analyte, and Interpretive Findings at 37%. Coverage of Observable Functions was 16% and Laboratory Findings was lowest at 5%.

These results suggest that the Observable and Findings hierarchies of SNOMED are most in need of additional content for modeling of NORMEX pathways. A minor portion of the concepts used in the NORMEX model reside in a subset of SNOMED (31%), while a majority would exist as an extension of SNOMED (69%).

Table of Content

List of Figures

List of Tables

Introduction

SNOMED-CT currently contains over 350,000 concepts that are located in eighteen hierarchies.[1-4] Concepts are defined by their associated collection of roles and value concepts.[5] With approximately 1,500,000 sets of these object-attribute-value triples, SNOMED is very expressive in its ability to relate concepts to one another.[2, 6] The *IsA* attribute is a part of the definition for all concepts and plays a key role in establishing the hierarchy structures of SNOMED. Many of the other SNOMED attribute concepts, when used in OAV triples, create interhierarchy relationships.[7]

A proposed format has been made for representing a sequence of concepts explaining clinical pathology laboratory result interpretations.[8] This model, the Normalized Medical Explanation (NORMEX), is syntactically defined by SNOMED hierarchies and is aligned with SNOMED's database storage format. A modeling tool was developed for the creation of clinical pathology explanation pathways in the NORMEX format.

Modeling explanatory pathways in the NORMEX format uses concepts focused on providing insight to altered pathophysiology. Reviews have shown SNOMED to be very capable of representing medical concepts for use in electronic medical records.[4, 9-14] However, other work has shown that SNOMED coverage of concepts relevant to clinical pathology may be lower than reported in other areas of medicine.[15] So, while the NORMEX model is aligned syntactically and from a storage standpoint with SNOMED, it is not clear to what degree SNOMED can provide actual concept content for the model. Ideally, a knowledge-modeling tool to assist users in representing domain explanatory knowledge will guide the syntactical format and storage of such knowledge but additionally point to a resource providing standardized medical concepts. SNOMED would be a logical choice as such a resource, but first SNOMED's coverage of explanatory concepts must be assessed.

This experiment examines SNOMED's coverage of explanatory concepts used in creating NORMEX pathways. From this work, areas of SNOMED needing additional content to represent explanation pathways will be identified, helping to focus future SNOMED content development efforts. Additionally, this work will provide preliminary estimates of the distribution of explanatory concepts between SNOMED subsets and extensions.

Methods

A statistical software package[lxiii] was used for the statistical calculations of this work. Syntax classification of location-nonspecific concepts was tabulated from a database table[lxiv] previously created in storage of NORMEX pathways.[8] Concepts were

[lxiii] Minitab. Minitab statistical software. 13.32 ed: Minitab Inc; 2002
[lxiv] Microsoft. Microsoft access 2000. 9.0.2720 ed: Microsoft Corporation; 1999

sorted into 6 groups according to syntax classification: ((T)op, (O)bservable Analyte, (I)nterpretive Finding, (L)aboratory Finding, Observable (F)unction, (D)isorder). Concepts of the syntactical type (T)op, (O)bservable Analyte, and (I)nterpretive Finding were combined to form a grouping of concepts named Group 1. Concept Groups 2, 3, and 4 were created from the concepts of the syntactical types (L)aboratory Finding, Observable (F)unction, and (D)isorder respectively. Following the creation of the four concept groups, each concept was assigned a unique identifier for purposes of randomization. Thirty concepts from each group were randomly selected and placed in a spreadsheet[lxv] for storage.

The randomly selected concepts from each of the four groups were independently mapped by the author and another SNOMED terminology expert to the best semantically matching SNOMED[lxvi] concept.[10] Mapping was done using both a desktop[lxvii] and web-based SNOMED browser[lxviii]. Reviewers assigned a "quality of match" score for each mapping attempt. A match score of (1) – problem - was recorded for a sample-group concept when a semantically equivalent SNOMED concept was not found. A match score of (2) – good - was assigned when a semantically equivalent SNOMED was found. SNOMED hierarchy classification was also recorded for instances of good matches. Calculation of agreement statistics (number matched, percentage matched, and Kappa value) between reviewers was done following scoring completion.[16-18] Quality of match scores were tabulated for each reviewer and for each grouping of concepts. SNOMED sample coverage was calculated for each sample grouping. Sample coverage was calculated by dividing the tabulated number of good scores in a sample grouping by the number of concepts in the sample group. The average SNOMED sample coverage was then determined. Averaged sample coverage was calculated by combining both individual reviewer's sample coverage scores and dividing by two. Projected tallies for the quality of match scores comprising all concepts in a grouping were determined. Projected good quality of match group tallies were calculated by multiplying the average sample coverage by the total number of concepts in a group. Lastly, the average SNOMED group coverage was derived by dividing the number of projected good scores in a grouping by the total number of concepts in the group. Chi-square test for non-association between quality categories and concept groups was made using the projected number of good and bad quality scores in the four concept groups.

Results

There were 468 unique concepts in the concept table storing the NORMEX knowledge. The tally of syntactical types of concepts comprising these 468 concepts is

[lxv] Microsoft. Microsoft excel 2000. 9.0.2720 ed: Microsoft Corporation; 1999
[lxvi] College of American Pathologists. SNOMED clinical terms. 2002-01Q ed: SNOMED International; 2002
[lxvii] Clinical Information Consultancy. CIC Look Up Engine (CLUE). 5.0.0047 ed: Clinical Information Consultancy; 2002
[lxviii] Wilcke JR. SNOMED CT Browser. 2.0 ed. Blacksburg: Drug Information Laboratory, Virginia Maryland Regional College of Veterinary Medicine; 2003

shown in Table 12. Systematic classification of concepts used in forming the four Groupings of concepts, as well as the total number of concepts in each of these Groupings, is shown in Table 12.

Individual reviewer scores for each group of concepts are shown in Table 12. Between-reviewer agreement for quality of mapping score matched for more than 80% of concepts and Kappa value exceeded 0.60 for all four groups (Table 12). The averaged coverage for Group 4 (D) was highest at 61.9%, with Group 1 (TOI) next at 37.5%, followed by Group 3 (F) at 16.7%, and Group 2 (L) had the fewest matches at 5.4% (Table 12). Averaged SNOMED group coverage for all four Groups was 30.4% (Table 12).

NORMEX Syntax	Num Con in Syntax Class	Grp ID	Num in Grp	Num in Sample	Quality Score Cat	Reviewer 1		Reviewer 2		Agreement			Average		
						Num in Sample	Cover % in Sample	Num in Sample	Cover % in Sample	Num Match in Sample	% Match in Sample	Kap	Cover % in Sample	Num in Grp	Cover % in Grp
T	1	1	8	8	1	5	37.5	5	37.5	8	100.0	1.000	37.5	5	37.5
O	3				2	3		3						3	
I	4														
L	130	2	130	30	1	29	3.3	28	6.7	29	96.7	0.649	5	124	5.4
					2	1		2						7	
F	120	3	120	30	1	24	20.0	26	13.3	28	93.3	0.760	16.7	100	16.7
					2	6		4						20	
D	210	4	210	30	1	12	60.0	11	63.3	25	83.3	0.648	61.7	80	61.9
					2	18		19						130	
Total / Average	**468**		**468**	**98**			**30.2**		**30.2**				**30.2**	**468**	**30.4**

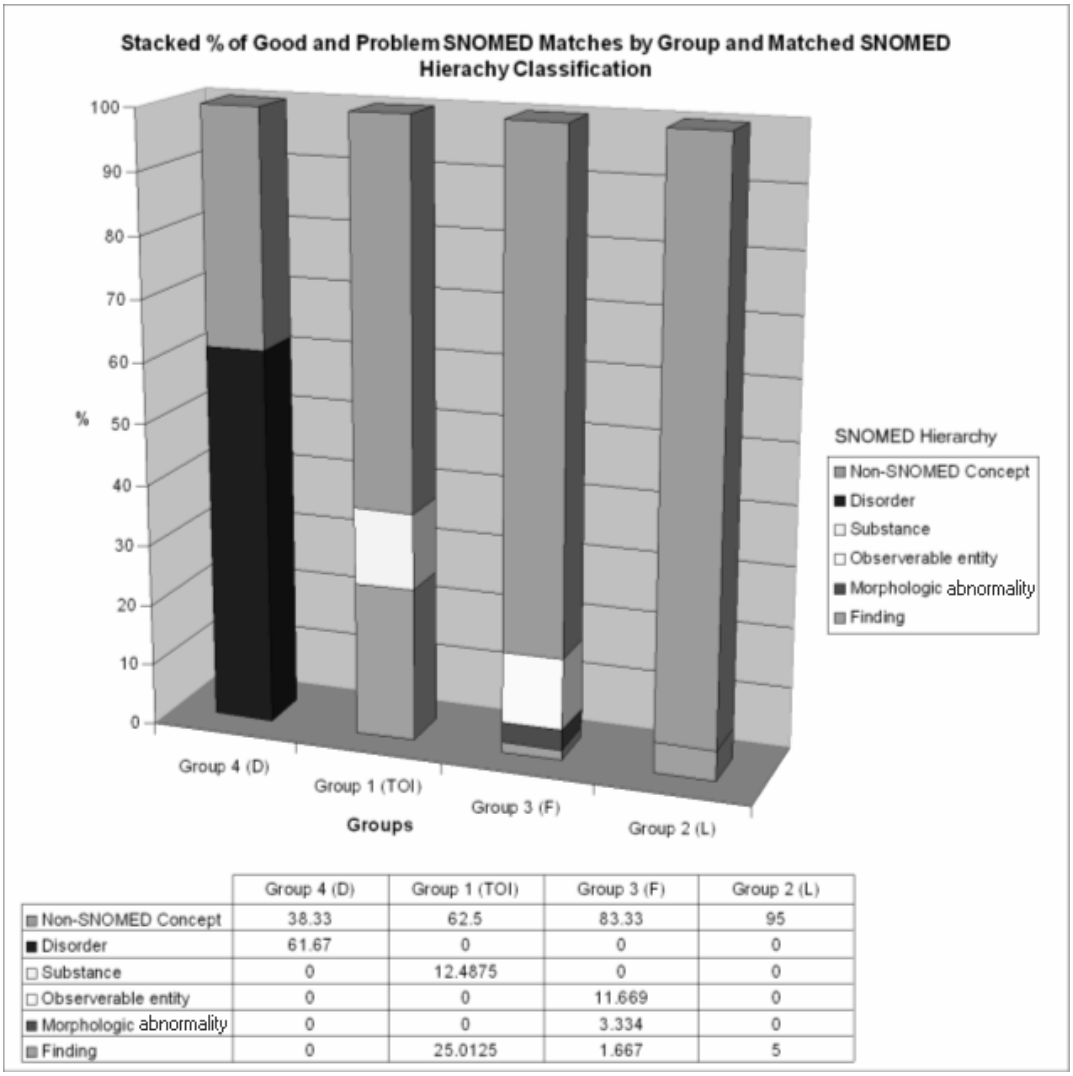
Table 12 NORMEX concepts quality score assignments, reviewer agreement and SNOMED coverage

Chi-square test for non-association P-Value of 0.000 strongly indicates an association between the quality score assignments and concept group categories (Table 13). A visible difference in the percentage of good matches (2) between group categories can be seen in Figure 12. SNOMED hierarchy location for concepts with good mapping scores is also shown in Figure 12. Group 4 (D) good SNOMED matching concepts came from the Disorder hierarchy. In Group 1 (TOI), most SNOMED source of good matching concepts came from the Findings hierarchy, with a smaller number from the Substance hierarchy. In Group 3 (F), three SNOMED hierarchies provided most of the matching concepts: Observable entity, Morphological abnormality, and Finding (in order of contribution significance). Lastly, within Group 2, all good matching SNOMED concepts were found to originate from the Finding TLC hierarchy.

Table 13 Chi Square test for non-association between Concept Groups and match score quality[lxix]

Score		Group 1 (T.O.I.)	Group 2 (L)	Group 3 (F)	Group 4 (D)	Total
1	Actual	5	124	100	80	309
	Expected	5.27	86.3	79.06	138.36	*
2	Actual	3	7	20	130	160
	Expected	2.73	44.69	40.94	71.64	*
Total		8	131	120	210	469
Chi-Sq	136.694					
DF	3					
P-Value	0.000					

[lxix] Score 1 = bad SNOMED match, score 2 = good SNOMED match



	Group 4 (D)	Group 1 (TOI)	Group 3 (F)	Group 2 (L)
Non-SNOMED Concept	38.33	62.5	83.33	95
Disorder	61.67	0	0	0
Substance	0	12.4875	0	0
Observerable entity	0	0	11.669	0
Morphologic abnormality	0	0	3.334	0
Finding	0	25.0125	1.667	5

Figure 12 distribution of good and bad NORMEX - SNOMED matches and SNOMED hierarchy source for good matched concepts[lxx]

[lxx] Organized by NORMEX syntax groups

Discussion

In the original modeling of the NORMEX knowledge used in this experiment, all concepts and relationships were created without reference to SNOMED.[8] However, steps were taken to align the syntax with SNOMED as it is currently organized. Following completion of the NORMEX model this review was undertaken to identify SNOMED concepts that might have been applicable. Such concepts would be a subset of SNOMED.[2] Necessary concepts that could not be located in SNOMED would exist as an extension of SNOMED.[2] Similarly, attributes that have no functional classification role in SNOMED might remain outside the nomenclature. Wide variability was found in SNOMED coverage of NORMEX concepts, ranging from 5-62% depending on NORMEX-syntactical classifications.

It was not surprising that this wide spectrum of coverage existed for the concepts from the various syntactic elements. SNOMED has had a long history of development from its initial days as Systematized Nomenclature of Pathology (SNOP) with its focus on pathologic findings.[19] With SNOMED's continued evolution into its current form, content expansion for SNOMED-CT has been an important design goal, in accordance with basic desiderata for terminology development.[20, 21] While this content expansion touches upon many areas of medicine, SNOMED is still best known for its ability to represent viewable clinically relevant concepts.

A clinical and disorder content focus was supported by the coverage findings, which showed that the concepts of the Disorder (D) syntactical class had the highest coverage (62%), followed by concepts from the Top, Observable Analyte, and Interpretive Finding (T.O.I) syntactical categories (37%). This latter percentage is lower than expected, but attributed to the low number of concepts in this category as a whole and the resulting disproportionate impact of failure to discover several of the few Interpretive syntactical element concepts used in the model. As anticipated, the syntactical categories that were least directly "clinical" in nature were much less well covered (Observable Functions 16% and Laboratory Finding 5%).

This distribution of concept coverage for the various elements of the NORMEX model indicates which portions of SNOMED most need additional content to support direct modeling of NORMEX pathways. While adding content to the specific hierarchies is one possible solution to meet NORMEX modeling needs, it is also reasonable to speculate that appropriate extension sets could be developed for this purpose.[2] Regardless of the approach taken to represent the missing content, efficient modeling of explanation pathways relevant to clinical pathology will require a moderate amount of additional SNOMED content (69% of the concepts used in the NORMEX model).

Acknowledgment

This work has been funded by the National Institutes of Health, National Library of Medicine Informatics Research Fellowship Award, grant number 1 F37 LM07699.

Bibliography

1. College of American Pathologists. SNOMED CT product profile: SNOMED International; 2003.
2. College of American Pathologists. SNOMED clinical terms consultation document: core structure: SNOMED International; 2000.
3. Kudla K. A new generation of SNOMED for computer-based patient records. Paper presented at: ACP News, 1999.
4. Kudla KM, Rallins MC. SNOMED: a controlled vocabulary for computer-based patient records. J Ahima 1998;69(5):40-4; quiz 45-6.
5. Rector AL. Terminology and concept representation languages: where are we? Artif Intell Med 1999;15(1):1-4.
6. Sowa JF. Knowledge representation: logical, philosophical, and computational foundations. Pacific Grove: Brooks/Cole; 2000.
7. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp 1997:640-4.
8. Zimmerman K. Normalized Medical Explanation (NORMEX) Syntax to Integrate Causal Reasoning with SNOMED [Blacksburg, VA: Virginia Tech; 2003.
9. Campbell JR, Payne TH. A comparison of four schemes for codification of problem lists. Proc Annu Symp Comput Appl Med Care 1994:201-5.
10. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. J Am Med Inf Assoc 1997;4(3):238-51.
11. Berman JJ, Moore GW, Donnelly WH, Massey JK, Craig B. A SNOMED analysis of three years' accessioned cases (40,124) of a surgical pathology department: implications for pathology-based demographic studies. Proc Annu Symp Comput Appl Med Care 1994:188-92.
12. Henry SB, Holzemer WL. Can SNOMED International represent patients' perceptions of health-related problems for the computer-based patient record? Proc Annu Symp Comput Appl Med Care 1994:184-7.
13. Lussier YA, Bourque M. Comparing SNOMED and ICPC retrieval accuracies using relational database models. Proc AMIA Annu Fall Symp 1997:514-8.
14. Sinha U, Yaghmai A, Dai B, Thompson L, Taira R, Dionisio J, et al. Evaluation of SNOMED3.5 in representing concepts in chest radiology reports: integration of a SNOMED mapper with a radiology reporting workstation. Proc AMIA Symp 2000;20(Suppl):799-803.
15. Zimmerman K. SNOMED Representation of explanatory knowledge in veterinary clinical pathology [Blacksburg, VA: Virginia Tech; 2003.
16. Minitab. Minitab statistical software. 13.32 ed: Minitab Inc; 2002.
17. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford; Boston: Blackwell Scientific Publications; 1994.
18. Baumann RP, Moret J. Weighted kappa statistics for measuring the divergence of rapid section and definitive diagnosis. Pathologe 1997;18(5):412-6.

19. Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
20. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(Nov(4-5)):394-403.
21. Chute CG, Elkin PL, Sherertz DD, Tuttle MS. Desiderata for a clinical terminology server. Proc AMIA Symp 1999:42-6.

Comparing Predictive Performance Of Naïve Versus NORMEX-Based Bayesian Network In Clinical Pathology

by
Kurt Lee Zimmerman

Virginia Polytechnic Institute & State University
Department of Biomedical Sciences and Pathobiology
540.231.3692
kzimmerm@vt.edu

Keywords: Bayesian Network, Expert Systems, NORMEX, SNOMED, Controlled Terminologies

Abstract

Bayesian-based expert systems have been used very successfully in medicine. While their usefulness and accuracy are recognized, their construction is still challenging. Common approaches in creating Bayesian networks for expert systems tend to compromise either the inclusion of potentially meaningful relationships or the pace of network construction. This work demonstrates the feasibility of extracting Normalized Medical Explanation (NORMEX) formatted knowledge from a storage architecture designed for medical nomenclature. Specifically, a Bayesian network was created to support analysis of veterinary pathology clinical chemistry results and associated pathophysiology.

Relevant concepts were derived from the text of pathology reports referencing pre-clinical studies of candidate pharmaceuticals given to laboratory rats. From this list of 97 user concepts, 22 were found to have semantically equivalent concepts in a set of previously constructed SNOMED-based explanation pathways. These 22 concepts were used to create a Naïve Bayesian network where the identified concepts each were directly linked to analyte test result evidence nodes. The same 22 concepts were located in 77 distinct explanation pathways and used to assemble a NORMEX-based Bayesian network. Six prediction nodes (6 of the 22 concepts identified) with the highest rate of positive occurrence in the training and testing datasets were identified for detailed evaluation. Network performance, defined by accuracy and area under the receiver-operating-characteristic (ROC) curve constructed from sensitivity and specificity data for the six prediction nodes, was determined and then compared for the NORMEX-based and Naïve Network.

Significant differences (alpha 0.05) suggesting superior network performance were seen for the following individual nodes of the pathway-based network: hepatic necrosis (ROC curve area – 74.6), hepatic inflammation (accuracy – 88.2%, ROC curve area – 76.3), altered blood hemostatic measurement (ROC curve area – 74.6. The overall network performance of the pathway-based network was significantly better (alpha of 0.10, roc curve area – 79.1). For the Naive network, superior performance was seen for skeletal muscle necrosis (accuracy – 85.7%).

This work demonstrates successful selective retrieval of explanatory knowledge stored in a SNOMED-like architecture for use to create a Bayesian network. Retrieval of the NORMEX knowledge was relatively easy and produced a network whose performance equaled or exceeded that of a Naïve model.

Table of Contents

List of Figures

List of Tables

Introduction

Medical expert systems such as Pathfinder, DXplain and Illiad have been used in medicine for some time.[2-8] These tools assist professionals in medical decision-making by bringing attention to significant findings or generating specific recommendations for actions. Bayesian-based expert systems, a subset of this class of tools, create a quantitative interaction between their stored knowledge and data describing particular clinical situations.[9, 10]

This subset of expert systems relies on Bayesian probability principles to analyze a particular scenario in light of existing knowledge and information.[11-15] Using extensions of the basic theorem, alerts and recommendations generated by Bayesian-based expert systems can have a very high degree of accuracy (60-96%) when compared with those of domain experts.[3-5, 16] Part of this success can be attributed to the importance that these types of systems give to the consideration of uncertainty, an inherent component of medical reasoning.[13, 14, 17, 18] The process of arriving at a specific diagnosis or plan for treatment is seldom black and white, a condition well recognized in the world of Bayesian probability theory.

A feature of Bayesian expert systems is that their knowledge can be represented graphically in the form of a Bayesian network. This network displays a collection of nodes (vertices, concepts) connected by arcs (edges), which from a simplistic standpoint implies a causal relationship between a prediction node (parent node at the tail of the arc) and the evidence node (child node at the head of the arc).[19, 20] Probabilistic inferencing (the process of arriving at new belief conditions -posterior probability- for prediction nodes based upon current evidence) is computationally possible due to three key features of the network: prior probabilities, (in)dependency between nodes, and a directed acyclic graph (DAG).[20-24] Prior probabilities may be either directly supplied by experts or learned from retrospective data sets.[25] These prior probabilities represent the likelihood for a given node state (prevalence rate) without the benefit of having specific pieces of evidence influencing the state of belief about the node. The establishment of clear (in)dependency between nodes was formalized by Pearl.[22] These (in)dependencies are depicted by the graph edges. Additionally, the graph edges provide the direction of causal influence within the directed graph. Efficient inferencing of posterior probabilities would not be possible without the assumption of conditional independency between nodes and the avoidance of directed cycles within the network.

An important aspect of a Bayesian network is the construction of the domain model. There are two general methods by which this is done: 1) directly by a user, 2) by use of a machine-learning approach.[8, 26] In user-created Bayesian networks, users may interact with a knowledge engineer whose job it is to extract the necessary causal relationships in the topic area and then display them graphically as concept nodes with appropriately connected arcs.[3, 15, 21] There are also Bayesian network development tools which the user can use directly to create the causal model for their domain of interest.[27-31]

Both of these approaches can result in a complex network capable of providing visual insight into cause-effect relationships within the topic area.

The second development approach to a Bayesian network is quite different. In this method, large collections of data containing outcomes, along with numerous other observable variables associated with that data, are explored with knowledge-discovery algorithms to find statistical relationships existing between observations and outcomes.[25, 26, 32] There are many variations of these knowledge-discovery (data-mining) algorithms and the reader is referred to several sources that provide an overview of these processes.[23, 29, 33, 34] Once the relationships among variables have been discovered, it is possible to automate their integration and assemblage into a Bayesian network.[25, 26, 29] Even more simply, the network can be constructed in a completely Naïve manner with all the possible outcomes being independently related to the pieces of evidence.[3, 29, 35, 36] Either variation of this general development approach can be very fast once the original dataset is prepared for data-mining and individual variables have been identified.

However, two non-overlapping problems are presented to developers of Bayesian-based expert systems using either of these development methods. In user-created networks, the chore of representing domain-expert causation knowledge can be quite costly.[3, 8] Machine-learning approaches can significantly reduce these costs but generally with a loss of understandable causation meaning within the network itself. This loss of understandability results in expert systems incapable of providing justification and supportive reasoning. This lack of supporting information is a large obstacle associated with user acceptance of such systems.[3] There is also work suggesting that expert knowledge can have a positive impact on cost and performance in using data-mining learning algorithms.[37]

Normalized Medical Explanation (NORMEX) is a syntax model developed for the expression and storage of explanatory knowledge using the storage format and representational syntax of SNOMED-CT.[1, 38] This experiment assessed the feasibility of semiautomatic creation of a Bayesian network specific to the needs of a group of users by using causation knowledge represented in the NORMEX model. Investigative pathologists at a pharmaceutical company were selected as the target group for whom a user-specific Bayesian network would be constructed, using a pre-existing set of concepts and relationships in NORMEX syntax. The target group had previously interpreted clinical pathology data generated in the investigation of new pharmaceutical compounds in terms of their biological safety and impact on biologic physiology. Concepts were harvested from their reports. The pre-existing knowledge base includes an exhaustive list of explanatory pathways for three serum chemistry analytes. It was hypothesized that a Bayesian network constructed by this approach would retain prediction performance similar to that of a naïve model, while gaining the advantages of greater understandability.

Methods

Data Preparation

Reports selected for this project included: 1) toxicological studies of 5-14 days duration 2) conducted with female, Fisher 344 rats, less then 90 days of age, 3) completed in the last 12 months 4) which contained a statistically significant alteration (alpha 0.05) in measured alanine aminotransferase activity in at least one animal of the study. Each report was an analysis of four dosage groups. Textual discussion of serum chemistry test results and actual test result values for total serum bilirubin (TBil), alanine aminotransferase (ALT), and alkaline phosphatase (ALP) were collected from reports matching the listed criteria.

The text of the selected reports was parsed, using a barrier-word technique, to identify a list of semantically unique concepts used in each report.[39-41] The list of concepts drawn from each report was then reviewed to eliminate concepts that were not interpreted findings. Specifically, these interpretive or explanatory concepts justified mention of particular facts or observations (e.g., "increased ALT activity" - a removed fact concept - versus "liver failure" – a retained interpretive concept).

Assigning Concepts to Appropriate Dosage Groups

Unique concepts for each report were individually identified by study and by dosage group. The data was recorded as for a cumulative toxicology report. That is, if a concept was recorded for a dosage group, it was also identified as being present in all animals that were members of higher dosage groups for that study. Each report was reviewed by hand to locate the no observable adverse effect (NOAEL) dosage level.[42] The NOAEL and lower dosage groups were disregarded in subsequent steps.

Mapping Report Concepts to Explanatory Pathway Concepts

The list of unique interpretive report concepts was hand-compared to the concepts table of the explanatory pathways extension set. Report concepts were considered to be matches for explanatory pathway concepts based on the semantic equivalence of the concepts. Concepts deemed appropriate for each group were assigned to each animal in the group.

Preparation of Training and Test Dataset

A spreadsheet file was created in which each row represented an individual animal from a study (Figure 13). For each animal the following data were recorded: sequential number ID, animal ID, study ID, values for the three analytes, and a true/false state for each explanatory concept. Results were recorded in sequence from control group through high dosage group for each study. A frequency of true state (indicating that the concept had been used to describe the animal) was calculated for each concept. Once all animals from all studies had been listed, the order of animals was randomized.

The randomized raw data spreadsheet was then divided into ten test datasets (*test n*, where n = 1-10), and ten training datasets (*train n*). Division of the randomized raw data file into ten equal portions, with each portion representing one of the *test n* files, created the test datasets (Table 14). For any one *test n* file the remaining portion of the raw data file represented the respective *train n* file (n of test = n of train).

Random ID	139	167	102	91	111	156	166	114	209	84	23
Case ID	70	71	72	73	74	75	76	77	78	79	80
Study ID	7	7	7	7	7	7	7	7	7	7	7
Group ID	1	2	2	2	2	3	3	3	3	4	4
Animal ID	1053	2051	2052	2053	2054	3051	3052	3053	3054	4051	4052
TBIL mg/dl	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	*	0.09
ALP u/l	160	252	371	188	260	198	209	181	193	*	197
ALT u/l	39	204	745	186	178	50	50	41	47	*	314
HEPATIC_FAILURE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
DEC_BLOOD_HEMO_MEASURE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
INC_HEP_CELLULAR_INFLAM	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
INC_HEP_CELLULAR_INJURY	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
INC_HEP_CELLULAR_NECROSIS	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
INC_SKM_CELLULAR_NECROSIS	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Figure 13 Example of true / false concept states for cases 70-80

Table 14 Schema for sorting true / false state data for *test n* and *train n* files

Run Number	Test n			Train n		
	n	Random ID	Number of Cases	n	Random ID	Number of Cases
1	1	1-24	24	1	25-235	211
2	2	25-48	24	2	1-24, 49-235	211
3	3	49-72	24	3	1-48, 73-235	211
4	4	73-96	23	4	1-72, 97-235	212
5	5	97-120	24	5	1-97, 121-235	211
6	6	121-144	24	6	1-120, 145-235	211
7	7	145-168	24	7	1-144, 169-235	211
8	8	169-192	24	8	1-168, 193-235	211
9	9	193-216	24	9	1-192, 217-235	211
10	10	212-236	25	10	1-211	210

Identifying Relevant Explanatory Pathways

An application was created[lxxi] to assist with this portion of the project and was used to perform the following processes.[1] All explanatory pathways implied by the relationships and concepts tables of the explanatory pathways extension set were created.[1] Reassembled pathways containing at least one explanatory concept were selected. Pathways were truncated to the first occurrence of a mapped concept (Figure 14). Removing shorter pathways that were subparts of, or implied by, longer pathways further reduced the total number of pathways selected. Resulting selected explanatory pathways were stored as strings in a database table[lxxii].

[lxxi] Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000
[lxxii] Microsoft. Microsoft access 2000. 9.0.2720 ed: Microsoft Corporation; 1999

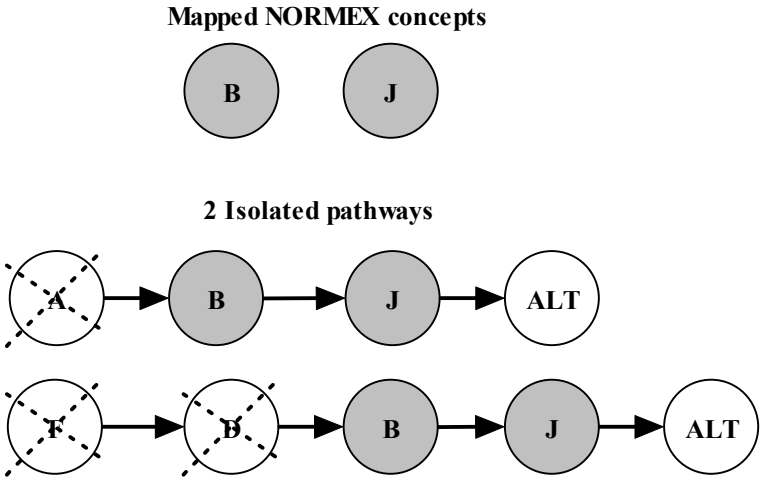


Figure 14 Example of pathway truncation

Creation of Bayesian Networks

Naive and NORMEX-based versions of a Bayesian network were created, using commercial Bayesian network design software[lxxiii] loaded with the mapped explanatory pathway concepts from the *report pathway mapping* spreadsheet. Imported nodes were assigned *node type* (discrete or continuous), *range of values*, and *state* associated with range values.[43] TBil, ALT, and ALP were created as continuous type for both networks. Each of these nodes was created as having three states associated with a range of result measurements: low (TBil, 0-0.01; ALT, 0-313; ALP, 0-28), target (Tbil, 0.01-0.1; ALT, 313-456; ALP, 28-59), high (TBil, 0.01-infinity; ALT, 456-infinity; ALP, 59-infinity). Explanatory concept prediction nodes were created as discrete types having two value states (True / False) in both networks. Prediction nodes (explanatory concepts) were linked to the evidence nodes (analytes) as parents. However, the networks differed in the method used to place arcs between nodes. For the Naive network, each prediction node was connected directly to each of the three analyte concepts. For the NORMEX-based network, the selected explanatory pathways, with their implied causal attributes, were used to draw the appropriate linking arcs. The Bayesian network design software then "summed out" or "absorbed" from the built networks those nodes with a true state frequency of occurrence in the lower 75 percentile.[43-45]

Network Evaluation

Each network was trained and tested by the same methodology. Ten runs were conducted on each network, with one run consisting of training with the $training_n$ dataset and then testing with the respective $test_n$ dataset. The Bayesian network software used the *test n* file to output a probability prediction result file for the likelihood of a true state condition for each prediction node for each animal in the *test n* file. Thresholds for considering probability predictions as being true were established at integer values of 0-100 (interval 1). The sensitivity, specificity, and accuracy at each threshold value for the *test n* file were calculated using the predicted probabilities provided by the network, the known status of the particular query node for the *test n* animals, and results stored in *result n* file. Sensitivity and 1-specificity information from each *result n* file were used to create a receiver operating characteristic (ROC) curve and determination of area under the ROC curve (trapezoid method) for each prediction node.[13, 29, 46, 47]

A commercial statistical software package[lxxiv] was used to compare the performance of the NORMEX-based Bayesian network versus the Naive Bayesian.[48] The ten values of accuracy and area from each network were checked for normality using Anderson-Darling method.[48, 49] A paired t-test was used for normally distributed data and

[lxxiii] Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002
[lxxiv] Minitab. Minitab statistical software. 13.32 ed: Minitab Inc; 2002

the Mann-Whitney test for data that failed a normality test.[49] All tests were conducted at an alpha value of 0.05.

Results

Concept and Explanatory Pathway Isolation

Sixteen pathology reports meeting the established selection criteria contained information regarding 236 animals. Isolation of interpretive concepts from the sixteen studies identified a total of 259 lexically unique interpretive concepts (Table 15). Of these 259 concepts, 97 were semantically unique. Of these 97 concepts, 22 were mapped as semantically equivalent to concepts in the NORMEX extension set (Table 16).

Table 15 Example of isolated concepts from study 1

Study Number	Concept
1	Endothelium vacuolation
1	Iris vacuolation
1	Alveolar histiocyte vacuolation
1	Muscle necrosis
1	Muscle injury
1	Malabsorption
1	Phospholipidosis
1	Stress leukogram
1	SKM vacuolation
1	Acinar cell vacuolation

Table 16 Example of mapping between report and pathway concepts

Report Concept ID	Unique Report Concept	Mapped Explanation Pathway Concept
16	Brown fat atrophy	NULL
17	Central nervous system toxicity	NULL
18	Cholestasis	DECREASED BILE PASSAGE BILE FLOW (L)
19	Decrease food consumption	DECREASED ORGANISM FOOD INTAKE (D)
20	Decreased erythroid parameters	DECREASED BLOOD RED BLOOD CELL MASS (L)
21	Decreased erythropoiesis	DECREASED MARROW ERYTHROPOIESIS PROCESS (L)
22	Decreased gastrointestinal integrity	NULL
23	Decreased hepatic protein production	HEPATIC FAILURE (D)
24	Decreased platelet production	NULL
25	Decreased renal concentrating capacity	NULL
26	Decreased renal function	DECREASED RENAL FUNCTION MEASUREMENT (L)

Table 17 shows the frequency of mapped explanatory concepts. The six concepts with the highest positive occurrence frequency were: increased hepatic cellular injury, increased hepatic cellular necrosis, decreased blood hemostasis measurement, increased skeletal muscle cell necrosis, hepatic failure, and increased hepatic inflammation. Concept names were abbreviated to accommodate node-naming restrictions imposed by the Bayesian design software.

Table 17 Frequency of explanatory concept occurrence[lxxv]

Concept	Abbreviated Network Node Name	Frequency
Increased Hepatic Cellular Injury	INC_HEP_CELLULAR_INJURY	45
Increased Hepatic Cellular Necrosis	INC_HEP_CELLULAR_NECROSIS	41
Decreased Blood Hemostasis Measurement	DEC_BLOOD_HEMO_MEASURE	27
Increased Skeletal Muscle Cellular Necrosis	INC_SKM_CELLULAR_NECROSIS	21
Hepatic Failure	HEPATIC_FAILURE	19
Increased Hepatic Inflammation	INC_HEP_CELLULAR_INFLAM	19
Decreased Organism Food Intake	DEC_ORG_FOOD_INTAKE	15
Cholestasis	CHOLESTASIS	13
Decreased Renal Function Measurement	DEC_RENAL_FUN_MEASUREMENT	13
Decreased Hepatocellular Membrane Integrity	DEC_HEP_CELL_MEM_INTEGRITY	12
Increased Bone Architecture Remodeling	INC_BONE_ARCH_REMODELING	9
Increased Bone Osteobast Activity	INC_BONE_OSTEOBLAST_ACTIVITY	8
Increased Skeletal Muscle Cellular Injury	INC_SKM_CELLULAR_INJURY	8
Decreased Marrow Erythropoiesis Productivity	DEC_MAR_ERYTHROPOIES_PROC	5
Decreased Skeletal Muscle Cellular Membrane Integrity	DEC_SKM_CELL_MEM_INTEGRITY	5
Decreased Marrow Erythroid Numbers	DEC_MARROW_ERYTHROID_NUM	4
Decreased Blood Rbc Mass	DEC_BLOOD_RBC_MASS	3
Gastrointestinal Ulcer	GI_ULCER	3
Hyperbilirubinemia	HEYPERBILIRUBINEMIA	3
Increased Marrow Cytotoxic Insult	INC_MAR_CYTOTOXIC_INSULT	3
Increased Hepatocellular Apoptosis	INC_HEP_APOPTOSIS	0
Bile Passage Trauma	BILE_PASSAGE_TRAUMA	0

[lxxv] Non-shaded concepts = top quartile. Abbreviate concept name need to accommodate naming conventions used in Bayesian design software

The Visual Basic application created 3561 explanatory pathways in the LOIT format, using the explanatory pathway extension set tables.[1] Of these pathways, 3490 (98%) contained one or more of the 23 explanatory pathway concepts listed in Table 4. The removal of shorter pathways implied by longer pathways left a total of 77 pathways (Table 18 Example of one of the 77 isolated pathways).

Table 18 Example of one of the 77 isolated pathways[lxxvi]

Explanatory Pathway
TOXOPLASMOSIS (D)<>INCREASED HEPATIC CELLULAR INFLAMMATION (L)<>INCREASED HEPATIC FREE RADICAL CONCENTRATION (L)<>INCREASED HEPATIC CELLULAR INJURY (L)<>INCREASED HEPATIC CELLULAR NECROSIS (L)<>DECREASED HEPATIC CELL NUMBERS (L)<>INCREASED BLOOD UNCONJUGATED BILIRUBIN CONCENTRATION (L)<>HYPERBILIRUBINEMIA (I)<>TOTAL SERUM BILIRUBIN CONCENTRATION (O)<>ANALYTE (T)

[lxxvi] Concepts classified syntactically as Observable Functions have been removed for the displayed example. This removal is discussed in the discussion section of this paper.

Network Creation

Results for Naïve and NORMEX-based networks before and after node absorption are shown in Figure 15 and Figure 16.

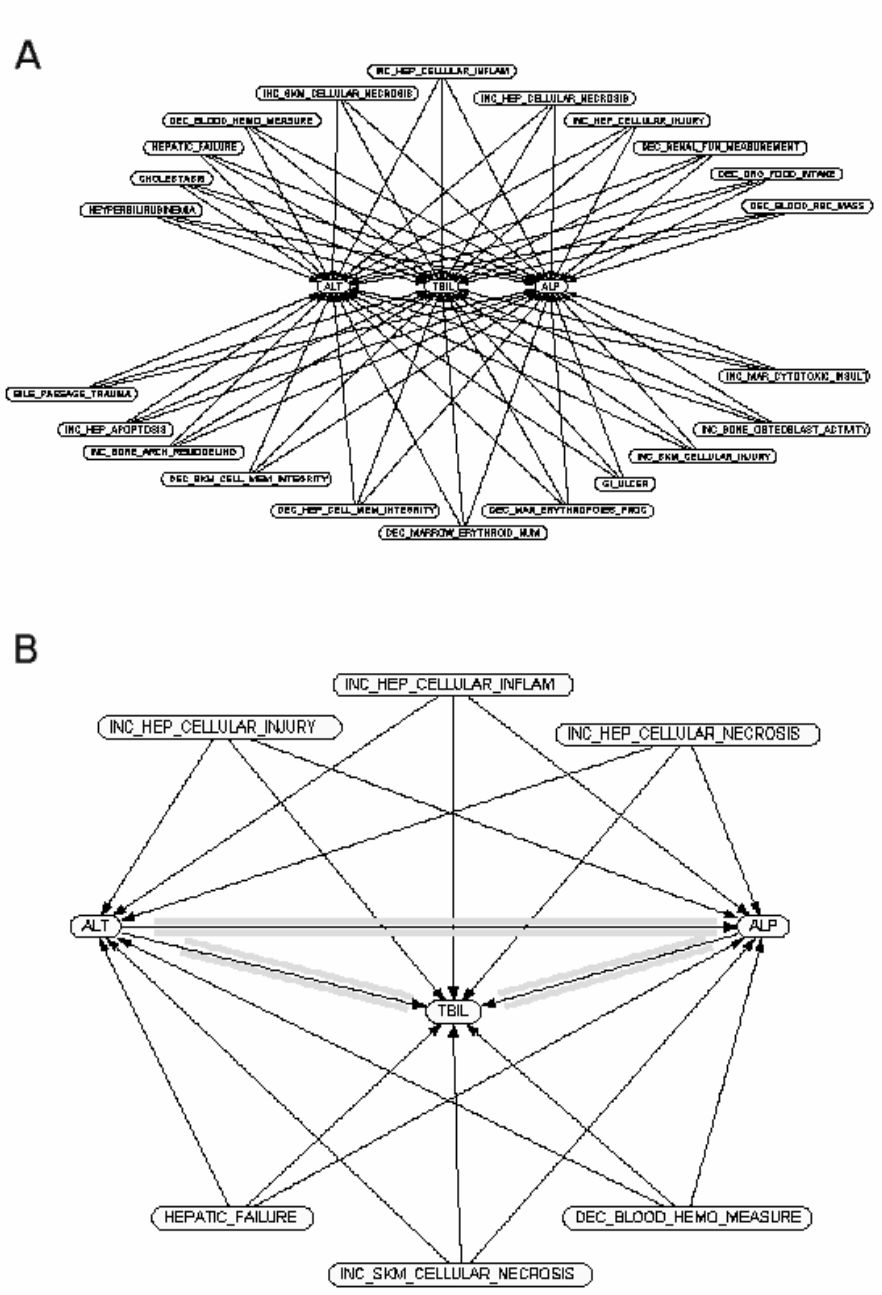


Figure 15 A) Naive network prior to node absorbing. B) Network after node absorbing[lxxvii]

[lxxvii] New arcs highlighted

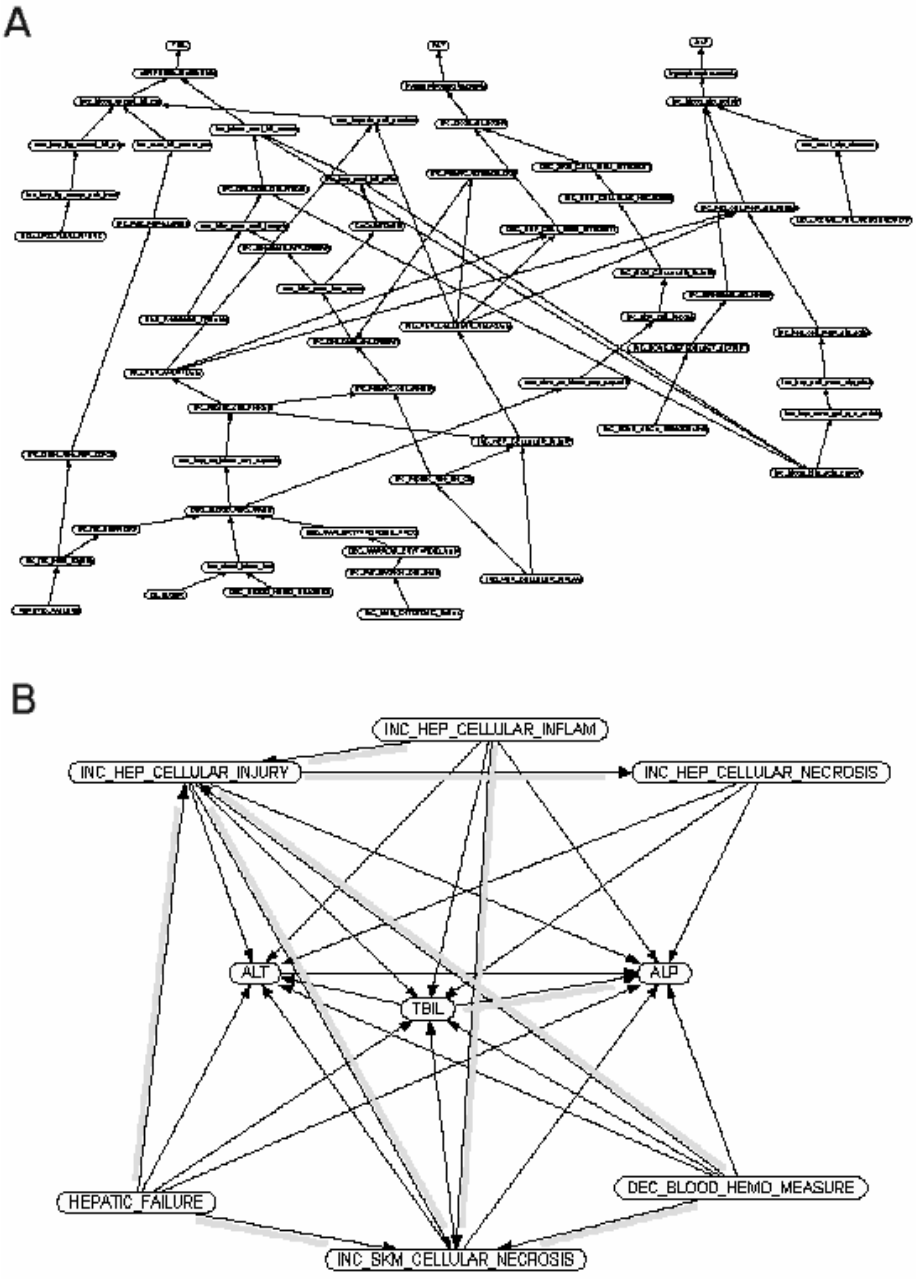


Figure 16 (A) NORMEX-based network prior to node absorbing. (B) Network after node absorbing[lxxviii]

[lxxviii] Arcs differing from Naive network highlighted

Network Evaluation

Predictions for the probability of a true state for each of the six explanatory-concept nodes were calculated for both networks. Results for individual concept nodes and cumulative performance of each network were compared as shown in Table 19. Significant differences were tested at alpha = 0.05. Network performance could be judged superior on the basis of accuracy, ROC curve area or both. Superior performance was seen for the following explanatory-concept nodes of the NORMEX-based network: hepatic necrosis (ROC), hepatic inflammation (accuracy & ROC), altered blood hemostatic measurement (ROC). For the Naive network, superior performance was seen for skeletal muscle necrosis (accuracy). In addition, the overall network performance of the NORMEX-based Bayesian network, as judged by the average ROC curve area calculated from consideration of all six concept nodes, was significantly better at an alpha of 0.10 (ROC curve area – 79.1). The combined curve of the entire network is shown in Figure 17.

Table 19 Summary of comparison of accuracy and ROC curve area between NORMEX and Naive network[lxxix]

Significantly different at p-value 0.05	Mean Accuracy %		Mean ROC Curve Area	
Significantly different at p-value 0.10	**Naive**	**NORMEX**	**Naive**	**NORMEX**
Altered Blood Hemostatic Measurement	81.6%	82.0%	71.4	**74.6**
Hepatic Failure	87.4%	87.1%	86.5	90.0
Hepatic Injury	80.2%	81.0%	89.8	89.8
Hepatic Necrosis	73.8%	73.8%	57.3	**70.9**
Hepatic Inflammation	87.9%	**88.2%**	72.4	**76.3**
Skeletal Muscle Necrosis	**85.7%**	84.7%	72.8	73.1
Mean	82.7%	82.8%	75.0	**79.1**

[lxxix] Light grey items significantly different at alpha 0.05; dark grey items significantly different at alpha 0.10; items in bold are of significantly greater value

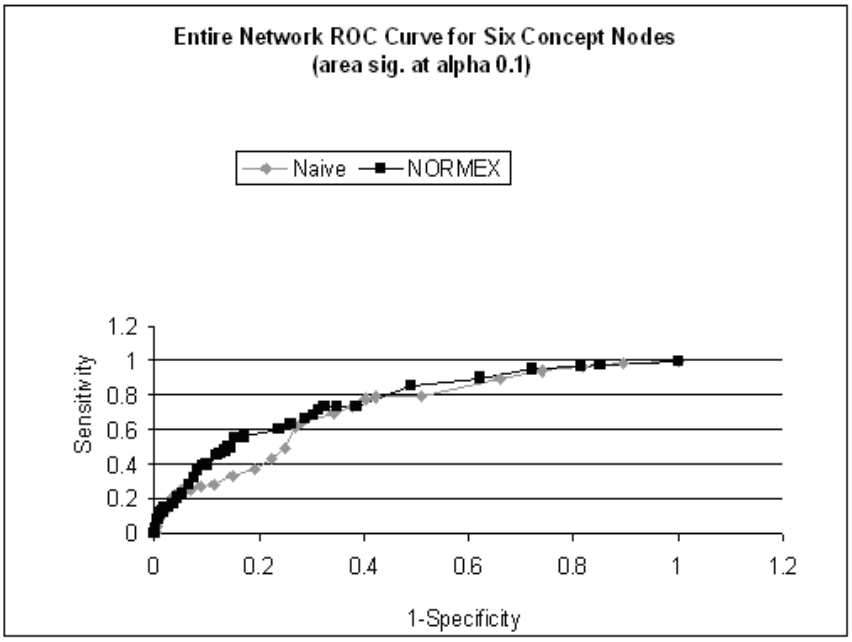


Figure 17 NORMEX and Naive ROC curve graphs for the six individual explanatory pathway concept nodes and for the combined entire network of six nodes

Discussion

This paper demonstrates that user-specific expert knowledge (NORMEX format) can be extracted from a purpose-built extension of SNOMED and used to assemble a Bayesian network. The prediction performance associated with the NORMEX-based network equaled or exceeded that of the Naïve version when assembled using a Bayesian network design tool.

The NORMEX model has been used to represent explanatory knowledge within the field of clinical pathology.[1] Further, this model of knowledge can then be stored in an object-attribute-value format compatible with the storage format used by the representational terminology SNOMED.[1] SNOMED provides for aggregating related concepts in complex searches and retrieval operations.[50-54] This power is similarly conferred to extension sets adhering to SNOMED logical design and physical storage formats. Therefore, explanatory knowledge modeled as an extension of SNOMED is selectively retrievable based upon individual user needs. Selective retrieval, combined with the cause-and-effect knowledge conveyed through the NORMEX syntax, is a powerful resource for the construction of Bayesian networks capable of addressing user-specific needs.

Traditionally, user-relevant expert knowledge has been captured by a knowledge engineer.[3, 8] Extensive interviews and time are necessary with this approach as the knowledge engineer captures expert knowledge about a topic area.[6, 8, 30] The knowledge engineer then represents the knowledge in a formalized representation format. This work illustrates another approach to the knowledge-modeling process. Here the expert knowledge was already expressed in a standardized manner (NORMEX) and the task became one of identifying those portions of the knowledge relevant to the target user group. In this work, relevant user concepts were extracted from the findings and the discussion sections of pathology reports. Each report was produced by the efforts of several pathologists, and reports were created from studies associated with a variety investigative compounds. This variety helped to assure that the reports contained a wide range of concepts. A barrier-word technique was used to parse the document sections and to isolate individual concepts.[14, 39, 40] The type of knowledge these concepts was intended to convey was examined in an earlier manuscript.[55] Once isolated, individual concepts were related to semantically equivalent standardized versions of the concepts within the NORMEX extension set. Although this approach was moderately labor-intensive, reports stored in a controlled terminology would have been easily processed and automated applications currently exist to map text phrases to medical nomenclatures.[39, 56-63]

The relatively low number (23) of NORMEX extension-set concepts mapped to the isolated user concepts (97) was not surprising. The current NORMEX model represents a very small portion of the explanatory knowledge associated with clinical pathology. Specifically, the model is limited to explanation pathways dealing with three

serum analytes (ALT, ALP, and TBil). In the pathology reports, the finding and discussion sections reported on results from a serum chemistry panel consisting of 12 analytes, along with hemogram results. Concepts associated with explanatory knowledge related to these other test results were not present within the NORMEX extension set, accounting for the low positive mapping rate.

An algorithm described in a previous paper was used to recreate the explanatory pathways of knowledge from the NORMEX extension set.[1] Concepts within these reconstructed pathways were singularly connected, had a single occurrence in an individual pathway, and had multiple occurrences within the entire tree.[20] Reflecting the architecture of the originally modeled causal-dependency hierarchies, the recreated polytree architecture was composed of linear and diverging d-separated trails of concepts.[20, 22]

As implied in an earlier manuscript, the NORMEX syntax is bi-directional.[1] However, to be consistent with representations in Bayesian networks (implied causality) storage and processing order was important to the final result. The polytree structure resulting from use of the LOIT pathway format was best suited for creating a network to answer questions of interest to the targeted users. In a Bayesian network, two general classes of nodes exist, one set referred to as the prediction (also known as the hypothesis or root nodes) and the other set being the evidence nodes, with the tails of arcs depicting causation extending from the predictor to the evidence nodes.[20, 21, 25] These two categories of node types have relevance to the concepts identified in the pathology reports. The explanatory or interpretive concepts from the reports were found in the discussion section of the document. These concepts represented the conclusions being drawn from the observations of the findings sections. These conclusions correlate with the types of concepts at the more proximal end of the NORMEX represented in the LOIT format. In terms of a Bayesian network, these conclusion concepts correlate with prediction nodes within the network with arcs extending to the evidence nodes. The evidence nodes correlate with the finding sections concepts, as expressed in more distal NORMEX (LOIT format) concepts.

It should be pointed out that expressing the pathways in the TOIL format for use within a Bayesian network could also allow one to determine the likelihood of occurrence of what would then be considered the evidence nodes for given state values of root nodes. To accomplish this, the network's arcs would have to be reversed, converting the evidence nodes to the prediction nodes. This transformation would have required additional computing resources and likely would have resulted in a network of increased complexity in terms of arc connections.[43, 64] Since the algorithm for recreating the explanatory pathways could express the causation knowledge in either the TOIL or LOIT, format there was no reason to incur the additional processing cost associated with the TOIL version.

An algorithm identified those pathways (3,490) in which one of the 23 mapped NORMEX concepts occurred. While the pathways represented d-separated trails, it was still possible to remove from consideration those pathways expressing causal information already contained in other pathways. This greatly reduced the number of pathways (77) needed to construct the NORMEX Bayesian network. Removing concepts occurring prior to the first occurrence of a mapped NORMEX concept shortened these remaining pathways. If these removed concepts had been allowed to remain in the pathway, they would have been represented as prediction nodes within the Bayesian network. Since there were neither any prior probabilities for these concepts, nor any expressed interest of the user group in terms of predicting the occurrence of these concepts, they were removed prior to adding the pathway to the network.

Function concepts were fundamental to the creation of the original pathways in the NORMEX syntax. For this application, however, Laboratory Finding concepts adequately capture the state consequence and its anatomic location of physiologic functions. Therefore, expressions of functions in an unperturbed state became meaningless. Function concepts were removed from explanation pathways.[1]

Bayesian networks have been created using two different approaches. Either the network is based on domain knowledge of human experts, or data mining techniques can isolate variables and relationships within large datasets.[23, 25, 29] These two approaches will produce networks with distinctly different appearances. Bayesian networks produced by data mining may not provide physiologic explanations for the connected nodes. Other investigators have shown that justification of explanations is a very important factor in user acceptance of expert systems.[6, 30] Thus, while the use of data-mining principles for creating a Bayesian-based expert system is relatively easy, understanding the implications of the relationships in the model can be very difficult. Similarly, Naïve Bayesian networks are easy to create but fail to provide meaningful justification for network relationships. In spite of this lack of justification of network arcs, Naïve and data-mining networks have been shown to have quite good prediction performance.[9, 65-67] This knowledge is the basis for the selection of a Naïve Bayesian network as the gold standard to which the NORMEX-based network would be compared.

Once identified, the NORMEX pathways were assembled into the NORMEX-based Bayesian network. Because concepts were used more than once throughout the original NORMEX-based causal-dependency hierarchy, polyconnected concepts were present in the resulting network.[20] There were also converging and diverging pathways, along with a few undirected cycles, but there were no directed cycles in the network. Directed cycles were prevented by enforcement of relative logical ordering of concepts in the originally modeled NORMEX knowledge.[1] This helped to minimize cycles as a whole and assured creation of a DAG when the pathways were assembled. The DAG structure prevented the development of unmanageable complexity associated with inferencing calculations in the network.

Nodes representing the NORMEX concepts within the network needed to be discretized prior to using the network for making predictions.[29] The discretizing method was simple for the prediction nodes. For these nodes, two states were defined, true or false. The choice of a discretizing method for the evidence nodes (analyte results) was more difficult. Many discretizing methods exist: unsupervised-equal interval binning, equal frequency; supervised-minimum descriptive length principle, and error-based, as a few examples.[23, 29] The discretizing choice has ramifications for a network's prediction performance. An entirely new project could be created around the task of finding the optimal discretizing method for the NORMEX-based network. Since the goal of this work was to compare the performance of the NORMEX-based network with that of the Naïve model, a critical comparison was possible as long as similar discretizing methods were applied to both networks. In this work, evidence nodes were discretized into three categories: below the reference range, within the reference range, and above the reference range. Numeric values used to establish the three intervals categories for each evidence node (analyte) were obtained from the user group's established laboratory reference intervals.

Following network assembly and node discretization, one final manipulation was performed on the networks. The frequency distribution for true/ false states of the prediction concepts from the networks were examined in the prepared dataset comprising evidence and prediction concepts. All the prediction concepts appearing in the Naïve model were present as prediction concepts in the dataset. This was not the case in the NORMEX network. Here, NORMEX concepts appeared as part of a causal pathway but were not included as prediction concepts in the dataset. These novel concepts were implied by the NORMEX model itself, not derived from the reports. These additional concepts provided much of the justification explaining how the prediction concepts related to the evidence concepts, which represented a targeted goal of this work. However, in terms of training the network and making predictions, there was no historic data in the dataset to provide prior probabilities for these novel concept nodes. Therefore, while their presence influenced the creation of arcs, no evidence existed for their state to allow for their direct influence on the inference of posterior probabilities in the network. The software used to construct the networks permitted the use of a common method referred to as "summing out a variable" or "node absorption" to hide such nodes.[43-45] While this summing-out visually removed such nodes from the network, the overall network conditional probability distribution implied by the causal relationships of the NORMEX pathways remained unchanged. This node absorption process transformed the network.[43] Network transformations are manifested by the removal and or creation of new arcs as necessary to maintain the same overall network conditional probabilities. It was because of the causal relationships (arcs) supplied by the NORMEX pathways that the transformed NORMEX Bayesian network had a different arrangement of arcs as compared to the Naïve network. It should be emphasized again that even though these nodes were visually removed, the overall prediction model for the network remained unchanged, and the absorbed nodes could have been returned to the network as easily as they were removed.

The frequency of distribution of concepts was not only used to identify predictive network concepts not present in the training dataset, but also to identify those concepts with significant differences in their occurrence frequency. It is desirable, during the training of a network, that events being predicted occur with equal frequency. Lack of similarity in occurrence frequency leads to bias in the prediction results when new cases are encountered.[29, 32] Those concepts with lower occurrence rates create negative prediction bias, while those with higher occurrence rates engender positive prediction bias. In examining the testing and training dataset, far more False cases than True cases were seen. This distribution establishes an overall negative bias in prediction performance. For this reason, only concepts in the top twenty-fifth percentile of True state occurrence were used for training purposes. While the distribution of occurrences for the top six identified concepts was not equal, this reduced set of concepts with higher frequency of occurrence contained reasonable numbers of positive cases to help counter the overall negative bias of the dataset in the prediction of new outcome events.

Evaluation of the area under the ROC curve created from sensitivity and 1-specify test result data is a standard method for comparing performance between two tests.[13, 29, 47, 67] The predictions provided by the two networks represented predictions of likelihood or strengths of belief that an event had occurred, based on the given evidence. These strengths of belief, expressed as probabilities, had to be converted into actual predictions for the state of the prediction nodes. Probability threshold above which a prediction would be considered true could be set at any point between 0 and 1. Since the ideal threshold for maximum test performance was not initially known, sensitivity and specificity results were calculated for 101 evenly divided segments between the values of 0 and 1 for each test case within a given run, and then averaged for the entire run. These averaged run results were then themselves averaged for the 10 runs.

The assessment of sensitivity and specificity over a wide range of threshold settings permitted graphical visualization of where the networks differed in terms of their areas under the ROC curves. ROC curves of the individual nodes illustrate that most of the test-performance gains realized by the NORMEX network occurred at the higher probability threshold settings (high specificity), where sensitivity for the NORMEX network out-performed that of the Naïve network. The overall high average specificity of the network (91%) can best be leveraged as a screening test for nodes expected to have lower true state frequency rates in the data set, thus maximizing the negative predictive power of the NORMEX network for these types of nodes.[13, 68, 69] Obtaining a dose-response pattern in the chemistry data was an important goal when compound dosing schedules were created for the experiments from which the pathology reports were created.[42] With a control group and a low, middle, and high dose groups, a majority of the animals would ideally have failed to exhibit dosing compound effects classified as exceeding the NOAEL. Therefore, in this domain, numerous prediction nodes will have a low frequency of true state occurrence, suggesting that the NORMEX network could serve as an important tool for screening adverse compound effects.

The sensitivity of both the NORMEX and Naïve networks likely suffered as a result of using the NOAEL to establish the presence of a TRUE / FALSE state of a concept condition. Individual animals in a dosage group below the NOAEL, with significant alteration of chemistry values were classified as normal, creating false-negative cases in the dataset. It is likely that if the network performance had been assessed on the state of individual animals, ignoring the implied NOAEL state, test sensitivity would have been better than seen in this experiment.

It was interesting that only the accuracy prediction associated with skeletal muscle necrosis (SKMN) was found to be greater in the Naïve as compared to the NORMEX network (Table 8). Again, the area under the ROC curve for SKMN was considered to be a better evaluation of the two networks and no statistical difference was noted in this comparison. Explanatory knowledge in the NORMEX model was focused on explaining laboratory alterations associate with the liver. There is some overlap in the analyte markers used in evaluating liver disease and skeletal muscle disease.[68, 69] Therefore, some explanatory information was presented that dealt with SKMN. However, several analytes (creatine kinase, aspartame transaminase) associated with SKMN were not part of the NORMEX knowedge.[68] The prediction accuracy for SKMN might have been improved if the NORMEX model had included explanatory knowledge associated with these two additional analytes.

Conclusion

This work demonstrates a method to selectively retrieve explanatory knowledge stored in a SNOMED-like format for use in a Bayesian network. Incorporation of the knowledge stored in NORMEX-based pathways produced a network whose performance equaled or exceeded that of the Naïve model. Naïve Bayesian networks and those derived from data-mining approaches are relatively quick and easy to assemble and have been shown to have good prediction accuracy. Drawing on the explanatory knowledge contained in the pathways, user-relevant explanatory knowledge was retrieved and made available to a Bayesian network without loss of performance.

This work justifies further investigations of a formal relationship between controlled terminologies and Bayesian based expert systems. Using the NORMEX knowledge model extended the capacity of SNOMED to provide explanatory knowledge for the creation of Bayesian-based expert systems.

Acknowledgment

This work has been funded by the National Institutes of Health, National Library of Medicine Informatics Research Fellowship Award, grant number 1 F37 LM07699 and Lilly Research Laboratories, Greenfield, IN.

Bibliography

1. Zimmerman K. Normalized Medical Explanation (NORMEX) Syntax to Integrate Causal Reasoning with SNOMED [Blacksburg, VA: Virginia Tech; 2003.
2. Diez FJ, Mira J, Iturralde E, Zubillaga S. DIAVAL, a Bayesian expert system for echocardiography. Artif Intell Med 1997;10(1):59-73.
3. Berner ES. Clinical decision support systems: theory and practice. New York: Springer; 1999.
4. Heckerman DE, Horvitz EJ, Nathwani BN. Toward normative expert systems: Part I. The Pathfinder project. Methods Inf Med 1992;31(2):90-105.
5. Heckerman DE, Nathwani BN. Toward normative expert systems: Part II. Probability-based representations for efficient knowledge acquisition and inference. Methods Inf Med 1992;31(2):106-16.
6. Munsen MA, Shahar Y, Shortliffe EH. Clinical Decision support systems. In: Shortliffe EH, editor. Medical informatics. 2nd ed. New York: Sringer-Verlag; 2001. p. 573-609.
7. Tamai S. Expert systems and automatic diagnostic systems in histopathology--a review. Rinsho Byori 1999;47(2):126-31.
8. Warner HR, Sorenson DK, Bouhaddou O. Knowledge engineering in health informatics. New York: Springer; 1997.
9. Spiegelhalter D, Dawld A, Lauritzen S, Cowell R. Bayesian analysis in expert systems. Statistical Science 1993;8(5):219-83.
10. Andreassen S, Riekehr C, Kristensen B, Schonheyder HC, Leibovici L. Using probabilistic and decision-theoretic methods in treatment and prognosis modeling. Artif Intell Med 1999;15(2):121-34.
11. Parmigiani G. Modeling in medical decision making - a Bayesian approach. West Sussex, England: John Wiley & Sons, Ltd; 2002.
12. Morgan MG, Henrion M, Small M. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge; New York: Cambridge University Press; 1990.
13. Sox HC. Medical decision making. Boston: Butterworths; 1988.
14. Ward CD. Meditations on uncertainty. In: Kohout L, Bandler W, editors. Knowledge representation in medicine and clinical behavioral science. Cambridge, Mass.: Abacus Press; 1986. p. 19-25.
15. Hamilton PW, Montironi R, Abmayr W, Bibbo M, Anderson N, Thompson D, et al. Clinical applications of Bayesian belief networks in pathology. Pathologica 1995;87(3):237-45.
16. Nathwani BN, Clarke K, Lincoln T, Berard C, Taylor C, Ng KC, et al. Evaluation of an expert system on lymph node pathology. Hum Pathol 1997;28(9):1097-110.
17. Baxt WG. Complexity, chaos and human physiology: the justification for non-linear neural computational analysis. Caner Lett 1994;77:85-93.
18. Owens DK, Sox HC. Medical decision making: probabilistic medical reasoning. In: Shortliffe EH, editor. Medical informatics. 2nd ed. New York: Sringer-Verlag; 2001. p. 76-131.

19. Bartels PH, Thompson D, Bibbo M, Weber JE. Bayesian belief networks in quantitative histopathology. Anal Quant Cytol Histol 1992;14(6):459-73.
20. Charniak E. Bayesian network without tears. AI Magazine 1991;Winter:50-63.
21. Montironi R, Whimster W, Collan Y, Hammilton P, Thompson D, Bartels P. How to develop and use a Bayesian belief network. J Clin Pathol 1996;49:194-201.
22. Geiger D, Verma T, Pearl J. Identifying independence in Bayesian networks. Networks 1990;20:507-34.
23. Han J, kamber M. Data Mining: concepts and techniques. San Francisco, CA: Moragn Kaufmann Publishers; 2001.
24. Basilio S, Nicolas S, Pedro L, Eliseo PJ, Inaki I, Juan JJ, et al. Using Bayesian networks in construction of a bi-level multi-classifier. A case study using intensive care unit patients data. Artif Intell Med 2001;22:233-48.
25. Heckerman D. Bayesian networks for data mining. Data Mining and Knowledge Discovery 1997;1(1):79-119.
26. Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learning 1995;20:197-243.
27. Turban E, Aronson J. Decision support systems and intelligent systems. 5 ed. Upper Saddle River, NJ: Prentice-Hall; 1995.
28. Haddawy P, Jacobson J, Kahn CE. BANTER: a Bayesian network tutoring shell. Artif Intell Med 1997;10(2):177-200.
29. Witten I, Frank E. Data mining - practical machine learning tools and techniques with java implementations. San Francisco, CA: Morgan Kaufman; 2000.
30. Shortliffe EH. Medical informatics: computer applications in health care and biomedicine. 2nd ed. New York: Springer; 2001.
31. Heckerman D, Chickering D, Meek C, Rounthwaite R, Kadie C. Dependency networks for inference, collaborative filtering, and data visualization. Journal of Machine Learning Research 2000;1:49-75.
32. Fayyad U, Piatetsky-Shapir G, Smyth P, Uthurusamy R, editors. Advances in knowledge discovery and data mining. Menlo Park, CA: AAAI Press / The MIT Press; 1996.
33. Kukar M, Kononenko I, Silvester T. Machine learning in prognosis of the femoral neck fracture recovery. Artif Intell Med 1996;8(5):431-51.
34. Ngan PS, Wong ML, Lam W, Leung KS, Cheng JC. Medical data mining using evolutionary computation. Artif Intell Med 1999;16(1):73-96.
35. Cawsey A. The essence of artificial intelligence. London ; New York: Prentice Hall; 1997.
36. Ramoni M, Riva A, Stefanelli M, Patel V. An ignorant belief network to forecast glucose concentration from clinical databases. Artif Intell Med 1995;7(6):541-59.
37. Wilcox AB, Hripcsak G. The role of domain knowledge in automating medical text report classification. J Am Med Inf Assoc 2003;10(4):330-38.
38. College of American Pathologists. SNOMED CT product profile: SNOMED International; 2003.

39. Moore GW, Berman JJ. Automatic SNOMED coding. Proc Annu Symp Comput Appl Med Care 1994:225-9.
40. Moore GW, Miller RE, Hutchins GM. Indexing by mesh titles of natural-language pathology phrases identified on 1st encounter using the barrier word method. Mod Pathol 1988;1(1):A110-A10.
41. Microsoft. Microsoft visual basic 6.0 (sp4). 8862 ed: Microsoft Corporation; 2000.
42. Klaassen C, editor. Casarett & Doull's toxicology, the basic science of poisons. 5 ed. New York, NY: McGrall-Hill Health Professions Division; 1996.
43. Norsys. Netica. 2.05 ed: Norsys Software Corp; 2002.
44. Geiger D, Heckerman D. Knowledge representation and inference in similarity networks and Bayesian multinets. Artificial Intelligence 1996;82(1):45-74.
45. Baker M, Boult T. Pruning Bayesian networks for efficient computation. Paper presented at: 6th conference on uncertainty in artificial intelligence, 1990; Cambridge, Mass.
46. Margolis DJ, Bilker W, Boston R, Localio R, Berlin JA. Statistical characteristics of area under the receiver operating characteristic curve for a simple prognostic model using traditional and bootstrapped approaches. J Clin Epidemiol 2002;55(5):518-24.
47. Bradley AP. The use of the area under the ROC curve in the evaluatiuon of machine learning algorithms. Pattern Recognition 1997;30(11):45-59.
48. Minitab. Minitab statistical software. 13.32 ed: Minitab Inc; 2002.
49. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford; Boston: Blackwell Scientific Publications; 1994.
50. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(Nov(4-5)):394-403.
51. Lussier YA, Bourque M. Comparing SNOMED and ICPC retrieval accuracies using relational database models. Proc AMIA Annu Fall Symp 1997:514-8.
52. Kudla KM, Rallins MC. SNOMED: a controlled vocabulary for computer-based patient records. J Ahima 1998;69(5):40-4; quiz 45-6.
53. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. J Am Med Inf Assoc 1994;1(3):218-32.
54. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp 1997:640-4.
55. Zimmerman K. SNOMED Representation of explanatory knowledge in veterinary clinical pathology [Blacksburg, VA: Virginia Tech; 2003.
56. Zweigenbaum P, Courtois P. Acquisition of lexical resources from SNOMED for medical language processing. Medinfo 1998;9(Pt 1):586-90.
57. Wingert F. Automated indexing of SNOMED statements into ICD. Methods Inf Med 1987;26(3):93-8.
58. Wingert F. An indexing system for SNOMED. Methods Inf Med 1986;25(1):22-30.
59. Wingert F. Automated indexing based on SNOMED. Methods Inf Med 1985;24(1):27-34.

60. Satomura Y, do Amaral MB. Automated diagnostic indexing by natural language processing. Med Inform 1992;17(3):149-63.
61. Sager N, Lyman M, Nhan NT, Tick LJ. Automatic encoding into SNOMED III: a preliminary investigation. Proc Annu Symp Comput Appl Med Care 1994:230-4.
62. Oliver DE, Altman RB. Extraction of SNOMED concepts from medical record texts. Proc Annu Symp Comput Appl Med Care 1994:179-83.
63. Franz P, Zaiss A, Schulz S, Hahn U, Clar R. Automated coding of diagnoses - three methods compared. Proc AMIA Symp 2000;(20(Suppl)):250-4.
64. Owens D, Shachter R, Nease Jr R. Representation and analysis of medical decision problems with influence diagrams. Med Decis Making 1997;17(5):241-62.
65. Firriolo FJ, Levy BA. Computer expert system for the histopathologic diagnosis of salivary gland neoplasms. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1996;82(2):179-86.
66. Diamond LW, Mishka VG, Seal AH, Nguyen DT. Are normative expert systems appropriate for diagnostic pathology? J Am Med Inform Assoc 1995;2(2):85-93.
67. Li YC, Haug PJ. Evaluating the quality of a probabilistic diagnostic system using different inferencing strategies. Proc Annu Symp Comput Appl Med Care 1993:471-7.
68. Duncan JR, Prasse KW, Mahaffey EA. Veterinary laboratory medicine. 3rd ed. Ames: Iowa State University Press; 1994.
69. Stockham SL, Scott MA. Fundamentals of veterinary clinical pathology. 1st ed. Ames, Iowa: Iowa State Press; 2002.

Conclusion

This work was undertaken to examine the feasibility of using a controlled terminology to provide content, relationships, and a storage format for supporting the creation of a Bayesian network. Four general phases were associated with this investigation, as represented by the presented manuscripts in this dissertation. As a starting point, SNOMED's representation of explanatory concepts, as might be contained in an expert system dealing with clinical pathology, was examined by using several text sources from the domain. After gaining a glimpse of SNOMED's capacity for expressing explanatory-knowledge, work continued on forming a syntax and knowledge representation format aligned with the controlled terminology. A newly derived syntax was then used to model explanatory knowledge within a limited domain of clinical pathology. A comparison of this domain knowledge to SNOMED content provided a focused assessment of SNOMED's capacity to represent explanatory concepts. Lastly, performance was evaluated for a Bayesian network created by extraction of the modeled knowledge.

The first phase of work demonstrated limitations in SNOMED's ability to provide concept content and to express explanatory knowledge. SNOMED is currently lacking much of the content and relationships needed to depict pathways of influence amongst explanatory concepts. The Observable hierarchy (functions and events that can be observed), a logical realm for concepts expressing physiologic processes and biological functions, seemed particularly deficient. This class of concepts was frequently used in the examined text sources to explain why clinical finding events occurred. This missing information is confined to a relative small area of the terminology, indicating a fruitful area on which to focus future efforts to develop and refine SNOMED for use in medical expert systems.

Realizing that creation of explanatory information stored as facts plays a vital part in the development of expert systems served as a motivation for the next stage of work. It is hoped that a standard SNOMED-aligned syntax to represent explanations will assist in creation and reuse of explanation knowledge. Using a purpose-built ontology tool, explanatory pathways adhering to the syntactical pattern $TOIL(FL)_nFD$, defined by using SNOMED-CT hierarchies, were assembled. A specific set of attributes, whose usage was stipulated by the location of the various syntactical elements, was created to form relationships that were true rather than probably true. Explanation pathways modeled using the syntax were stored as OAV triples and reassembled to demonstrate alignment with SNOMED, and were recovered with no loss of meaning.

During the development of the syntax, storage format and the explanations themselves, no effort was made to rely upon SNOMED for concepts. Earlier work in the first phases of this project investigated SNOMED capacity for representing these types of explanatory concepts, but the text sources used in this evaluation contained a mixture of

concept types. Evaluation of the specific explanatory knowledge created for this project provided an additional assessment of SNOMEDS coverage. Results of this analysis were in agreement with those conducted for the text sources. Thirty-one percent of the concepts in the modeled knowledge were located in SNOMED. The "missing" concepts can either be incorporated in SNOMED or could continue to exist outside of SNOMED as an extension.

Evaluation of a Bayesian network created by extracting previously modeled knowledge provided two practical assessments of the work. First, the syntax could be used to assemble a functional Bayesian network. Overall performance of the network based on the extracted knowledge was reasonably good (accuracy = 80%). Second, the expert knowledge provided through the syntax and assembly process had a positive impact on performance of the network. While improvements to the prediction performance may have been possible by choice of more suitable datasets and discretization methods, in its most crude form the derived Bayesian network still outperformed the Naïve model in terms of ROC curve area (79 versus 75). This demonstrates the relevancy of expert knowledge in the creation of Bayesian networks over naïve design approaches and those reliant on data-mining principles. Based on these successful results, it was speculated that explanatory knowledge modeled in the proposed syntax format could facilitate the efficient dissemination and reuse of such information. Further, this work establishes a more formal relationship between controlled terminologies and Bayesian-based expert systems.

THE END

Vita

Kurt Lee Zimmerman was born December 8th, 1959 in Portland, Indiana to Earl and Sondra Zimmerman. He, along with his two sisters and brother, grew up in the Portland area on the family farm of several generations.

Kurt graduated from Jay County High School, Portland, Indiana in 1978. He enrolled as a preveterinary student at Purdue University, West Lafayette, Indiana and was accepted into the Doctor of Veterinary Medicine program in 1980. Throughout college he continued to work on the family farm during summer sessions. Kurt received his DVM degree from Purdue in 1984.

Following graduation, Kurt worked for one year at the Maple Crest Equine Clinic in Westfield, Indiana as an associate veterinarian. In 1985, he joined the mixed animal practice, Apex Veterinary Hospital, in Apex, North Carolina as a clinical veterinarian and partner.

Kurt moved to Blacksburg, Virginia in July of 1997. He began to pursue a Ph.D. in Veterinary Medical Sciences along with a residency in clinical pathology at Virginia Tech - Virginia Maryland Regional College of Veterinary Medicine (VMRCVM) in January of 1998. He became a diplomate of the American College of Veterinary Pathologists in 2001. In 2002, Kurt received an Informatics Research Fellowship Award from the National Library of Medicine in support of his PhD work entitled “Semantic Model of SNOMED as a Bayesian Network”. After completion of his graduate work Kurt will start as an assistant professor in the department of Biomedical Sciences and Pathobiology at VMRCVM.