

Reproducibility of Interpreting “and” and “or” in Terminology Systems

Eneida A. Mendonça, M.D.¹, James J. Cimino, M.D.¹,
Keith E. Campbell, M.D., Ph.D.², Kent A. Spackman, M.D., Ph.D.^{3,4}
¹Columbia University, ²Kaiser Permanente,
³College of American Pathologists, ⁴Oregon Health Sciences University

High quality terminologies are a fundamental requirement in a range of health care applications. To ensure high quality terminologies we should reflect about the understandability, reproducibility and utility criteria within a terminology. This paper describes efforts to improve the understandability of SNOMED. We describe the problem related to the grammatical conjunctions “and” and “or” and how we applied basic semantic rules defined by the SNOMED Editorial Board. The results show that the meaning of “and” and “or” in SNOMED can be made explicit in almost all cases and can be done in a reasonable, reliable, and reproducible manner.

INTRODUCTION

As we start performing more detailed and wide-ranging types of analysis to quantify the quality of health care, we must consider whether the terminologies we use are of sufficiently high quality to support the intended data analysis. In reviewing and developing terminologies over the last few years, we have identified three properties that we apply to the development and evaluation of terminologies: understandability, reproducibility, and utility.

The first property, understandability, is a face validity test. Can you read the term and understand what it means? For example, if you encounter the term “relieving suffering” in a terminology, can you understand the concept that is represented by the phrase? Second is reproducibility. Will others (individuals who enter or analyze data) apply the term in the same way you might based upon their understanding of what the term means? At the surface, terms such as “relieving suffering” may seem to be reproducible, but if you recognize that a particular terminology may also include the term “alieving suffering” as a distinct concept, reproducibly applying the intensional of the terms based upon simply the phrase itself is problematic¹. Instances where we have informally polled individuals on the difference between alieving and relieving have demonstrated to us that the phrases were not sufficiently precise to be reproducible.

Finally, if we were to determine that the phrases are reproducibly applicable to patient records, do the distinctions represented by the phrases provide some utility for some purpose? For example, is the difference between “relieving” and “alieving” of interest to a decision support application or some data analysis application? We think that asking about the understandability, reproducibility, and utility of terms within a terminology is a helpful way to ensure that the terminologies will be of high quality.

As we have tried to apply the understandability, reproducibility, and utility criterion to terminology, we have continually encountered problems with phrases that include “and” and “or”. We believe that we must answer the understandability, reproducibility, and utility questions raised by “and” and “or” before we can delve into the deeper semantic issues posed by other terms such as “relieving” and “alieving.”

In the *Systematized Nomenclature of Medicine (SNOMED)*², for example, we can find terms such as “F-A2080 Sense of position and movement”. Can you read the term and understand it? Does the term refer to sense of position and sense of movement both present at the same time? Or would the presence of one or the other be sufficient to this term being used for coding a medical record?

In an effort to improve the understandability of SNOMED, work is under way to convert SNOMED terms to a logic-based representation form^{3,4}. As part of this work, the SNOMED Editorial Board has determined that grammatical conjunctions must follow basic semantic rules. “And” should be used when both subjects in the term must be present (logical and), “and/or” when one or both must be present (logical inclusive or), and “either_or” when one or the other, but not both, must be present (exclusive or).

We have applied these rules to SNOMED in order to determine if they are sufficient for representing intended meaning and if they can be applied in a reproducible way.

METHODOLOGY

SNOMED International version 3.4 contains 150,543 terms of which many contain the grammatical conjunctions “and” and/or “or”. The objectives of the study were to identify all grammatical conjunctions present in the SNOMED terms and determine their appropriate semantics.

SNOMED is comprised of twelve modules (axes), each one consisting of chapter headings, preferred terms, and synonyms. From each axis, four new files were obtained: (a) preferred terms and synonyms containing “and”, (b) chapter headings containing “and”, (c) preferred terms and synonyms containing “or”, and (d) chapter headings containing “or”. The chapter headings and preferred terms were analyzed separately because of our impression that the meaning and use of conjunctions differs between the two types.

The files were manually reviewed to identify the semantics of each term. We defined the conjunctions according to the SNOMED Editorial Board’s recommendations.

Two independent reviews were done. The first review evaluated the entire SNOMED and was done by two researchers. Initially, the first researcher evaluated the terms in the files and suggested appropriate assignments. The results were presented to the second researcher who reviewed the assignments. In cases of disagreement, the terms were discussed by both to reach a final agreement.

A second independent review was done in one axis, *Diseases/Diagnoses*, by a third researcher. We then performed a simple comparison of the differences between this second review and the results for the *Diseases/Diagnoses* axis from the first review.

RESULTS

SNOMED has a total of 1,956 terms (including preferred terms and synonyms) and 953 heading terms containing the grammatical conjunction “and”. In addition, it has 1,078 terms (including preferred terms and synonyms) and 24 heading terms containing the conjunction “or”. For this project, eleven axes were reviewed by two researchers. Tables 1 and 2 present the results, showing in detail the number of assignments suggested for each axis. It is important to point out that some SNOMED terms have more than one conjunction. The table summarizes the results taking into account the number of conjunctions, not the number of terms in each axis. We found a total of 2,971 “ands” and 1,120 “ors”, including chapter heading, preferred terms and synonyms.

We judged the heading terms to have an “and/or” meaning in 93.68% of the cases (948 heading terms, including terms containing “ands” and “ors”). The remainder 6.32% we judged to be “ands” (59 cases - 5.83%) and “ors” (5 cases - 0.49%). Figure 1 shows an example of four heading term interpretations. Tables 1 and 2 show the detailed assignment of heading terms.

Table 1. Assignment of 2,971 “and” terms to logical conjunctions.

	New assignments					
	Preferred + synonyms			Heading terms		
	and	or	and/or	AND	OR	AND/OR
Topography	70	-	85	-	-	75
Morphology	51	-	11	-	-	39
Function	36	-	58	-	-	187
Living organisms	9	-	5			41
Chemicals, Drugs and Biological Products	50	2	25	59	-	149
Manufacturers of pharmaceutical products	-	-	-	-	-	-
Physical agents, forces and activities	3	-	9	-	-	33
Occupations	17	-	307	-	-	215
Social context	1	-	-	-	-	14
General linkage / Modifiers	3	-	-	-	-	4
Diseases / Diagnoses	766	62	414	-	-	171
Total	1,006	64	914	59	-	928

Table 2. Assignment of 1120 “or” terms to logical conjunctions.

	New assignments					
	Preferred + synonyms			Heading terms		
	or	and	and/or	OR	AND	AND/OR
Topography*	8	2	26	-	-	-
Morphology	9	-	-	3	-	-
Function	7	-	31	-	-	1
Living organisms	-	-	-	-	-	2
Chemicals, Drugs and Biological Products	-	-	15	-	-	-
Manufacturers of pharmaceutical products	-	-	-	-	-	-
Physical agents, forces and activities	1	-	4	-	-	-
Occupations	-	1	8	-	-	1
Social context	1	-	4	-	-	1
General linkage / Modifiers	-	-	2	-	-	-
Diseases / Diagnoses	524	-	452	2	-	15
Procedures						
Total	550	3	542	5	-	20

* We assigned one “or” to “of”.

A. Original: D4-32000 CONGENITAL ANOMALIES OF THE AORTA AND CORONARY ARTERIES

Interpretation: CONGENITAL ANOMALIES OF THE AORTA AND/OR CORONARY ARTERIES

Supporting terms:

D4-32000 Congenital anomaly of aorta, NOS

D4-32010 Congenital anomaly of aortic arch, NOS

D4-32500 Congenital anomaly of coronary artery, NOS

D4-32501 Anomalous origin of coronary artery

Explanation: Note that “Anomalous origin of coronary artery” is an anomaly of coronary artery *and* an anomaly of the aorta.

B. Original: F-52840 Nausea and vomiting

Interpretation: Nausea AND vomiting

Supporting terms:

F-52772 Acute vomiting

F-52843 Increased nausea and vomiting

F-52845 Absence of nausea and vomiting

F-52850 Nausea, vomiting and diarrhea

F-0A610 Alteration in bowel elimination: diarrhea

Explanation: In this case both symptoms must be present. Otherwise, F-52850 could be used.

C. Original: D1-6024 Subacute osteomyelitis with or without periostitis

Interpretation: Subacute osteomyelitis with OR without periostitis

Explanation: We believe “exclusive or” is the only sensible interpretation.

D. Original: D1-21060 Localized osteoarthritis uncertain if primary or secondary

Interpretation: Localized osteoarthritis uncertain if primary OR secondary

Explanation: In this case or means is related to *one or the other*, that is, the location is primary or secondary.

Figure 1. Application of rules to SNOMED terms.

A. Heading term containing “and”. B. Preferred term containing “and”.

C. Preferred term containing “or”. D. Preferred term containing “or”.

The preferred terms and synonyms containing the conjunction “and” were judged to be a logical and in 50.70% (1,006) of the cases, an “and/or” in 46.07% (914) of the cases, and an exclusive or in 3.23% (64). Figure 1 illustrates a case where “and” meant logical and.

The preferred terms and synonyms containing “or” were judged to be a logical exclusive or in 50.23% (550), a logical inclusive or in 49.50% (542) and an “and” in 0.27% (3) of the cases. We judged the meaning to be an exclusive-or when one of these patterns was present: (1) “one or the other”, (2) “with or without”, (3) “not X or Y “(could be “not X and not Y”), (3) “XY or other Y”, where only one Y

would occur per patient, (4) “X or Y affecting W”, and (5) “one or more”. Figure 1 shows two examples where “or” can be defined as an exclusive-or meaning (“with or without” and “one or the other”).

The disagreement between the two researchers in the first review was 1% (35 cases), and these were resolved readily through discussion.

The *Diseases/Diagnoses* axis was independently evaluated by a third researcher as described above. The total number of agreements was 1,177. We disagreed in 236 cases (16.7%). Table 3 shows the comparison between the first and second review.

Table 3. Comparison between reviewers 1,2 and 3.

	Reviewers 1 and 2	Reviewer 2	Total
Agreement	AND	AND	732
	AND/OR	AND/OR	438
	AND	NA*	7
Disagreement	AND/OR	AND	147
	AND	AND/OR	27
	OR	AND	61
	OR	AND/OR	1

* Not applicable: The third reviewer used NA when the components of the phrase do not have meaning on their own (e.g., half and half nail syndrome) or when the operator “and” was between two proper names (e.g., Jervell and Lange-Nielson syndrome).

DISCUSSION

Searches of the medical, medical informatics, and computer science literature failed to yield any relevant work on the logical semantics of conjunctions. We therefore used manual methods to apply the rules as defined by SNOMED Editorial Board to make explicit the meaning of 4091 “ands” and “ors” in SNOMED. From a purely mechanical standpoint, human reviewer generally found the rules sufficient to cover the explicit meanings and that it was to decide which meaning was intended. Inter-rater disagreement showed however that these rules are not so readily applied consistently.

We also found that, in general, when an “and” appears in a heading term, it usually means “and/or”. However, when it appears in a preferred term or synonym, half the time it means “and” and half the time “and/or”. As for the grammatical conjunction “or”, we found that, in general, when “or” appears in

a heading term, it usually means “and/or” (inclusive or). However, when it appears in a preferred term or synonym, half the time it means “exclusive or” and half the time “inclusive or”.

We found that two researchers working together were able to reach agreement rapidly. However, we also found that rules were hard to apply with complete consistency and reproducibility across different groups. The two groups differed on 16%. In some instances, general rules about meaning can be interpreted more than one way, and in these cases, the Editorial Board must make an arbitrary decision about the intended meaning of the term.

When we reviewed the discrepancies, we found that some SNOMED terms were actually ambiguous. For example, “D8-70200 Disorder relating to short gestation and unspecified low birthweight”. In this particular case one reviewer interpreted the meaning as an “exclusive or” and the other one initially

interpreted as a “logical and”. In resolving this discrepancy, the word “unspecified” was dropped because of the ambiguity it introduces, and the final intended meaning is designated as “Disorder relating to short gestation and/or low birthweight”.

This study presents one effort at resolving the ambiguity of SNOMED terms⁵. For decades, SNOMED terms containing “and” and “or” have been used in casual ways that left interpretation up to individual users. As SNOMED evolves to make such meanings more explicit, it is our expectation that understandability will improve. As this occurs, we expect concomitant improvements in reproducibility and utility.

We believe that improving understandability, reproducibility and utility in SNOMED, we would be able to reduce the disparity between interpretations, improving the quality of the vocabulary. As consequence, we expect that SNOMED will better support data analysis and that coding errors will be reduced.

CONCLUSION

We found that the meaning of “and” and “or” in SNOMED could be made explicit in almost all cases and could be done in a reasonable, reliable, and reproducible manner. Further work is needed to clarify how semantic rules should be applied and to remove ambiguity where it exists in the terms themselves.

Acknowledgments

This work was support in part by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil, grant 200557/95-5.

References

1. Campbell KE, Oliver DF, Spackman KA, Shortliffe EH. Representing thoughts, words, and things in the UMLS. *JAMIA* 1998; 5(5):. (in press)
2. Côté, RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L. eds. *The Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International*. Northfield, Illinois. College of American Pathologists, 1993.
3. Campbell KE, Cohn SP, Shortliffe EH, Rennels G. Scalable methodologies for distributed development of logic-based convergent medical terminology. *Methods of Information in Medicine* 1998.(in press.)
4. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. Masys DR, ed. *The emergence of 'Internetable' health care. Systems that really work. Proceedings of the AMIA Annual Fall Symposium 1997*. Philadelphia: Hanley & Belfus, Inc., 1997: 640-44.
5. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an Introspective, Controlled Medical Vocabulary. Kingsland LW, ed.: *Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care; Washington, D.C.; 1989: 513-518.*