

Projection and Inheritance of SNOMED CT Relations between MeSH Terms

Tayeb MERABTI^{a,b,1}, Catherine LETORD^a, Hocine ABDOUNE^c, Thierry LECROQ^b,
Michel JOUBERT^c, Stéfan J. DARMONI^{a,b}

^a *CISMeF, Rouen University Hospital, France*

^b *TIBS, LITIS EA 4108, Institute of Biomedical Research, University of Rouen, France*

^c *LERTIM, Faculty of Medicine, University of Mediterranee, Marseille, France*

Abstract. This paper proposes a methodology to achieve the automatic inheritance of SNOMED CT relations applied to MeSH preferred terms using UMLS as knowledge source server. We propose an interoperability wildcard to achieve this objective. A quantitative and a qualitative analysis were performed on top four SNOMED CT relations inherited between MeSH preferred terms. A total of 12,030 couples of MeSH preferred terms are in relation via at least one SNOMED CT relationship. For the top-four relations inherited between MeSH preferred terms, overall 79.25% of them are relevant, 16.25% as intermediate and 4.5% as irrelevant, as judged by a medical librarian. This work should lead to an optimization of multi-terminology indexing tools, multi-terminology information retrieval and navigation among a multi-terminology server.

Keywords. algorithm, terminology, SNOMED CT, Medical Subject Headings, relationships

1. Introduction

Currently in health informatics, a large number of biomedical terminologies are developed and created for different purposes and specific treatment. Another particularity is that healthcare increasingly changes from isolated treatment towards a continuous treatment process involving multiple healthcare professionals and various institutions. These changes will involve the co-ordination of all the biomedical terminologies and making them interoperable as much as possible. Methods and tools to improve matching between terminologies are thus required. The UMLS (Unified Medical Language System) [1] developed by the “US National Library of Medicine” is an example of such a tool proposed to match health terminologies. Currently, UMLS is the largest database of medical terms from more than 130 terminologies. The SMTS (French acronym of Health Multi-Terminology Server) [2] is another example that will allow semantic interoperability between French terminologies. Different methods are also proposed that allow automatic matching between terminologies where the UMLS is used as a source of knowledge, for example Cimino et al. [3] described how to use the UMLS to convert ICD9-CM (International Classification of Diseases, 9th Edition, Clinical Modifications) terms into MeSH (Medical Subject Headings), and Bodenreider

¹ Corresponding Author: Tayeb Merabti, Medical Library-University Hospital, 1 Rue Germont, 76031 Rouen Cedex; E-mail: tayeb.merabti@chu-rouen.fr

et al. [4] propose a method exploiting the UMLS semantic network to match terminologies. This latter paper proposes a methodology to achieve the automatic inheritance of SNOMED CT (Systematized NOMenclature of MEDicine Clinical Terms) [5] relations applied to MeSH [6] using UMLS as knowledge source server. The focus of this work is clearly librarian related (e.g., information retrieval on the Internet) and not the interoperability of the EHR systems which should be more adhering to strict ontological principles. One key application to this projection and inheritance of these SNOMED CT relations between MeSH terms should optimize information retrieval allowing more specifically expansion and limit of initial queries in all web using the MeSH thesaurus (e.g., MEDLINE/PubMed [7], CISMef [8], Intute [9], NGC (National Guideline Clearinghouse) [10]). This work partakes to a project named InterSTIS (French acronym of Semantic Interoperability of terminologies in French Health Information Systems) funded by the French Research Agency (Health Technologies program)². The goal of InterSTIS is to make interoperable the main French medical terminologies within a “Health Multi-Terminology Server” (HMTS).

2. Material and Methods

2.1. SNOMED CT

SNOMED CT is a comprehensive terminology that provides clinical content and expressivity for clinical documentation and reporting. A total of 1,051,085 terms of SNOMED CT are present in the UMLS, with a number of 308,893 preferred terms (PTs), and a total of 61 relationships are defined in SNOMED CT [5] divided into four types of relationships: First, “Defining characteristics” are the “IS_A” relationship and defining attributes. They are considered defining because they defined formally a SNOMED CT concept by establishing its relationships with other concepts. For example the SNOMED CT concept “*Fracture of tarsal bone*” is defined as *Fracture of foot*, of which the FINDING SITE is *Bone structure of tarsus* and ASSOCIATED MORPHOLOGY is *Fracture*. Second, “Qualifying characteristics” are used to create more complex concepts such as severity or laterality relationships. Third, “Historical relationships” relate inactive concepts to active concepts. Finally, “Additional relationships” are other non-defining characteristics, such as “PART-OF” which is retained for backward compatibility with SNOMED RT [11].

2.2. MeSH

MeSH is the National Library of Medicine’s controlled vocabulary thesaurus used for indexing and cataloging, and searching for biomedical and health-related information and documents. There are 24,767 descriptors and 83 qualifiers (and subheadings) in 2008. In addition to these MeSH headings, there are over 180,000 Supplementary Concept Records within a separate thesaurus. There are also over 97,000 entry terms (or synonyms) that assist in finding the most appropriate MeSH Headings.

In this study, we used PT for the two terminologies SNOMED CT and MeSH. The PT is the term describing a unique medical concept in each terminology. The PT is defined as less ambiguous, more specific and self-descriptive as possible. Any

² Agence Nationale de la Recherche, grant ANR-07-TECSAN-010.

SNOMED CT relation linking two preferred terms implies that these relations also link all the terms that belong to the class of terms represented by each preferred term.

2.3. Quantitative and Qualitative Analyses

To achieve the projection of the SNOMED relations between MeSH terms we used the 2007AB version of the UMLS and we applied a wildcard pattern presented in Figure 1 to realize this mapping [12].

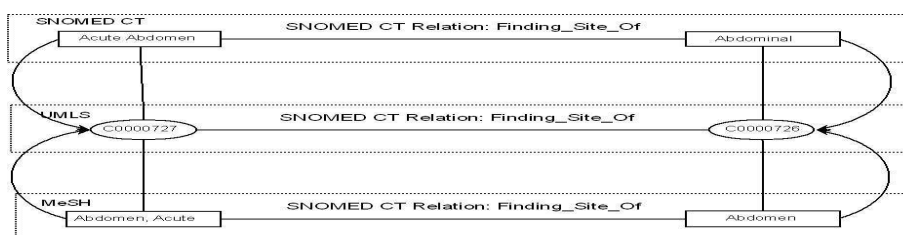


Figure 1. The SNOMED CT relations inheritance between MeSH terms methodology

First, we extract all UMLS Concepts linked by at least one SNOMED CT relations. For example, the two UMLS concepts C0000727, C0000726 associated two SNOMED CT PTs “acute abdomen”, “abdominal” respectively, linked by the SNOMED CT relation “IS_A”. The second step was to map the SNOMED CT relations to the MeSH terminology. To do this we need to project all UMLS concepts founded in first to all MeSH PTs. Finally, we obtained a set of couples of MeSH PTs linked by the SNOMED CT relations. Formally, our wildcard pattern for interoperability is as follows:

- We have A, B two SNOMED CT PTs linked by at least one SNOMED CT relationship.
- We have A', B', two MeSH PTs corresponding to two SNOMED CT preferred term A, B using UMLS.

Therefore, if we have some relation R1 between A and B in SNOMED CT it can be deduced that this (R1) exists between the two preferred term A', B'.

After this quantitative study, a qualitative study was performed on the top-four SNOMED CT relations (IS_A, Finding_Site_of, Associated_Morphology, and Causative_Agent) projected between MeSH PTs. These relations are also considered as Defining characteristics relationships. We constructed a set of all MeSH PTs in relation according to one of this relation, and for each set we selected the 100 first couples obtained. Then, each set was evaluated manually by an expert medical librarian of the CISMef [8] team (CL). The rating was performed using a 3-point Likert scale (relevant, intermediate, and irrelevant) to rate each MeSH PTs couple.

However, a specific treatment was applied in order to refine the “IS_A” relationship to eliminate an explicit “Parent relationship”³ existing between MeSH terms. For example, in MeSH there is no direct “parent relationship” between the two terms “Bronchial diseases” and “Asthma, exercise-induced”. But there is an explicit “Parent relationship” between them because the first is a parent of the second in SNOMED CT.

³ The default MeSH relationship “IS_A”, represented into the UMLS by “PAR”

3. Results: Quantitative and qualitative analyses

A total of 12,030 couples of MeSH PTs are in relation via at least one SNOMED CT relation. Table 1 displays the top-five SNOMED CT relations inherited by MeSH PTs according to our method. Among the 61 SNOMED CT relations, only four of them (IS_A, Finding_Site_of, Associated_Morphology, Causative_Agent) are inherited by more than 300 couples of MeSH PTs. Therefore, the qualitative analysis was performed on these top-four relations. **Erreur ! Référence non valide pour un signet.**² displays the result of the evaluation for each relationship. For the SNOMED CT relation “Associated_Morphology”, overall 90% of inherited relations are relevant when 8% are irrelevant. Whereas for the “Causative_agent” relationship only 64% of the inherited relations are relevant.

Table 1. Numbers of pairs of MeSH PTs inherited from 5 principal SNOMED CT relations

SNOMED CT Relations	Couples of MeSH PTs in relation
IS_A	6871
Finding_Site_of	2512
Associated_Morphology	1080
Causative_Agent	328
Associated_With	53

Table 2. Relevance of the four SNOMED CT relationships inherited from MeSH PTs

SNOMED CT Relationships	Relevant (%)	Intermediate (%)	Irrelevant (%)
IS_A	75	17	8
Finding_Site_Of	88	10	2
Associated_Morphology	90	2	8
Causative_Agent	64	36	0
Average	79.25	16.25	4.5

4. Discussion, Conclusion and Perspectives

We focused on this study on the MeSH which is the current terminology used in CISMef⁴ [8] (a web site dedicated to Catalog and Index Health Resources in French). We applied a comparable methodology in a previous work [12] to two other health terminologies: SNOMED 3.5 [13] and ICD10 [14].

For the top four relations (IS_A, Finding_Site_of, Associated_Morphology, and Causative_Agent) the qualitative evaluation performed by a CISMef medical librarian was considered as very encouraging. Therefore, the CISMef team has decided to implement it in its R&D version to optimize information retrieval. This optimization will allow to limit or expand a query: e.g., the query ‘acute abdomen’ will propose an expansion/limit with Localization: Abdomen. Expansion will expand the two terms (acute abdomen or abdomen) where the limit will generate the following query: acute abdomen and abdomen. This optimization could be used in any Web site using the MeSH thesaurus [7–10]. This study allows establishing new relations between MeSH terms (e.g., the MeSH term “acute abdomen” is linked by the SNOMED CT relation “Finding_Site_of” to the MeSH term “abdomen”). This study will also improve navigation in the Health Multi-Terminology Server integrating these relationships. The methodology proposed in this paper is very dependent on two points: a) the validity of SNOMED CT relationships and b) the perfect match provided by the UMLS. In fact, wrong attribution of SNOMED CT relations between two concepts can lead to

⁴ www.cismef.org

incorrect inferences in the other terminologies. An example of the “IS_A” relationship. For instance, the SNOMED CT concept “Neoplasm of uterus” is wrongly subsumed by the SNOMED CT concept “Neoplasm of abdomen” which will lead an incorrect deduction in MeSH. There are a lot of such problems in SNOMED CT [15] due to: improper treatment of negation (e.g., the concept “Dupuytren’s disease of palm, nodules with no contracture” is subsumed by the concept “contracture of palmar disease”) or to improper treatment of the partial/complete distinction. Our main perspective is to apply the same methodology to all French-speaking terminologies integrated into the Health Multi-Terminology Server. The maintenance of the inheritance of SNOMED CT relations between MeSH terms, and between terms of other terminologies, will be insured by the Health Multi-Terminology Server.

Acknowledgment. This work is partially granted by the project “Semantic Interoperability of terminologies in French Health Information Systems” (French Acronym InterSTIS) project funded by the French National Research Agency (French acronym ANR).

References

- [1] Lindberg, C. (1990) The Unified Medical Language System (UMLS) of the National Library of Medicine. *Journal of the American Medical Record Association* 61(5):40–42.
- [2] Joubert, M., Dahamna, B., Delahousse, J., Fieschi, M., Darmoni, S.J. (2009) SMTS®: Un Serveur Multi-Terminologies de Santé. In Fieschi, M. et al. (Eds.) *Risques, technologies de l’information pour les pratiques médicales: Proceedings of JFIM 2009*, Informatique et Santé, Springer, NICE, 17:47–56.
- [3] Cimino, J., Johnson, S., Peng, P., Aguirre, A. (1993) From ICD9-CM to MeSH using the UMLS: A how-to guide. *Proc Annu Symp Comput Appl Med Care* 730-734.
- [4] Bodenreider, O., Nelson, S.J. (1998) Beyond synonymy: Exploiting the UMLS Semantics in Mapping Vocabularies. *AMIA Annual Symposium Proceedings 1998*, 815–819.
- [5] College of American Pathologists. *SNOMED CT- User Guide*, July 2008 Release.
- [6] Nelson, S.J., Johnson, W.D., Humphreys, B.L. (2001) Relationships in Medical Subject Headings. In: Bean CA, Green R. (eds.) *Relationships in the organization of knowledge*. Kluwer Academic Publishers, New York. p.171-84.
- [7] PubMed/Medline, <http://www.ncbi.nlm.nih.gov/pubmed/>.
- [8] Darmoni, S.J., Leroy, J.P., Baudic, F., Douyère, M., Piot, J., Thirion, B. (2000) CISMef: A structured health resource guide. *Methods of Information in Medicine* 39(1):30–35.
- [9] Williams, C. (2006) Intute: The New Best of the Web. *ARIADNE* issue 48, <http://www.ariadne.ac.uk/issue48/williams/>.
- [10] Goolsby, M.J. (2003) National Guideline Clearinghouse guideline syntheses and comparisons. *Journal of the American Academy of Nurse Practitioners* 15(3):98–101.
- [11] Spackman, K.A., Campbell, K.E., Côté, R.A. (1993) SNOMED RT: A reference terminology for health care. *AMIA Annual Fall Symposium 1993*, 640–644.
- [12] Merabti, T., Abdoune, H., Lecroq, T., Joubert, M., Darmoni, S.J. (2009) Projection des relations SNOMED CT entre les termes de deux terminologies (CIM10 et SNOMED 3.5). In Fieschi, M. et al. (Eds.) *Risques, technologies de l’information pour les pratiques médicales: Proceedings of JFIM 2009*, Informatique et Santé, Springer, NICE, 17:79-88.
- [13] Côté, R.A., Rothwell, D.J., Patolay, J.L., Beckett, R.S. Brochu, L. (Eds.) (1993) *The Systematised Nomenclature of Human and Veterinary Medicine: SNOMED International*. College of American Pathologists, Northfield.
- [14] ICD10. <http://www.who.int/classifications/icd/en/>.
- [15] Ceusters, W., Smith, B., Kumar, A., Dhaen, C. (2004) Ontology-based error detection in SNOMED-CT®. In Fieschi, M. et al. (Eds.) *Proceedings of Medinfo 2004*, IOS Press, Amsterdam, 482–486.