

Relating UMLS semantic types and task-based ontology to computer-interpretable clinical practice guidelines

Anand Kumar^a, Paolo Ciccarese^a, Silvana Quaglini^a, Mario Stefanelli^a, Ezio Caffi^b, Lorenzo Boiocchi^b

^aLaboratory of Medical Informatics, Department of Computer Science, University of Pavia, Pavia, Italy.

^bConsorzio di Bioingegneria e Informatica Medica, Pavia, Italy.

Abstract

Medical knowledge in clinical practice guideline (GL) texts is the source of task-based computer-interpretable clinical guideline models (CIGMs). We have used Unified Medical Language System (UMLS) semantic types (STs) to understand the percentage of GL text which belongs to a particular ST. We also use UMLS semantic network together with the CIGM-specific ontology to derive a semantic meaning behind the GL text. In order to achieve this objective, we took nine GL texts from the National Guideline Clearinghouse (NGC) and marked up the text dealing with a particular ST. The STs we took into consideration were restricted taking into account the requirements of a task-based CIGM. We used DARPA Agent Markup Language and Ontology Inference Layer (DAML+OIL) to create the UMLS and CIGM specific semantic network. For the latter, as a bench test, we used the 1999 WHO-International Society of Hypertension Guidelines for the Management of Hypertension. We took into consideration the UMLS STs closest to the clinical tasks. The percentage of the GL text dealing with the ST "Health Care Activity" and subtypes "Laboratory Procedure", "Diagnostic Procedure" and "Therapeutic or Preventive Procedure" were measured. The parts of text belonging to other STs or comments were separated. A mapping of terms belonging to other STs was done to the STs under "HCA" for representation in DAML+OIL. As a result, we found that the three STs under "HCA" were the predominant STs present in the GL text. In cases where the terms of related STs existed, they were mapped into one of the three STs. The DAML+OIL representation was able to describe the hierarchy in task-based CIGMs. To conclude, we understood that the three STs could be used to represent the semantic network of the task-based CIGMs. We identified some mapping operators which could be used for the mapping of other STs into these.

Keywords: Computer-interpretable guideline, clinical practice guideline, Ontology, Semantic type.

1. Introduction

Creation of task-based CIGMs [1] requires effective medical ontologies and terminologies [2]. These support formal reasoning and have been developed for different purposes by different organisations. Task-based CIGMs are based on execution of clinical activities in the healthcare setup. UMLS [3] mentions the terms and the STs related to them and has been used as a knowledge acquisition tool for rule-based clinical decision support systems. From the point of view of CIGMs, the STs needed for their representation can be restricted. This not only helps understand the task specification in

CIGMs but also helps create the ontology. It also falls in line with the problem-solving approach where the tasks should be subdivided into subtasks in order to achieve a specific goal. This has been the approach in the different GL modelling tools.

We report the percentage of nine GL texts from NGC which deal with a particular ST. We take into consideration the UMLS semantic network [3] and map the terms and the STs using DAML+OIL [4] based on the concepts of Default Reasoning.

2. Materials and methods

a.) UMLS Metathesaurus and GLs

The UMLS Metathesaurus [3] consists of terms from its controlled source vocabularies. The meaning of each term is defined by its source, explicitly by definition or annotation; by context (its place in a hierarchy); by synonyms and other stated relationships between terms; and by its use in description, classification, or indexing. To every Metathesaurus concept is assigned at least one ST. From the GL point of view, most of the GLs suggested actions can be mapped into the terminology associated with the STs – Laboratory Procedure (LP), Diagnostic Procedure (DP) and Therapeutic or Preventive Procedure (TPP).[Table 1] These STs are subclassifications of the ST Health Care Activity (HCA).

Thus in terms of the GL-based task determination, the tasks suggested in the GLs can be linked to one of these STs. Other STs which are used relatively less frequently in GLs are related to “Occupational Activity”, which contains “Educational Activity”, “Governmental or Regulatory Activity” and “Research Activity”. And even these are related to “HCA”. For example, “Research Activity” actually helps to determine the “HCA” by giving the “Strength of Evidence”.

b.) GL text markup considering the UMLS STs

Nine GL texts from the NGC were chosen belonging to different pathologies and the pieces of text dealing with the different STs were marked manually. This was done to demonstrate the portions of GL text which belonged to the different UMLS STs.

The results confirm that most of the GL text deals with STs mentioned under the “HCA”. A high percentage of OOA (Occupational Activities or Explanations) in general means that the GL contained a lot of explanation for the tasks as they were counted as OOA. These included research activities including surveys or explanation of pathophysiology etc. behind the task. This in turn signifies that it is possible to create task-based CIGMs using certain specific STs of UMLS. In situations where other STs were specified, we used default reasoning to map them. For example,

Term – Hemoglobin level ST- Laboratory or Test Result
is mapped to Term – Hemoglobin measurement ST- Laboratory Procedure
because according to the UMLS Semantic Network
Laboratory or Test Result |result of| Laboratory Procedure

c.) UMLS's semantic Network and DAML+OIL

The UMLS Semantic Network provides a consistent categorization of all concepts represented in the UMLS Metathesaurus and the important relationships between them.

All information about specific concepts is found in the Metathesaurus; the Network provides information about the set of basic STs, or categories, which may be assigned to these concepts, and it defines the set of relationships that may hold between the STs.

We used OILed [4], the DAML+OIL editor, in order to specify the relationships between the several UMLS STs and the rest of three STs mentioned above- "LP", "DP" and "TPP". [Fig 1] The main features of OIL are class and slot definitions. A class definition (class-def) associates a class name with the class description. The class-def contains slot constraints, which is a list of one or more constraints applied to a slot (a binary relation). A slot definition (slot-def) associates a slot name with a slot description. We used DAML+OIL because it has a frame-based structure which helps to specify the medical context and can support reverse relationships between the classes. The UMLS Semantic Network [3] defines first order relationships between the STs, which we extended using the frame description and inverse relationships of DAML+OIL.

For example, to represent that "LP" analyzes "Chemical", we use the following class and slot definitions. There is also a possibility to indicate the inverse relationship.

```
class-def defined Laboratory Procedure
      subclass-of Health Care Activity
      slot-constraint analyzes
            value-type Chemical
class def defined Chemical
      subclass-of Substance
slot-def analyzes
      inverse is analyzed by
```

d.) GL-specific ontology

We used the 1999 WHO-International Society of Hypertension Guidelines for the Management of Hypertension. The STs mentioned in the GL text were mapped to the three STs mentioned above using some operators. These operators signify the relationship among the ST's as mentioned in the UMLS Semantic Network. For example, for the term "Proteinuria", we did the following mapping.

```
Term – Proteinuria      ST - Laboratory or Test Result, Disease or Syndrome
Operator – DOF (Determination of)
Term – DOF Proteinuria      New ST – Laboratory Procedure
```

The operator "DOF" is used since a corresponding term with ST as "LP" was not present in this case.

4. Results

The knowledge in the GL texts could be related to the three STs under "HCA" for the most part. This also included the mapping among the STs as mentioned above. With the use of DAML+OIL, it was possible to create an ontology for the task-based CIGMs. We were able to create a task-subtask network which followed the UMLS STs.

5. Discussion

The work underlines that for the task-based CIGMs, we can restrict the UMLS STs. Considering the findings from the nine GL texts, on an average, more than 80% of the texts dealt directly with the three UMLS STs – “LP”, “DP”, “TPP” mentioned under “HCA”, directly or indirectly after mapping. This is significant considering that there are more than 100 STs originally mentioned in the UMLS for the terminology.

We created two separate ontologies. The first was generic which was derived from the UMLS Semantic Network. This itself had almost 400 slot constraints for the three STs under “HCA”. The second was an ontology specific to a particular GL text, whose complexity depends on the particular GL text under consideration.

Thus, in addition to the complexity of GL texts, there are a large number of STs and slot constraints. The reduction in the number of the STs and task-based hierarchy helped to represent the medical knowledge formally and to understand the semantic relationships among the various tasks mentioned.

Considering that the implementation of GLs needs an execution engine, this effort would help to understand the meaning behind the clinical tasks performed in the wake of UMLS’s formal terminology and semantic network and DAML+OIL’s strength to represent the ontological relationships.

6. Conclusion

The developments in the areas of semantic network and clinical decision support systems have triggered efforts towards the use of common standards as regards terminology, ontology, knowledge sources etc. With the adaptation of the generic standards like UMLS and DAML+OIL for task based CIGMs, our effort is on the lines of the use of those standards in the GL text based knowledge representation and creation of decision support tools.

7. References

- [1] Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R., Johnson PD, Jones N., Kumar A., Miksch S., Quaglino S., Seyfang A., Shortliffe EH, and Stefanelli M. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. *J. Am. Med. Inform. Assoc.* Preprint doi:10.1197/jamia.M1135
- [2] Nigel S., Michel C, and Jean PB. Which coding system for therapeutic information in evidence-based medicine. *Computer Methods and Programs in Biomedicine* 2002; 68(1):73-85.
- [3] Achour LS, Dojat M, Rieux C, Bierling P, and Lepage E. A UMLS-based Knowledge Acquisition Tool for Rule-based Clinical Decision Support System Development. *J Am Med Inform Assoc* 2001;8(4):351-360.
- [4] Bechhofer S, Horrocks I, Goble C, Robert S. OilEd: a Reason-able Ontology Editor for the Semantic Web. *Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, September 19-21, Vienna. Springer-Verlag LNAI Vol. 2174, pp 396--408. 2001.*

8. Address for correspondence

Anand Kumar; Laboratory of Medical Informatics; Department of Computer Science; Via Ferrata 1; University of Pavia; Pavia ; PV 27100; Italy. Email: anand@aim.unipv.it

UMLS ST	Definition
Diagnostic Procedure	A method, procedure or technique used to determine the nature or identity of a disease or disorder. This excludes procedures which are primarily carried out on specimens in a laboratory.
Laboratory Procedure	A procedure method or technique used to determine the composition, quality, or concentration of a specimen, and which is carried out in a clinical laboratory. Included here are procedures which measure the times and rates of reactions.
Therapeutic or Preventive Procedure	A procedure method or technique designed to prevent a disease or a disorder, or to improve physical function, or used in the process of treating a disease or injury.

Table 1. The definitions of the 3 UMLS STs used in the task-based CIGM.

The definitions were taken the UMLS Knowledge Source Server (UMLSKS), developed by the National Library of Medicine. Website <http://www.nlm.nih.gov/research/umls/>

GL Title	GL Source (Organisation)	HCA			OOA
		DP	LP	TPP	
National High Blood Pressure Education Program Working Group report on high blood pressure in pregnancy.	National High Blood Pressure Education Program/National Heart, Lung, and Blood Institute (U.S.).	16	10	43	31
Sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure.	National High Blood Pressure Education Program/National Heart, Lung, and Blood Institute (U.S.).	51	6	40	3
Safe handling of cytotoxic drugs: an independent study module.	Oncology Nursing Society	2	6	74	18
Diagnosis and management of asthma.	Institute for Clinical Systems Improvement	39	26	20	15
Standards for breast conservation therapy in the management of invasive breast carcinoma.	American College of Radiology	23	13	47	17
Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease.	National Heart, Lung, and Blood Institute	29	10	54	7
Diabetic retinopathy.	American Academy of Ophthalmology	23	5	43	29
The management of diabetes mellitus in the primary care setting.	American Diabetes Association.	37	13	26	24
Management of dyslipidemia in adults with diabetes.	American Diabetes Association	4	14	75	7

HCA = Health Care Activity ; DP = Diagnostic Procedure ; LP = Laboratory Procedure; TPP = Therapeutic or Preventive Procedure; OOA = Occupational Activities or Explanations

Table 2. The percentage of GL text which belongs to the different UMLS STs with the GL and the GL source. The GL texts were accessed from the NGC, a public resource for evidence-based clinical practice guidelines, sponsored by the Agency for Healthcare Research and Quality in partnership with the American Medical Association and the American Association of Health Plans. Website <http://www.guideline.gov/index.asp>

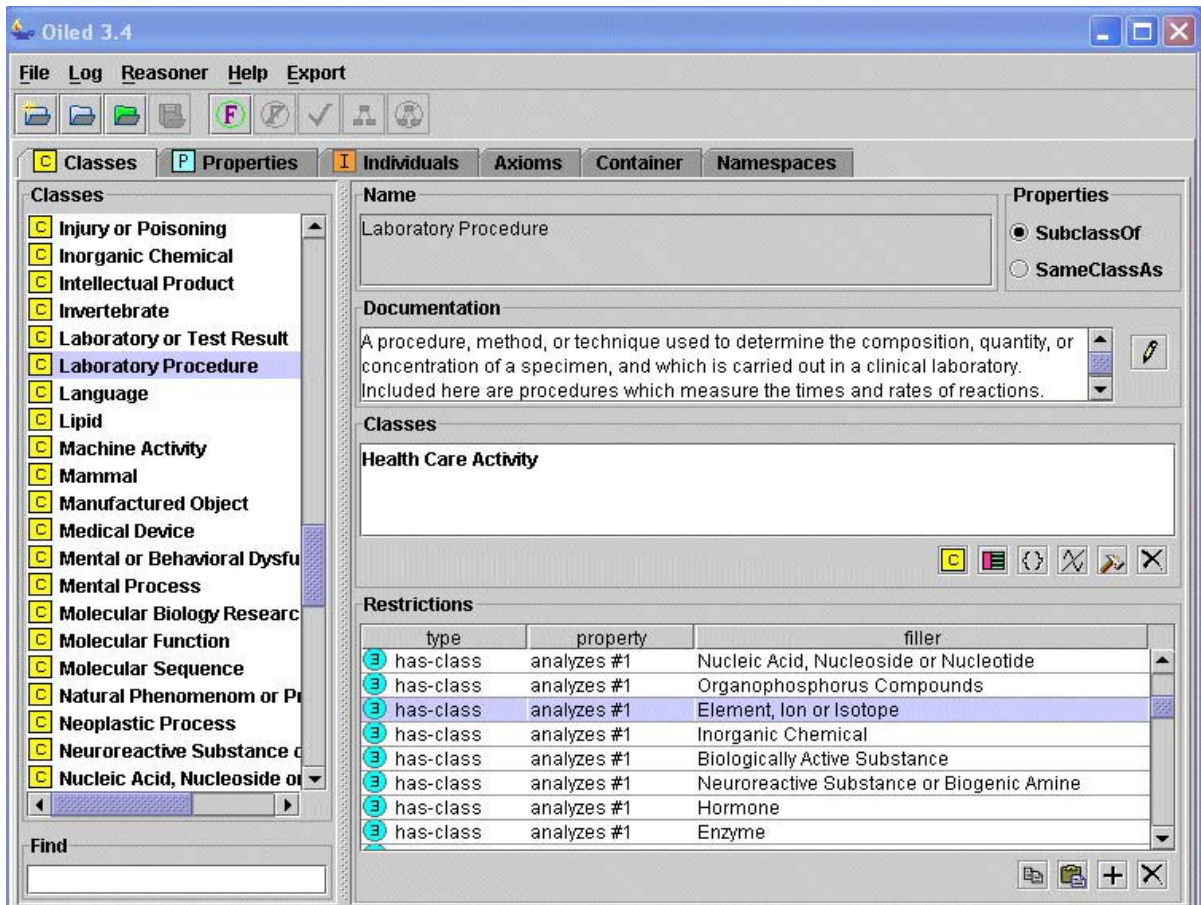


Fig 1. Oiled depiction of slot-constraints applied to the class “LP” using UMLS STs

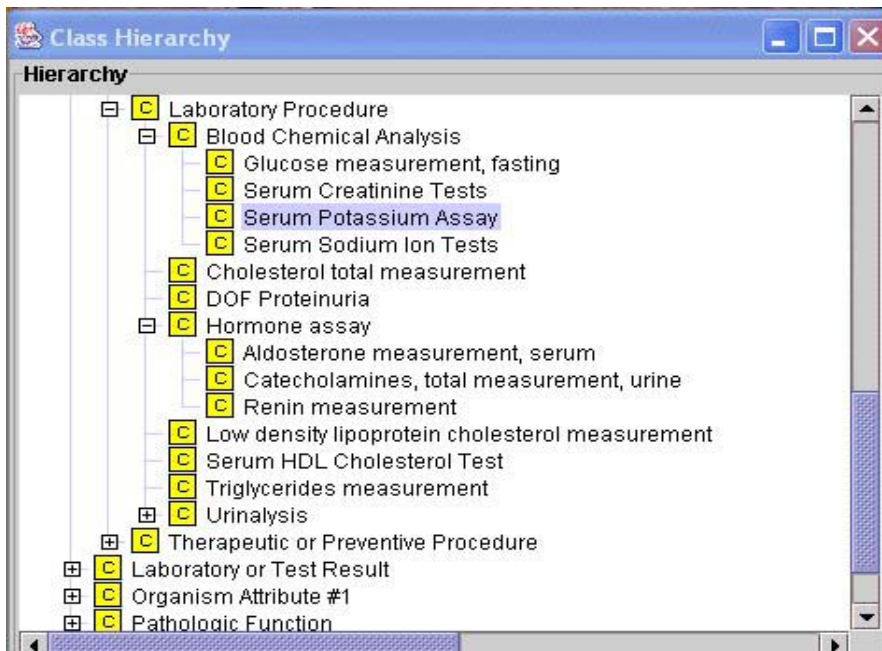


Fig 2. Task-based ontological representation of 1999 WHO-International Society of Hypertension Guidelines for the Management of Hypertension