Ontology for task-based clinical guidelines and the theory of granular partitions

Anand Kumar[a], Barry Smith[b]

[a] Laboratory of Medical Informatics, Department of Computer Science, University of Pavia, Italy.
[b] Institute for Formal Ontology and Medical Information Science, University of Leipzig, Germany and Department of Philosophy, University at Buffalo

Abstract. The theory of granular partitions (TGP) is a new approach to the understanding of ontologies and other classificatory systems. The paper explores the use of this new theory in the treatment of task-based clinical guidelines as a means for better understanding the relations between different clinical tasks, both within the framework of a single guideline and between related guidelines. We used as our starting point a DAML+OIL-based ontology for the WHO guideline for hypertension management, comparing this with related guidelines and attempting to show that TGP provides a flexible and highly expressive basis for the manipulation of ontologies of a sort which might be useful in providing more adequate Computer Interpretable Guideline Models (CIGMs) in the future.

1. Introduction

1.1 Clinical Practice Guidelines from an ontological point of view

Clinical Practice Guidelines (GLs) are 'systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.' [1] Their use in clinical decision-making is intended to improve the outcomes of clinical care. Given that most GLs are free texts or simple flowcharts, there is a growing need to create Computer Interpretable Guideline Models (CIGMs) [2]. For this, however, we require standardized terminologies based on coherent ontologies of clinical activities [3].The Unified Medical Language System (UMLS) of the National Library of Medicine integrates a number of standard medical terminologies into a single unified knowledge representation system. [4,5]

While the UMLS provides its terms with associated Semantic Types, in order to use the latter in CIGMs one needs to incorporate them within some ontological framework. Among the emerging standards in this field, the DARPA Agent Markup

Language and Ontology Interface Language (DAML+OIL) is a recent proposal for an ontology representation language suitable for such purposes [6].

1.2 Reference Ontologies and Applications Ontologies

DAML+OIL is an ontology language within the currently dominant paradigm, which views ontologies as *applications* capable of running in real time and exploiting the reasoning power of one or other variant of Description Logic. There is however a second paradigm in ontology – that of 'Reference Ontology' – whose proponents hold that the needs of terminology integration and standardization can be met only through the development of ontological theories marked by a high degree of descriptive adequacy.

One product of the Reference Ontology approach is the theory of granular partitions (TGP). This is designed to yield a framework within which both formal and informal representations of reality at different levels of granularity (for example molecule, cell, and whole-organism granularities) can be incorporated together [7, 8]. Our task here is to compare the results of adding to the pure DAML+OIL framework for analysis of guidelines the supplementary resources of TGP.

2. The UMLS Semantic Network, DAML+OIL and Guidelines

2.1 UMLS Semantic Types for Task-Based Guidelines

Most of the actions referred to in GLs can be mapped into that part of the UMLS terminology that is associated with the Semantic Types *Laboratory Procedure*, *Diagnostic Procedure* and *Therapeutic or Preventive Procedure*. All of these are sub-types of the Semantic Type *Health Care Activity*.

Other Semantic Types closely associated with *Health Care Activity* but used less frequently in GLs are: *Educational Activity*, *Governmental or Regulatory Activity* and *Research Activity*, all of which are subtypes of *Occupational Activity*. An instance of *Research Activity*, for example, is an instance of *Health Care Activity* marked in addition by the feature: *strength of evidence*.

2.2 The Case of Hypertension

The Guidelines for the Management of Hypertension prepared in 1999 by the WHO International Society of Hypertension were used as a basis for our DAML+OIL-based ontology for hypertension GLs [9,10]. The Semantic Types mentioned in the GL text were mapped to the three Semantic Types mentioned above, using operators such as 'Determination of' (abbreviated 'DOF') to signify the relationships among the Semantic Types in the UMLS Semantic Network. For example, the term 'Proteinuria' was assigned the following mapping:

Term – Proteinuria
Semantic Type – *Laboratory or Test Result, Disease or Syndrome*
    Operator – DOF (Determination of)
Term – DOF Proteinuria
New Semantic Type – *Laboratory Procedure*

According to this analysis, Proteinuria is either a laboratory or test result or a disease or syndrome. Since our GL ontology is restricted to the Semantic Type *Health Care Activity* we need to find a roundabout way of incorporating Proteinuria and similar terms, and we do this precisely by means of constructions such as: *determination of* (the presence of) proteinuria.

3. The Theory of Granular Partitions (TGP)

3.1 Background Rationale

When human beings engage in listing, mapping or classifying activities – for example when they seek to classify the domain of clinical activities in terms of UMLS Semantic Types or in terms of Guidelines or CIGMs – then they *partition* reality into *cells* of various sorts.

Perhaps the most important feature of TGP is that it recognizes that different partitions may represent cuts through the same reality at different levels, and even cuts through reality which are skew to each other.

Each partition consists of cells and subcells, the latter being nested within the former. Partitions can be hierarchical: they then consist of many layers of cells and subcells (for example in the animal kingdom the layers of genus, species, family, order, phylum, kingdom and so forth). The lowest layer of subcells corresponds to the finest grain of objects recognized by the partition in question.

3.2 The Axioms of TGP

The axioms of TGP can be given in partial and simplified form as follows. (We ignore here those aspects of the theory dealing with mereological structure and with vagueness of projection; for details see [7], [8]). For orientation one can think of the relation between an object and a cell in which it is located as analogous to the relation between an element and its singleton in set theory. The subcell relation is then a restricted version of the set-theoretical subset relation, formulated in such a way that each partition is isomorphic to a tree in the graph-theoretical sense:

A1: Every partition has a unique maximal or root cell in which all other cells are included as subcells.

A2: The subcell relation is reflexive, antisymmetric, and transitive.

A3: Each cell in a partition is connected to the root via a finite chain of immediate succeeding cells.

A4: If two cells within a partition overlap, then one is a subcell of the other.

These axioms relate to a granular partition as a system of cells.

4. Use of TGP as a Supplement to DAML+OIL-Based Ontologies

4.1 The Hypothesis

After translating our GL texts annotated in terms of UMLS Semantic Types into DAML+OIL ontologies, the results still need to be supplemented by machinery of the sort provided by TGP if they are to allow the comparison and manipulation of distinct

ontologies within a single framework. Our hypothesis is that we can achieve better (more natural, more scaleable, and more expressively powerful) results if we supplement the DAML+OIL framework with the resources of TGP.

4.2 Implementation based on TGP

By A1, a unique maximal cell contains as sub-cells all the cells present in the partition. In our present example, the domain of the partition is: the totality of activities in accordance with the given clinical guidelines. The maximal cell is then the UMLS Semantic Type *Health Care Activity*, which covers all the tasks specified in the guidelines. All other cells stand to this maximal cell in a subcell relation which satisfies A2. The immediate sub-cells of this maximal cell (as in the UMLS Semantic Network) are *Laboratory Procedure*, *Diagnostic Procedure* and *Therapeutic or Preventive Procedure*, which have further subcells depending on the GL text at issue.

By A2, A3 and A4, each distinct cell in a partition is connected to the root via a finite chain of immediately succeeding cells. This generates a nestedness of cells in the form of chains, terminating in the smallest cells, also called the *leaves* of the tree. For example, in Fig. 2, *determination of smoking* and *determination of women's age to be greater than 65* are leaves for the total partition.
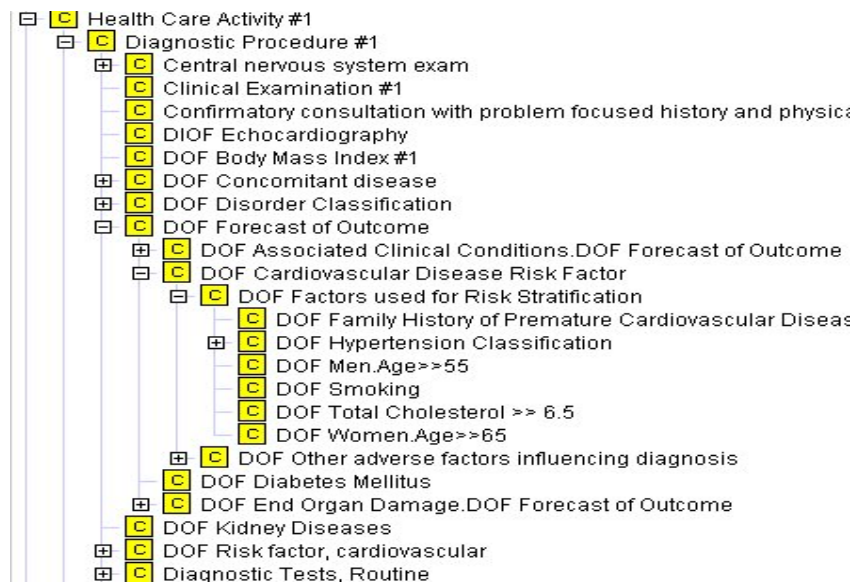

Fig. 2. Task-based ontological representation of the 1999 WHO International Society of Hypertension Guidelines for the Management of Hypertension

Leading out from *Diagnostic Procedure* via subcell relations we have cells for: *Determination of Forecast of Outcome, Determination of Cardiovascular Risk Factor, Determination of Factors Used in Risk Stratification* and finally *Determination of*

Hypertension Classification. This last is a leaf in the granular partition corresponding to the WHO GL for hypertension management.

By transitioning between taxonomical and partonomical partitions we can now represent the summation of subtasks at the same level in the granular hierarchy of a GL task-subtask structure, for example as follows: *DOF Family History of Premature Cardiovascular Disease* ∪ *DOF Hypertension Classification* ∪ *DOF Men.Age* >> 55 ∪ *DOF Total Cholesterol* >> 6.5 ∪ *DOF Women.Age* >> 65 together yield the task: *DOF Factors Used for Risk Stratification* on the next highest granularity level. When the latter is summed on this level with the cells *Central Nervous System Examination*, *Abdominal Examination* etc., then this yields the cell *Diagnostic Procedure*, which is present both in the UMLS Semantic Types and in the GL-specific ontology.

5 Conclusion

We have sketched how the theory of granular partitions can be used in the creation of ontologies for clinical practice guidelines by providing a framework within which we can transition between ontologies of different sorts and at different granular levels. We believe that this will produce a robust and flexible platform for the formulation of intuitive and easily extendible computer-interpretable guideline models.

References

1. Field M, Lohr KN. Attributes of good practice guidelines. In: Field M, Lohr KN, editors. Clinical practice guidelines: directions for a new program. Washington, DC: *National Academy Press*, 1990: 53-77.
2. Peleg M., Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R., Johnson PD, Jones N., Kumar A., Miksch S., Quaglini S., Seyfang A., Shortliffe EH, and Stefanelli M. Comparing Computer-Interpretable Guideline Models: A Case-Study Approach. *J Am Med Inform Assoc*. 2003 Jan-Feb;10(1): 52-68.
3. 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.
4. Humphreys BL, Lindberg DA, Schoolman HM, Barnett. The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc*. 1998 Jan-Feb; 5(1): 1-11.
5. UMLS website http://www.nlm.nih.gov/research/umls/
6. 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, 396-408. 2001.
7. Bittner, T. and Smith, B. (2003a). Granular Spatio-Temporal Ontologies, To appear in: *Proceedings of the AAAI Spring Symposium on Foundations and Applications of Spatio-Temporal Reasoning* (FASTR).

8. Bittner, T. and Smith, B. (2003). A Theory of Granular Partitions. In: *Foundations of Geographic Information Science*, M. Duckham, M. F. Goodchild and M. F. Worboys (eds.), London: Taylor & Francis, 117–151

9. Kumar A, Ciccarese P, Quaglini S, Stefanelli M, Caffi E, Boiocchi L. Relating UMLS semantic types and task-based ontology to computer-interpretable clinical practice guidelines. *Proc MIE 2003*.

10. WHO Hypertension Guideline. **http://www.bpcr.com/diet/who/**