Formalizing UMLS Relations Using Semantic Partitions in the Context of Task-Based Clinical Guidelines Model

Anand Kumar[a,b], Matteo Piazza[a], Barry Smith[b,c], Silvana Quaglini[a], Mario Stefanelli[a]

[a]*Laboratory of Medical Informatics, Department of Computer Science, University of Pavia, Italy*
[b]*Institute for Formal Ontology and Medical Science, Faculty of Medicine, University of Leipzig, Germany*
[c]*Department of Philosophy, University at Buffalo, New York, USA*

Abstract
An important part of the Unified Medical Language System (UMLS) is its Semantic Network, consisting of 134 Semantic Types connected to each other by edges formed by one or more of 54 distinct Relation Types. This Network is however for many purposes over-complex, and various groups have thus made attempts at simplification. Here we take this work further by simplifying the relations which involve the three Semantic Types – *Diagnostic Procedure*, *Laboratory Procedure* and *Therapeutic or Preventive Procedure*. We define operators which can be used to generate terms instantiating types from this selected set when applied to terms designating certain other Semantic Types, including almost all the terms specifying clinical tasks. Usage of such operators thus provides a useful and economical way of specifying clinical tasks. The operators allow us to define a mapping between those types within the UMLS which do not represent clinical tasks and those which do. This mapping then provides a basis for an ontology of clinical tasks that can be used in the formulation of computer-interpretable clinical guideline models.

Index Terms: UMLS, Semantic Types, Semantic Network, Graph Theory, Ontology, Terminology, Clinical Guidelines.

I. INTRODUCTION

The creation of task-based computer-interpretable guideline models requires effective medical ontologies and terminologies. [1,2] The Unified Medical Language System (UMLS), designed by the US National Library of Medicine, integrates the major standard terminologies into a single framework for knowledge representation. [3,4] It includes a concept repository, the Metathesaurus (META), and a Semantic Network, which serves as a

high-level abstraction designed to support navigation through META and through associated representation systems. The Semantic Network (UMLS Knowledge Source Server Version 4.0) consists of 134 Semantic Types and 54 types of relations which can hold between these types. [5] These together form a graph with a double tree structure, with *Event* and *Entity* forming the respective roots. The vertices consist of the Semantic Types and the edges consist of the links between them. The corresponding complete graph contains more than 6000 edges.

Unfortunately the large number of relations in the Semantic Network implies a certain degree of redundancy, and we here illustrate a method by means of which this number of relations can be reduced in a way which brings benefits of efficiency for purposes of automatic reasoning. We are here interested specifically in the ontology of clinical tasks. Hence we focus on those relations specific to the three Semantic Types **Diagnostic Procedure**, **Laboratory Procedure** and **Therapeutic or Preventive Procedure** and on the task of simplifying the relations which exist between these Semantic Types and those Types adjacent to them in the Semantic Network. Our simplification will rest on defining operators which can be used for transforming terms designating adjacent Semantic Types into terms designating one or other of the three Semantic Types on our list. This transformation is needed in part because certain terms representing clinical tasks are missing from the UMLS, so that guideline modelers often use non-task terms in their stead, employing for example "Oxygen mask" instead of "Use of Oxygen mask", or adding prefixes and suffixes to original terms without formally defining them. Our formalization provides a basis for creating an ontology specific to clinical tasks, which can support the development of computer-interpretable clinical guideline models.

We have published a series of papers containing criticism of the UMLS Semantic Network as it is presently structured [6,7], case studies of clinical guideline models [1] and discussion of the task ontologies needed to support such models [8,9,10]. The present communication is an extension of this work and shows how a part of the UMLS Semantic Network can be given a simplified representation and how the result can be used to define operators which are useful for the creation of task ontologies.

In section III, we compare our own adaptation of the UMLS Semantic Network with the representations proposed by other groups. In section IV, we deal with the issue of restricting the UMLS Semantic Network to just those Semantic Types needed to support the construction of task-based computer-interpretable guideline models. In section V we discuss the implications of our formalization. And in an appendix we outline some fundamentals of graph theory presupposed in the main text.

II. REPRESENTATION OF THE UMLS SEMANTIC NETWORK

The works of Chen, Perl *et al.* and of Geller, Perl *et al.* are important attempts to reduce the complexity of the Semantic Network. [11,12,13] In order to provide a simplified representation, they define the following types of entities:

Semantic Type Group: An abstract conceptual entity comprising the set of all Semantic Types standing in the exact same set of relationships.
Semantic Type Collection: An abstract conceptual entity representing a set of Semantic Types that is cohesive in virtue of its possession of a unique root.

Based on these concepts, they describe an "Induced Subnetwork" (see Appendix and [14]), a subgraph the Network which simplifies the visual representation of the content of the Network in a way which, as their evaluation shows, supports human understandability. [14]

The following relations (edges) exist within the Semantic Network:

Neuroreactive Substance or Biogenic Amine is-a$_{UMLS}$ **Biologically Active Substance**
Hormone is-a$_{UMLS}$ **Biologically Active Substance**
Enzyme is-a$_{UMLS}$ **Biologically Active Substance**
Vitamin is-a$_{UMLS}$ **Biologically Active Substance**
Immunologic Factor is-a$_{UMLS}$ **Biologically Active Substance**
Receptor is-a$_{UMLS}$ **Biologically Active Substance**

All of the classes mentioned in this list stand in the same three relations –**analyzes**, **assesses_effect_of** and **measures** – with the class **Diagnostic Procedure**. Thus in the case of **Biologically Active Substance** we have:

Diagnostic Procedure analyzes$_{UMLS}$ **Biologically Active Substance**
Diagnostic Procedure assesses_effect_of$_{UMLS}$ **Biologically Active Substance**
Diagnostic Procedure measures$_{UMLS}$ **Biologically Active Substance**

These same three relations are repeated within each subclass of **Biologically Active Substance**, as shown in Figure 1.

This enables Chen, Perl *et al.* and of Geller, Perl *et al.* to simplify their representation using induced subgraphs. Unfortunately however each of their induced subgraphs still represents all of the different kinds of edges between the various Semantic Types involved and thus the results of their work still remain complex for many purposes. Above all, they bring the limitation that not all the nodes within the network can be taken into account at the same time in an induced subgraph. Since all the terms within the UMLS are linked to Semantic Types, each loss of a Semantic Type brings the the loss of all terms associated with that Type. This is an undesirable feature for the purposes of creating clinical task ontologies.

Computer-interpretable clinical guideline models are constructed on the basis of guidelines formulated as free text. The latter are interpreted by experts, and ontologies are built using the UMLS or some other standard terminology system. These ontologies are then used in association for example with rule-based engines designed to execute queries in such a way as to point the user to the clinical tasks to be performed in given circumstances. [1,8,9,10] Various kinds of ontologies are needed for such models, including general ontologies representing entities such as *clinical task* and *healthcare organization*, as well as more specific ontologies representing the kinds of entities referred to in each given guideline text. The source text of clinical practice guidelines can include terms which belong to any of the UMLS Semantic Types. Thus to interpret such texts we need somehow to have access to all of the vertices of the UMLS Semantic Network. We now outline a Minimal Spanning Subnetwork of the Network (see Appendix), which is optimized for use specifically in the context of computer interpretable guideline models. This representation preserves all the nodes within the Network and yet substantially reduces the number of edges connecting them.
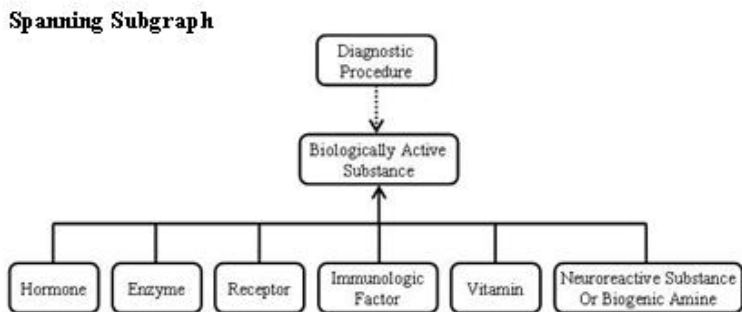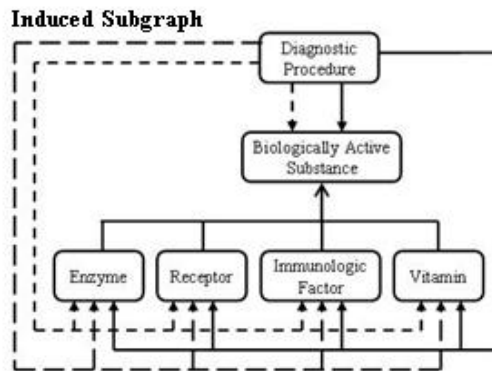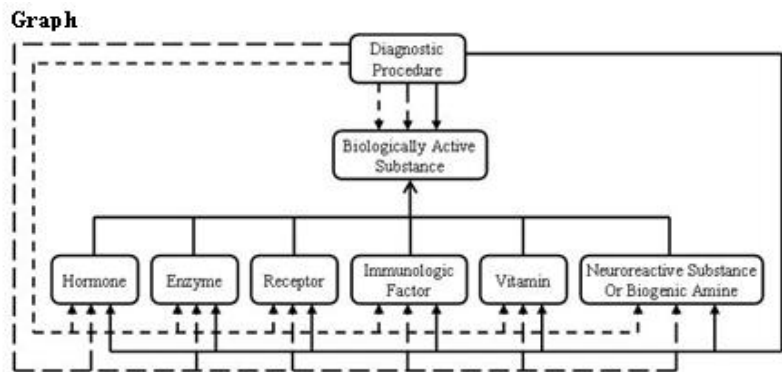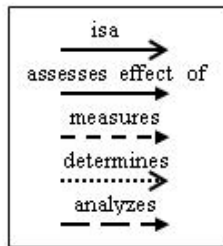
Figure 1. Representation of the Graph, Induced Subgraph and Spanning Subgraph for the Semantic Type Collection with root class **Biologically Active Substance**

III. RESTRICTION OF THE UMLS SEMANTIC NETWORK FOR TASK-BASED GUIDELINES

Almost all of the actions suggested in guidelines can be represented in terms of subclasses of just three Semantic Types, namely: **Laboratory Procedure**, **Diagnostic Procedure** and

Therapeutic or Preventive Procedure. [8] These Semantic Types are themselves subclasses of the Semantic Type **Health Care Activity**. The definitions are as follows:

> **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.

Other Semantic Types used less frequently in guidelines are **Occupational Activity** and its subtypes **Educational Activity**, **Governmental or Regulatory Activity** and **Research Activity**. Even these, however, are related to **Health Care Activity**. For example, **Research Activity** helps to determine the **Health Care Activity** by giving Strength of Evidence.

Our three selected Semantic Types can now be used as the basis for the creation of a Minimal Spanning Subnetwork which enables us to realize the goal of simplifying the UMLS Semantic Network by connecting every other Semantic Type to these three Semantic Types. We take into consideration the Semantic Type Collections already defined by Chen, Perl *et al.* and of Geller, Perl *et al.* and study the relations mentioned in the UMLS Semantic Network between these Semantic Type Collections and the three Semantic Types **Diagnostic Procedure**, **Laboratory Procedure** and **Therapeutic or Preventive Procedure**.

We found that many of the edges which connect Semantic Types within a given Semantic Type Collection to Types outside the Collection are sufficiently similar that they can be replaced by a single type of edge. Based on this idea, we formally defined operators which would enable the mapping of other Semantic Types to our three selected types.

IV. FORMALIZED RELATIONS

In the January 2003 edition of UMLS Semantic Network we find 6718 edges between the different Semantic Types, each edge representing one or other of the 54 distinct relations present within the network. When we consider just those relations connecting the three Semantic Types of concern to us here to their adjacent nodes, we found 179, 179 and 104 relations for **Diagnostic Procedure**, **Laboratory Procedure** and **Therapeutic or Preventive Procedure** respectively. This means that these relations comprise in all some 6.87% of the total number of edges linking nodes in the UMLS Semantic Network, with values of 2.66%, 2.66% and 1.55% for the three Semantic Types mentioned.

Of the 134 Semantic Types in the UMLS Semantic Network 79, 67 and 51 are adjacent to **Diagnostic Procedure**, **Laboratory Procedure** and **Therapeutic or Preventive Procedure** respectively. Thus, the percentages of all the nodes standing in relations to these three Semantic Types are 60.8%, 50.8% and 38.6%, putting the mean at 50.1%.

These numbers signify that, even without considering simplifications in terms of Semantic Type Collections or other varieties of subnetworks, we are able, with a mere 6.87% of the total 6718 edges, to take into account the relations which exist between half of all the Semantic Types in the network.

However, the 6.87% of edges which we consider still manifest many similarities in their definitions. For example, 'analyzes' and 'measures' in

Diagnostic Procedure analyzes_{UMLS} **Chemical**
Diagnostic Procedure measures_{UMLS} **Chemical**

can for many purposes be identified. Since these relations account for over 100 edges for each of the three Semantic Types considered, far more than would be useful for the purposes of defining a small and manageable list of operators, we can postulate the following rule for simplification of the Semantic Network and for formalizing operators:

> **Rule for relational inheritance**: *If a given relation holds for the root class in a Semantic Type Collection, then assume the relation holds for all its children, and thus for the entire Semantic Type Collection.*

We have derived this rule from examining how the subsumption (is-a) relation is formally defined (for details, see [15,16,17]). Each Semantic Type (other than the root class) within a given Semantic Type Collection stands in this relation to the root class within that collection, according to the usual definition:

$$R1: A \text{ is-a } B =_{def} \forall x(inst(x,A) \rightarrow inst(x,B))$$

> Example: **Hormone** is-a_{UMLS} **Biologically Active Substance**
> Interpretation: All individual instances of **hormone** are instances of **biologically active substance**.

We can then assert by contraposition:

$$R2: A \text{ is-a } B \rightarrow not \exists x(inst(x,A) \& not\ inst(x,B))$$

> Example: **Hormone** is-a_{UMLS} **Biologically Active Substance**
> Interpretation: There are no instances of **hormone** which are not instances of **biologically active substance**.

We can also assert that whenever some relation **R** holds between two classes A and B, then some counterpart relation R will hold between corresponding individual instances. The precise nature of this entailment differs for different types of relations. For some relations it is as follows:

$$R3: A \textbf{ R } B \rightarrow \forall x\ (inst(x,A) \rightarrow \exists y(inst(y,B) \& R(x,y))),$$

which holds where **R** is, for example, the relation *part-of* holding between classes, and R is the instance-level part relation holding between specific individual instances of those classes. [15]

We can assert also:

$$R4: A \textbf{ R } B \rightarrow \exists x\exists y\ (inst(x,A) \& inst(y,B) \& R(x,y))$$

For example if **R** is the relation *analyzes* holding between **Diagnostic Procedure** and **Biologically Active Substance** then R4 asserts that some instance of diagnostic procedure is an analysis of some instance of biologically active substance.

From **Diagnostic Procedure** analyzes_{UMLS} **Biologically Active Substance** and **Hormone** is-a **Biologically Active Substance** it can be inferred that **Diagnostic Procedure**

analyzes_{UMLS} **Hormone**. In this way the relevant relation is inherited when we move from some given class as target of **Diagnostic Procedure** to a subclass subsumed thereby.

As we saw in II, **Diagnostic Procedure** stands in three relations to each of the classes within the Semantic Type Collection **Biologically Active Substance**, making altogether 21 (7*3) relations or edges within the subgraph which contains the 7 classes within this collection together with the class **Diagnostic Procedure** itself. Adding the six subsumption relations present within the root class **Biologically Active Substance** and its subclasses, the total number of edges in this subgraph become 27 (21+6). By applying a synthesis in accordance with our rule for relational inheritance, we are able to reduce these relations or edges to 9 (3+6), 3 holding between **Diagnostic Procedure** and **Biologically Active Substance** and 6 holding between **Biologically Active Substance** and its subclasses.

But we can take the simplification still further. For on examining their definitions we see that the following three relations are in fact similar:

analyzes: Studies or examines using established quantitative or qualitative methods.
assesses_effect_of: Analyzes the influence or consequences of the function or action of.
measures: Ascertains or marks the dimensions, quantity, degree, or capacity of.

From the definition of **analyzes** and the fact that **analyzes** is a relation subject to (R4), we can infer further that:

R5: A analyzes B → ∃x∃y(inst(x,A) & (inst(y,B) & (studies(x,y) or examines(x,y))
& (uses-quantitative-methods(x) or uses-qualitative-methods (x)))

From the definitions of **analyzes** and **assesses_effect_of**, we can derive

R6: ∀x∀y(**assesses_effect_of**(x,y) → **analyzes**(x,y))

Unfortunately terms such as "influence of the function" or "marks", used to define the relations in R6, are themselves not defined within the UMLS, so that the analysis cannot be carried further. It would be useful to this end to have such terms defined, for example through connection to a generic ontology such as Wordnet. [19, 20]

Indeed, in addition to the relations mentioned above, there are also other relations within the Network with definitions similar to those of **analyzes**, for example:

evaluation_of: Judgment of the value or degree of some attribute or process.
diagnoses: Distinguishes or identifies the nature or characteristics of.

Within the UMLS Semantic Network, **evaluation-of** has an inverse relationship: **evaluates**. We consider **evaluates** instead of **evaluation_of**, since within the relations where **evaluation_of** exists, together with the three Semantic Types representing tasks, these Semantic Types are on the right hand side, unlike in other relations where these Semantic Types are on the left. The Network does not define the inverse relations and only mentions them in relations. Thus, there is no defintion of evaluates present within the network. The existing relations are:

Qualitative Concept evaluation_of_{UMLS} **Diagnostic Procedure**
Qualitative Concept evaluation_of_{UMLS} **Laboratory Procedure**
Qualitative Concept evaluation_of_{UMLS} **Therapeutic or Preventive Procedure**

Considering the Network considers evaluates as an inverse of evaluation_of, we interpret these relations as:

Diagnostic Procedure evaluates$_{UMLS}$ **Qualitative Concept**
Laboratory Procedure evaluates$_{UMLS}$ **Qualitative Concept**
Therapeutic or Preventive Procedure evaluates$_{UMLS}$ **Qualitative Concept**

Thus altogether there are five relation terms with similar definitions, namely, **analyzes**, **assesses_effect_of**, **measures**, **evaluates** and **diagnoses**. This suggests that we introduce a new relation, **determines**, defined by:

R7. A **determines** B = $_{def}$ A **analyzes** B or A **assesses_effect_of** B
or A **measures** B or A **evaluates** B or A **diagnoses** B

and satisfying:

R8: A **determines** B → ∀x (inst(x,A) → ∃y(inst(y,B) & (**analyzes**(x,y) or
assesses_effect_of(x,y) or **measures**(x,y) or **evaluates**(x,y) or **diagnoses**(x,y))))

We can then introduce a term-forming operator "Determination of" (DOF, for short), which satisfies the following axiom:

R9: A = DOF(B) ↔ A **determines** B.

Every time we encounter one or other of the five relations we can now substitute a term formed by applying the operator 'DOF' to the relevant target. DOF is, as it were, the generic guideline task of determining something by analyzing, measuring, etc., some target object.
We can similarly introduce other operators based on other Semantic Network relations, for example: COF (for: Cause of), UOF (for: Use of), MOF (for: Management of). (For definitions of the UMLS relations employed in defining these operators, see Table 1.)

Table 1 Some of the UMLS Relations connecting **Diagnostic Procedure** *to its adjacent nodes.*

UMLS Relations	Definition
affects	Produces a direct effect on. Implied here is the altering or influencing of an existing condition, state, situation, or entity. This includes has a role in, alters, influences, predisposes, catalyzes, stimulates, regulates, depresses, impedes, enhances, contributes to, leads to, and modifies.
complicates	Causes to become more severe or complex or results in adverse effects.
prevents	Stops, hinders or eliminates an action or condition.
result_of	The condition, product, or state occurring as a consequence, effect, or conclusion of an activity or process. This includes product of, effect of, sequel of, outcome of, culmination of, and completion of.
treats	Applies a remedy with the object of effecting a cure or managing a condition.

uses	Employs in the carrying out of some activity. This includes applies, utilizes, employs, and avails.

Thus for example MOF satisfies:

R10. A = MOF(B) ↔ A treats B or A prevents B.

These term-forming operators can now be used together with appropriate non-task terms in order to represent the corresponding tasks. For example, in the case of "Hyptertension", which belongs to the UMLS Semantic Type **Disease or Syndrome**, we can use "MOF Hypertension" to designate a task belonging to the Semantic Type **Therapeutic or Preventive Procedure**.

Going through this exercise provides two clear advantages.

1. The derived terminology helps to create an ontology consisting purely of clinical tasks, which are at the same time still related in perspicuous fashion to other Semantic Types. It serves to focus the ontology on just one aspect or partition of the vast domain covered by the Semantic Network, namely that of clinical tasks. We have claimed elsewhere that the creation of such partition-specific ontologies can help to make deductions possible which would otherwise be lost, to detect existing mistakes in existing terminology systems, and to provide a clear representation of the underlying knowledge. [18, 21]

2. Our approach clearly reduces the number of edges within the Semantic Network. For example, in the case of Semantic Type Collection with root class **Biologically Active Substance**, the 27 edges connecting **Diagnostic Procedure** to other Semantic Types can be reduced to just 7. Use of DOF leads to a reduction from 121 to 49 edges for **Laboratory Procedure** and from 106 to 45 edges for **Diagnostic Procedure** (a total 56.83% reduction). Use of COF leads to a reduction from 25 to 19 edges for **Laboratory Procedure**, from 26 to 20 edges for **Diagnostic Procedure**, and from 39 to 19 edges for **Therapeutic or Prventive Procedure** (a total 35.66% reduction). Use of MOF leads to a reduction from 17 to 11 edges for **Therapeutic or Preventive Procedure** (a 35.29% reduction). Altogether, the number of required edges is reduced by half as a result of applying the formal methods defined above. (For an overview of the formalized relationships, see figure 2.)

Our methods will also work in application to other relations connecting nodes adjacent to our three Semantic Types which have been not been taken into account here since they do not have a direct role within task-based guideline ontologies, for example: **location_of**, **issue_in**, and **associated_with**.
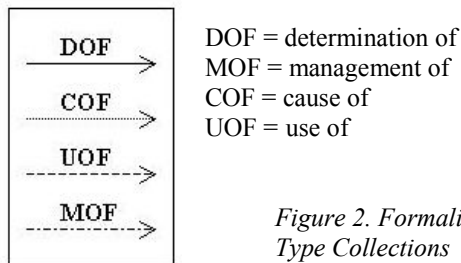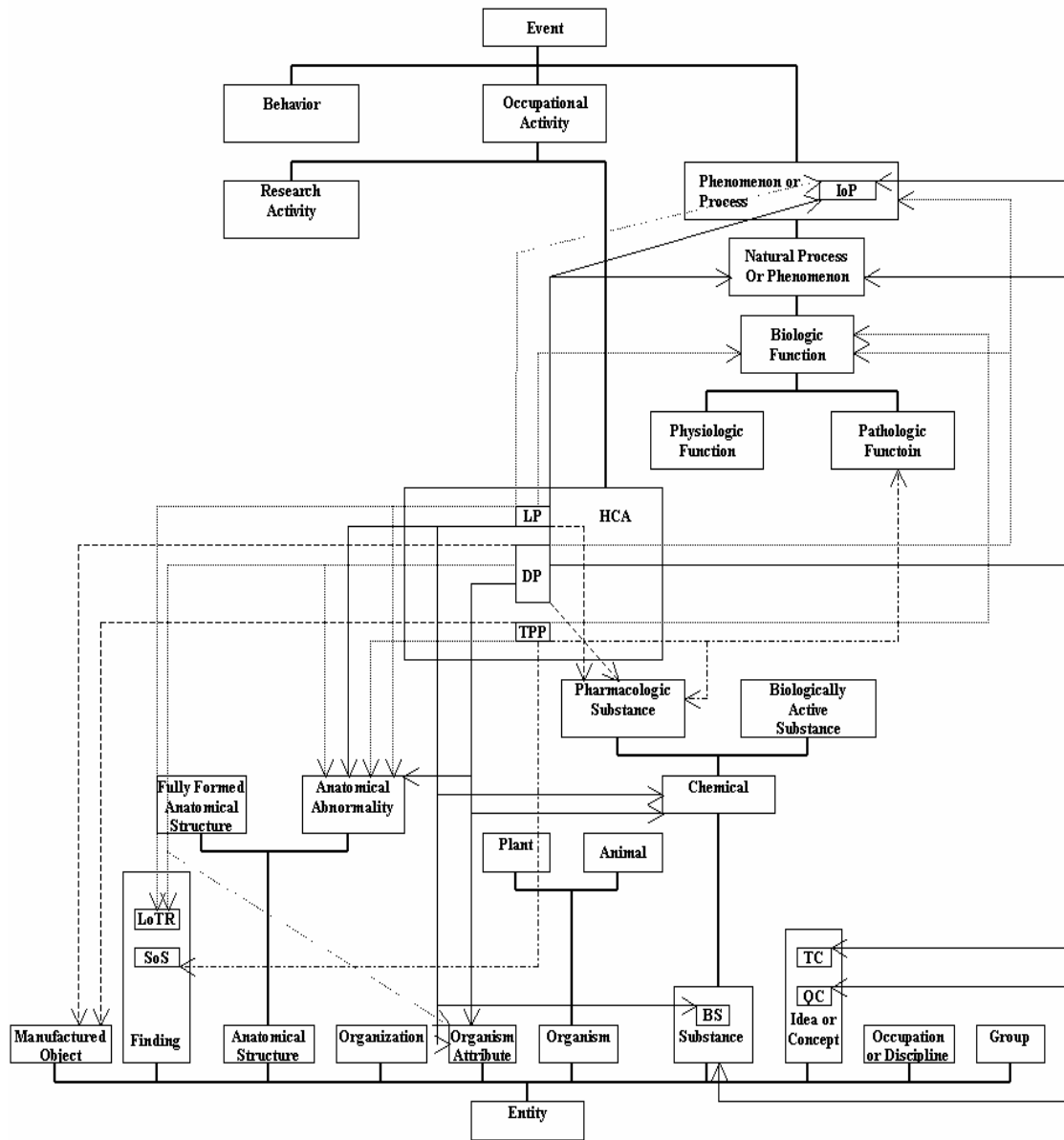
Figure 2. Formalized relationships described between the different Semantic Type Collections

V. DISCUSSION

Task ontologies form an important part of computer-interpretable guideline models. Such networks require a formal ontology which classifies tasks and provides an account of the parthood relations between complex tasks and the atomic tasks which are their components. Standard terminologies like UMLS deal with medical terms in a very general manner; here we have formalized relations which can help in creating an ontology specific to the clinical task domain.

Our formalization is independent of specific guideline modeling environments and it can thus be used as a tool to navigate between them. The methodology is furthermore not terminology-specific. Thus while in the application above we take terms and definitions from the standard UMLS terminology, the methodology can also be used where terminology does not exist within UMLS but needs to be created. For example, while **cardiac failure** exists within UMLS, "cardiac failure determination" is absent. While **Atenolol** exists, "Use of Atenolol" is absent. Such derived terms are often needed in work on task ontologies. As we saw, one common solution to this problem is to use a simple term like "Atenolol" where what is intended is in fact "Use of Atenolol". While this approach can be perfectly appropriate for human experts, it leads to problems when task ontologies are used for purposes of automatic data integration across patient records or in other contexts where human interpretation is not possible. Another approach is to use *ad hoc* operators introduced outside of any unifying formal framework. While this approach can work in relation to small models where only a small group of users are involved, such *ad hoc* measures do not work on a larger scale. It is for this reason that one needs to define operators in a formal way.

We have reduced the complexity of the UMLS Semantic Network by using derivations formalized in predicate logic. Given that each clinical guideline needs to be modeled manually, the possibility of performing this task with only a few specific relations and operators brings considerable benefits.

From the point of view of ontological structure, we have provided an untangled partition: an ontology built by taking into account just one single aspect of a complex domain. [18] This, too, brings computational benefits. There is an ongoing discussion within ontology, related to the issue of single vs. multiple inheritance, concerning the question whether subsumption relations within a single ontology should be based on just one or on a multiplicity of different types of criteria. **Neoplasm of colon**, for example, is usually represented as standing in an is-a relation both to **Disease of colon** and to **Neoplasm**. The first of these is a subsumption based on location, the second is based on pathology. If all such is-a relations are treated on a par, which means: if the underlying criteria are not recorded, then this leads to loss of knowledge, and to a polysemous use of 'is-a' that is often accompanied by coding errors. [22]

To see how this looks in relation to our present topic, we note that there are different ways in which tasks can be classified. Thus we can classify the tasks carried out within a particular healthcare organization based on the agents who perform them, on the hierarchy of the organization, on work schedules, and so on. The result is then quite different from the generic task ontology based on how tasks are represented within a clinical practice guideline text. [9,10] Our recommendation is that such heterogeneous ontologies should not be run together within a single structure, but rather that they should be constructed separately and associated with a technology for navigating between them. We have offered a framework for the creation of such untangled ontologies in the foregoing, with special reference to the domain of tasks as represented within clinical guideline texts.

Appendix: FUNDAMENTAL TERMS OF GRAPH THEORY

Graph: A graph is a pair G = (V,E) of sets satisfying E $\subseteq$ V × V; thus the elements of E are two-element subsets of V. The elements of V are the vertices or nodes of the graph G, and the elements of E are its edges [14].

Subgraph: Let G(V,E) and G′ = (V′,E′) be two graphs. If V′ $\subseteq$ V and E′ $\subseteq$ E, then G′ is a subgraph of G, represented as G′ $\subseteq$ G [Figure 1].

Induced subgraph: If G′ $\subseteq$ G, and if V′ $\subseteq$ V is such that for all pairs x, y of elements in V′, G′ contains the corresponding edges <x,y> from G, then G′ is the subgraph of G *induced by* V.

Spanning subgraph: A graph G′ is a spanning subgraph of graph G if G′ $\subseteq$ G and V′ = V (i.e. the vertices of G′ includes all the vertices of G).

Minimal spanning subgraph: A minimal spanning graph is a spanning subgraph such the sum of all edge costs is minimal, where *edge cost* is a quantitative measure of the distance between any two given vertices.

REFERENCES

[1] 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, Stefanelli M. Comparing computer-interpretable guideline models: a case-study approach. J Am Med Inform Assoc. 2003; 10(1): 52-68.

[2] Strang N, Cucherat M, Boissel JP. Which coding system for therapeutic information in evidence-based medicine. Computer Methods and Programs in Biomedicine 2002; 68(1): 73-85.

[3] Humphreys BL, Lindberg DAB, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc. 1998; 5(1): 1-11.

[4] Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993; 32: 281-91.

[5] UMLS website: http://www.nlm.nih.gov/research/umls/

[6] Kumar A, Smith B. The Unified Medical Language System and the Gene Ontology: Some critical reflections. (Lecture Notes in Computer Science 2821). 2003; 135 – 148.

[7] Schulze-Kremer S, Smith B, Kumar A. Revising the UMLS Semantic Network. Medinfo 2004. (In press)

[8] 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. Stud Health Technol Inform. 2003; 95: 469-74.

[9] Kumar A, Smith B. Ontology for task-based clinical guidelines and the theory of granular partitions. Artificial Intelligence in Medicine, Cyprus, 19-22 Oct. 2003.(Lecture Notes in Artitificial Intelligence 2780), 71-75.

[10] Kumar A, Ciccarese P, Smith B, Piazza M. Context-based task ontologies for clinical guidelines, in DM Pisanelli (ed.), Ontologies in Medicine: Proceedings of the Workshop on Medical Ontologies, Rome October 2003, Amsterdam: IOS Press. (In press).

[11] Geller J, Perl Y, Halper M, Chen Z, Gu H. Evaluation and application of a semantic network partition. IEEE Trans Inf Technol Biomed. 2002; 6(2): 109-15.

[12] Chen Z, Perl Y, Halper M, Geller J, Gu H. Partitioning the UMLS Semantic Network. IEEE Trans Inf Technol Biomed. 2002; 6(2): 102-8.

[13] Gu H, Perl Y, Halper M, Geller J, Kuo F, Cimino JJ. Partitioning an object-oriented terminology schema. Methods Inf Med. 2001; 40(3): 204-12.

[14] Diestel R. Graph theory. New York: Springer Verlag. 2000.

[15] Smith B, Rosse C. The role of foundational relations in the alignment of biomedical ontologies. Medinfo 2004. (In press).

[16] Smith B and Mulligan K, Framework for formal ontology, Topoi, 3, 73-85, 1983

[17] Smith B. The logic of biological classification and the foundations of biomedical ontology. In: Westerståhl D, editor. Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science, Oviedo, Spain, 2003: Elsevier-North-Holland; 2004 (In press).

[18] Bittner T, Smith B. A theory of granular partitions. In: Duckham M, Goodchild MF, Worboys MF, editors. Foundations of Geographic Information Science. London: Taylor & Francis; 2003. 117-151

[19] WordNet. An Electronic Lexical Database, C Fellbaum (ed.). MIT Press. 1998.

[20] http://www.cogsci.princeton.edu/~wn/papers.shtml

[21] Ceusters W, Smith B, KumarA, Dhaen C. Ontology-Based Error Detection in SNOMED-CT®. Medinfo 2004. (In press).

[22] Smith B, Köhler J, Kumar A. On the application of formal principles to life science data: A case study in the Gene Ontology. (Lecture Notes in Bioinformatics 2994). (In press).

Contact Information
Anand Kumar, IFOMIS, Faculty of Medicine, University of Leipzig, 16-18 Haertelstrasse, Leipzig 04107, Germany. Email: anand.kumar@ifomis.uni-leipzig.de