WWW-available Conceptual Integration of Medical Terminologies: the ONIONS Experience

Domenico M. Pisanelli, Aldo Gangemi, Geri Steve
Consiglio Nazionale delle Ricerche – Istituto Tecnologie Biomediche – Roma, Italy
{nico,aldo,geri}@color.irmkant.rm.cnr.it

We present the most applicable aspects of our research in the conceptual integration of terminologies. From past experience, we claim that the conceptualizations provided for terminological ontologies need to be philosophically and linguistically grounded. We developed ONIONS, a methodology for integrating domain terminologies by exploiting a library of generic ontologies. Our current focus is on flexible and cooperative modelling of terminological ontologies. We adopt modular and negotiable architectures of ontologies and some WWW-oriented tools, such as Ontolingua and Ontosaurus.

1. INTRODUCTION

Our research primarily concerns the integration and reuse of terminological ontologies in medicine. Terminological ontologies are crucial for activities such as vocabulary standardization[1], terminology server design[7,12], conceptual modelling of subdomains[17], and knowledge integration, sharing, and reuse[9,4,22].

A terminological ontology is defined as an explicit conceptualization of a vocabulary[24]. An explicit conceptualization specifies a term by providing some formal constraints to it.

First- or second-order logics are possible candidates as a formal language for ontologies, however some specialized languages have become available which are appropriately designed to support the activity of writing, browsing, editing, and revising an explicit conceptualization via WWW, as well as to support the creation of complex ontology libraries, with functional constraints among the component ontologies[5,22].

Van Heijst[24] distinguishes terminological ontologies in *representation*, *generic* and *domain* ontologies:

- *representation* ontologies specify the conceptualizations that underly knowledge representation formalisms[11];
- *generic* ontologies concern general, foundational aspects of knowledge: processes, part/whole structure, connexity, kinds of objects, quantities;
- *domain* ontologies specialize a subset of generic ontologies in a domain or subdomain: medicine, cardiology, clinical administration, protocols, etc.

It has been a common practice to define domain ontologies from scratch, i.e. without adopting formal methods of integration with generic ontologies. Good examples of such a kind of ontologies are: the UMLS Semantic Network (UMLS-SN)[12] and the "Core Model" developed in the GALEN Project (GCM)[16]. Some restricted generic and domain terminological ontologies have been defined as part of KBS development within the GAMES-II Project[4] and for Protégé-II[9].

In this paper we present our approach to build a library of generic ontologies and to explicitly conceptualize the domain ontologies of terminology systems in medicine.

A terminology system is a system which organizes the terms of a domain according to some structure, such as hierarchical indentations, codification or lookup tables.

Most medical terminology systems do not have a terminological ontology, for example ICD10[25], the Gabrieli Medical Nomenclature (GMN)[6], SNOMED-III[2]. This does not mean that terminology systems are not founded on a conceptualization, but only that their conceptualization is left to the interpretation of the experts who use the systems.

We also give a short description of our current ON9 ontology library, integrated from the terminological ontologies of five terminology systems. Finally, we provide an insight of the tools that fit our current efforts, which are focused on *cooperation* and *negotiation* in modelling ontologies.

2. THE ONIONS METHODOLOGY

Conceptualization poses severe problems to modellers when concepts must be shared by different users in different contexts. Local conceptualizations are not suitable to support the tasks of making standards, writing guide-lines, reusing, integrating or sharing knowledge.

To this aim we need: 1) to capture all and only the terminological knowledge needed for a purpose; 2) to explicitate the intended meaning of local conceptualizations; 3) to reuse or formalize generic ontologies; 4) to explicitate the intended meaning of the "meta-level categories" (e.g. "class", "slot", "property"); 5) a common language for expressing the resulting conceptualization.

We examine these issues in more detail in other papers[20,21]. A relevant effort in the direction of 5) has been done by Gruber[10]. The problems in 4) have been studied by several authors[8,10,11,18]. The issues in 1) and 5) have received little attention in AI, apart from our approach[20] and Uschold and King[23].

In medicine an important effort has being done by the European standardization body (CEN) which addresses issues 1) through 4) at various degrees of depth. The CANON group[3] mainly addressed 5); the MoSe pre-standard[1] provides some guidelines for 1) and 5), other groups are writing standard conceptual systems in medical sub-domains dealing with the issues in 2)[17]. The issue 3) has been treated by several authors[20,26,27] and it is mainly found in the AI, linguistics and philosophy literature.

A main concern of our research is to provide a terminological ontology to the most important terminology systems in medicine. To this purpose, we developed a methodology which addresses the above issues.

We initially defined ONIONS as a methodology to build the core model of a medical terminology server in the context of the GALEN Project[7]. Later ONIONS was revised with the goal of:[20]

- building (or re-using) a library of generic ontologies by formalizing ontologies from the literature in AI, philosophy, linguistics, cognitive science;
- generating a domain ontology for each terminology system.

ONIONS led to the successful integration of five terminology systems: the UMLS-SN (about 170 semantic types and relations, and their 890 defined combinations), SNOMED-III (about 600 most general concepts), GMN (about 700 most general concepts), ICD10 (about 250 most general concepts), and the GCM- 5g (about 2000 items).

We adopted Ontolingua[10] – a language developed from KIF[15] – as a formalism for representing the library of ontologies. It allows expressivity for both frame-like and axiomatic constraints.

The ONIONS methodology consists in five phases. Each one – labelled Mi – makes a terminology system evolve from a state Si into a state Si+1. Depending on the purpose of the integration, a terminology system may reach a "less evolute" state – e.g. S4 – and stay there without needing a further evolution.

Given in input a list of valid expressions, i.e. a terminology system in a state S1 (e.g. a flat list of terms), the phase M1 consists in the individuation of a possible taxonomic structure. The output of this phase is an S2 system, i.e. the flat list enriched with a taxonomy. Many 'classic' terminology systems (e.g. ICD10) can be classified as S2 systems.

In the phase M2 one must individuate the reason why the IS_A links in the taxonomy S2 exist and the reason why a term differs from its closest relatives in the taxonomy. For example, what distinguishes a "viral hepatitis" from the parent "hepatitis" is the fact that the viral hepatitis is due to a viral infection. Most vocabularies, glossaries and semantic nets in medicine fall into this category (e.g. UMLS-SN).

The methodological phase M3 consists in the formalization of the elements and the constraints in the S3 descriptions. The goal is to obtain S4 systems, i.e. systems in which the conceptual principles which motivate the description are formalized. In the example "viral hepatitis", the differentia "viral" calls for the formalization of some generic ontologies which specify axioms about organisms and causality.

The methodological phase M4 consists in the construction (or the reuse, if available) of a library of generic ontologies to account for the requirements in S4. This activity aims at building a motivated, well-grounded *top-level* to the extent the integration of the terminology systems needs. This phase leads to an S5 system. For example, "viral" can be defined in ON9 (see §3) by including and specializing the generic ontologies: "biologic objects" and "actants" (for causality). In this phase it is also given an assignment of meta-level categories to classes and relations defined.

The methodological phase M5 consists in the implementation of a domain ontology in a system which allows automatic classification.

At state S6 a domain ontology is implemented in an automatic classifier, e.g. Loom[13]. If we export a library of ontologies into Loom and make it classify the library, the results obtained can be shown by the Ontosaurus WWW interface[22] in order to allow negotiation with experts (see §3).

3. THE ON9 ONTOLOGY LIBRARY

ON9 – the current version of the ontology library – is available on the WWW and has been designed by means of the ONIONS methodology (http://saussure.irmkant.rm.cnr.it/ON9/index.html). Figure 1 shows the inclusion lattice of ON9 ontologies: the representation ontologies provided by default in Ontolingua[10] are the "frame-ontology" and the set of "kif-ontologies"[15]. We defined the ontologies: "metaontology", "semantic-field-ontology" and "structuring-concepts" in order to link the representation ontologies with the generic ontology library. The sets of "structural ontologies" and of "structuring ontologies" contain generic ontologies. They were designed adopting a minimalistic strategy: only some parts of some theories which are useful for the integration process are "bought". For example, given the need of buying some theory of parts and wholes, we chose a subset from the so-called Calculus of Individuals from the philosophical literature and some specific

notions of part from the cognitive science literature, formalizing a theory: "meronymy".
Generic ontologies are variously included in domain ontologies such as "biologic-objects", "clinical activities" and "surgical procedures"[17] (formula 1 reports an excerpt from the last one).
The "integrated-medical-ontology" (IMO) includes the generic ontologies and results from the integration of the terminology systems.
For example the semantic net of UMLS yields the "o-umls" ontology which consists in:
a) taxonomic constraints, for example: "Disease-or-Syndrome" IS_A "Pathologic-Function" which means that the former class is included in the latter one (in Ontolingua this is expressed by the relationship "subclass-of");
b) relationships between classes, for example: "Tissue sorrounds Cell". Such constructs are not provided with an explicit semantics; we can only say that a relationship R holds between instances of class A and instances of class B. The issues if the range of R is constrained by one or more alternative classes, as well as if the variable for instances of class B should be existentially or universally quantified, are not addressed.
"o-umls" and the ontologies derived from terminology systems are then integrated into the IMO by means of the ONIONS methodology.

It should be pointed out that modularity of the library does not prevent one to use in an ontology O concepts from an O' which is not included in O (although they do not become available for a classifier operating on O).
During the development of ON9 and its former versions (ON6-8[19,20]) we experimented several tools. In order to get flexible ontology architectures, we propose that an essential tool for cooperative modelling should be an environment where different modellers could experiment and face each other about the effects of ontological choices on terminology integration, as well as about the constraints posed by terminology integration on ontological choices.

Some tools for ontology editing and classification allow the user to access, to discuss, and to exploit the intermediate byproducts of ONIONS as well as the complete integrated ontology:
• Ontolingua seems to be the most experimented and compatible. We used it to build the ON9 ontology library. Ontolingua models can be made available also by means of a server for ontology browsing and editing with a friendly interface specially designed to support frame-like modelling[5]. The Ontolingua server is centralized and thus not always easy to access.

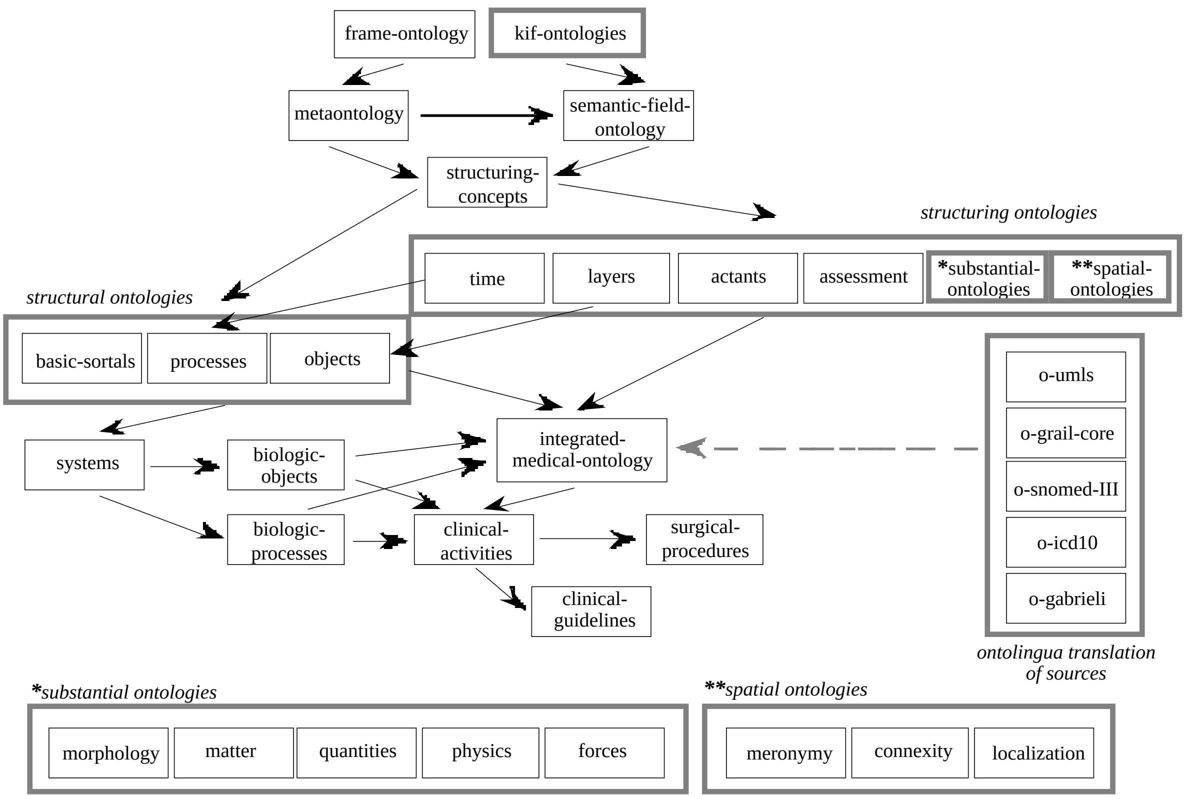

Figure 1. The ON9 library of ontologies (Ontolingua "theories"). Ontologies are represented by black frames. Thick grey frames are sets of ontologies (some explicitely show the elements). The semantics of black arrows is *included-in* (as defined in Ontolingua). The dashed grey arrow means *integrated-in*.

• Ontosaurus[22] is a recent powerful tool, which runs on top of a CL-HTTP server[14] and provides an interface to the Loom knowledge representation system[13](figure 2).

Ontosaurus features both accessibility, a friendly interface, ontology browsing and editing, and an automatic classifier. Groups can share ontologies (there called "contexts") with a quasi-real-time capability of browsing/editing. Within such an environment, ontologies can be negotiated and customized by users who accept a set of ontological definitions or reject some definitions and cooperate in order to define alternative ones.

4. CONCLUSIONS

The ONIONS experience in developing terminological ontologies shows that, from the conceptual integration point of view, ontologies defined using our methodology support:

• formal upgrading of terminology systems: all term classification and definitions are made available in a common, expressive and formal language;
• conceptual explicitness of terminology systems: all (local) term definitions are made available, even though the source does not include them explicitly;
• conceptual upgrading of terminology systems: all term classification and definitions are translated such that they can be included in an ontology library which has a subset constituted of motivated generic ontologies.

From the reuse and maintenance point of view, ontologies produced by means of ONIONS support:
• a motivated ontology library, developed from the integration of authoritative generic and domain sources;

```
(define-class surgical-move (?m)
(1)
  "a surgical-move is a surgical act performed on an object located in one region of an
   organism. It has the goal of having that structure in a different region of the organism.
   The movement is modelled here as position changing from old to new situations"
  :axiom-def (and (subclass-of surgical-move *surgical-act) (!type surgical-move))
  :def (exists      (?p ?r1 ?r2 ?s1 ?s2 ?org1 ?org2)
                (and  (object ?p) (region ?r1) (region ?r2) (situation ?s1) (situation ?s2)
                      (organism ?org1) (organism ?org2) (part ?r1 ?org1)
                      (or (part ?r2 ?org1) (part ?r2 ?org2))
                      (precedes ?s1 ?s2) (is-context-of ?s1 ?m)
                      (=> (is-true-in ?s1 "(and (has-position ?p ?r1) (embodies ?p ?m))")
                          (is-true-in ?s2 "(has-position ?p ?r2)")))))
```

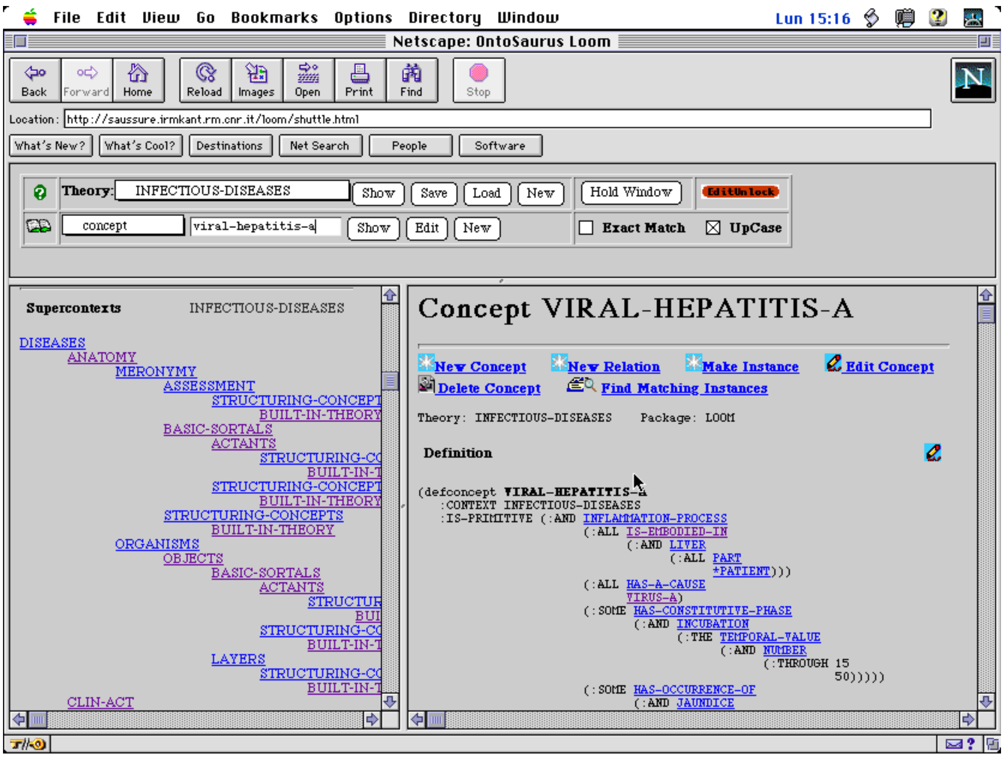

Figure 2. The Loom definition and the related context library for "viral-hepatitis-A" shown by the Ontosaurus.

• specialized domain ontologies which use some subset of ontologies from the ontology library;
• a refinement of the ontology library through the integration of other generic and domain sources.

WWW tools for ontology modelling and validation will play a significant role in the evolution of our libraries. They will allow a better integration and reuse of terminology systems, fostering also the standardization activities in medical terminologies.
There are still some bottlenecks in ontology modelling, mainly due to the necessity of off-line human intervention in the search, choice, and formalization of generic ontologies. In fact, the formalization of system theory (the usual configuration of component-state-event-process), widely available in the engineering domain does not fit the medical domain. If we want to understand the basic principles motivating the conceptualization of terminology in domains such as medicine, we need to refer to theories, such as those provided – usually provided only in informal way – by linguistics, philosophy, and cognitive science.

REFERENCES

[1] CEN prENV 12264:1995, “Medical Informatics - Categorial structure of systems of concepts - Model for representation of semantics”, Brussels, CEN, 1995.

[2] Coté RA, Rothwell DJ, Brochu L, eds. *SNOMED International* (3rd ed.), Northfield, Ill, College of American Pathologists, 1994.

[3] Evans DA, Cimino JJ, Huff SM, Bell DS. “Toward a Medical-Concept Representation Language”, *Journal of the American Medical Informatics Association* 1, 1994.

[4] Falasconi S, Stefanelli M, “A Library of Medical Ontologies”, *Proc. ECAI94 Workshop on Comparison of Implemented Ontologies,* 1994.

[5] Farquhar A, Fikes R, Rice J, “The Ontolingua Server: a Tool for Collaborative Ontology Construction”, *Proc. Knowledge Acquisition Workshop*, 1996.

[6] Gabrieli E, “A New Electronic Medical Nomenclature”, *J.* Medical Systems, 3, 1989.

[7] GALEN Project. Documentation available at the URL: http://www.cs.man.ac.uk/mig/galen

[8] Gangemi A, Steve G, Giacomelli F, “ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration”, *Proc. ECAI96 Workshop on Ontological Engineering,* 1996.

[9] Gennari JH, Tu SW, Rothenfluh TE, Musen M, “Mapping Domains to Methods in Support of Reuse”, *International Journal of Human-Computer Studies*, 41, 1994.

[10] Gruber T, “A Translation Approach to Portable Ontology Specifications”, *Knowledge Acquisition*; 5, 1993.

[11] Guarino N, Carrara M, Giaretta P, “An Ontology of Meta-Level Categories”, *Proc. Knowledge Representation and Reasoning,* Los Angeles, Morgan-Kaufmann, 1994.

[12] Humphreys BL, Lindberg DA, “The Unified Medical Language System Project”, *Proc. MEDINFO 92.*, Amsterdam, Elsevier, 1992.

[13] MacGregor RM, “A Description Classifier for the Predicate Calculus” *Proc. AAAI 94*, 1994.

[14] Mallery JC, “A Common LISP Hypermedia Server”, *Proc. WWW94*, 1994.

[15] Neches R et al, “Enabling Technology for Knowledge Sharing”, *AI Magazine*; 3, 1991.

[16] Rector A, Gangemi A, Galeazzi E, Glowinski A, Rossi-Mori A, “The GALEN CORE Model Schemata for Anatomy: Towards a Re-Usable Application-Independent Model of Medical Concepts”,*Proc. MIE94*, 1994.

[17] Rossi-Mori A, Gangemi A, Steve G, et al, “An Ontological Analysis of Surgical Deeds”, *Proc. AI in Medicine Europe*, 1997.

[18] Sowa JF, “Top-Level Ontological Categories”, *Journal of Human-Computer Studies*, 43, 1995.

[19] Steve G, Gangemi A, Rossi-Mori A, “Knowledge Integration of Medical Terminological Sources”, *Proc. FLAIRS*, 1996.

[20] Steve G, Gangemi A, “ONIONS Methodology and Ontological Commitment of Medical Ontology ON8.5”, *Proc. Knowledge Acquisition Workshop*, 1996.

[21] Steve G, Gangemi A, “Some Theses on Ontological Engineering in the Context of ONIONS Methodology”, *Proc. AAAI97 Spring Symposium Ontological Engineering*, 1997.

[22] Swartout B, Patil R, Knight K, Russ T, “Toward Distributed Use of Large-Scale Ontologies”, *Proc. Knowledge Acquisition Workshop*, 1996.

[23] Uschold M, King M, “Towards a Methodology for Building Ontologies”, *Proc. IJCAI95 Workshop on Basic Ontological Issues*, 1995.

[24] Van Heijst G, Schreiber ATh, Wielinga BG, “Using Explicit Ontologies in KBS Development”, *Int. Journal of Human-Computer Studies*, forthcoming.

[25] WHO, *International Classification of Diseases* (10th revision), Geneva, WHO, 1994.

[26] Grüninger M, “Designing and Evaluating Generic Ontologies”, *Proc. ECAI96 Workshop Ontological Engineering,* 1996.

[27] Borst P, Akkermans H, Top J, “Engineering Ontologies”, *Proc. Knowledge Acquisition Workshop*, 1996.