Integrating Medical Terminologies with ONIONS Methodology

Geri Steve, Aldo Gangemi, Domenico M. Pisanelli
Reparto Informatica Medica, Istituto Tecnologie Biomediche, CNR
Viale Marx 15-I 00137 Roma
{geri,aldo,nico}@color.irmkant.rm.cnr.it

ONIONS helps terminological ontology construction from existing, contextually heterogeneous terminologies. It is a methodology for integrating the context-dependent conceptualizations underlying conceptually heterogeneous terminology systems. We describe an application of this methodology to the medical domain with an example extracted from the UMLS system. We also give a short description of the current ontology library produced by means of ONIONS, and of its metaontology.

INTRODUCTION

Integration of large knowledge bases is a relevant issue and is gaining a constantly raising attention in literature (see for example: [Gruber, 93] [Gennari, Tu, Rothenfluh, Musen, 94] [Neches et al., 91] [Fankhauser, Kracker, Neuhold, 91] [Sujanski, Altman, 94] [EPISTOL, 94] [van Heijst, Schreiber, Wielinga, 97]).

Although much work has been already devoted to integration of data formats and even to integration of representation formalisms, a more challenging integration issue comes from the heterogeneity of intended meaning of concepts.

We introduce ONIONS (ONtologic Integration Of Naïve Sources), a methodology for the integration of terminological knowledge from repositories with heterogeneous conceptualizations. This methodology is founded on a philosophically account to a semantic theory (i.e. our theory of meaning) which drives the structure of the methodology itself [Gangemi, Steve, Rossi Mori, 95], [Steve, Gangemi, 96], [Gangemi, Steve, Pisanelli, Giacomelli, 97].

ONIONS creates a common framework to generalize and integrate the definitions that are used to organize a set of terminological sources. In other words, it allows to work out coherently a domain terminological ontology (a terminological ontology is usually defined as the explicit conceptualization of a vocabulary) for each source, which can be then compared with the others and mapped to an *integrated ontology library*.

Terminological ontologies are commonly distinguished into domain, generic and representation ontologies:

- domain ontologies concern specialized knowledge in a domain or subdomain: medicine, cardiology, clinical administration, protocols, etc. Some examples concern concurrent engineering, planning, medicine, law, military applications [McGuire, Kuokka, Weber et al, 93] [Tate, 96] [Gangemi, Steve, Giacomelli, 96] [Rossi Mori, Gangemi, Steve, et al., 97] [Valente, Breuker, 96] [Swartout, Patil, Knight, Russ, 96]. Also, some restricted generic and domain terminological ontologies have been defined as part of KBS development within the projects GAMES-II [Falasconi, Stefanelli, 94], Protégé-II [Gennari, Tu, Rothenfluh, Musen, 94], Kactus [Martil, Turner, Terpstra, 95] [Laresgoiti, Anjewierden, Bernaras, et al., 96].
- generic ontologies concern general, foundational aspects of knowledge: processes, part/whole structure, connexity, kinds of objects, quantities; good examples are [Sowa, 95] [Varzi, 96] [Borgo, Guarino, Masolo, 96] [Cohn, Randell, Cui, 96] [Gerstl, Pribbenow, 96] [Borst, Akkermans, Top, 97];
- representation ontologies specify the conceptualizations that underly knowledge representation formalisms [Guarino, Carrara, Giaretta, 94]; the frame-ontology [Gruber,

93] defined by the designers of Ontolingua is a typical example; a representation ontology is also considered a metaontology (it defines "meta-level" categories, §4.);

ONIONS has been applied to medicine, that can be itself seen as the integration of many heterogenous subdomains. The current resulting medical ontology is ON9 (§3.).

Our research is related to others in the same domain, such as vocabulary standardization [CEN, 95], natural language (lexical) processing [Bateman, 97], terminology server design [GALEN, 94] [Humphreys, Lindberg, 92] [Rector, Gangemi, Galeazzi, Glowinski, Rossi Mori, 94], conceptual modeling of subdomains [Rossi Mori, Gangemi, Steve, et al., 97], knowledge integration, sharing, and reuse [Evans, Cimino, Huff, Bell, 94] [Gennari, et al, 94] [Falasconi, Stefanelli, 94] [Swartout, Patil, Knight, Russ, 96] [Valente, Breuker, 96], [Steve, Gangemi 97] and multi-agent system development [Falasconi, Lanzola, Stefanelli, 96].

1. SOME PREMISES

There exist very few methodologies for terminological ontology construction (for example, Uschold, King, 95] [Valente, Breuker, 96]). In fact, in recent years ontological frameworks concentrated mainly on formal languages for expressing ontologies (*ontology representation*). We argue that the main reserve in discussing the methodology issue is linked to the difficulty of solving such problems as *context dependency*, *conceptual relevance* and *detail level*. We briefly summarize an assessment to these problems (an extensive discussion is in [Gangemi et al, 97]).

Context-dependency The analysis conducted on medical terminologies has shown that the conceptualization of a term is context-dependent, where "context" has to be intended in a wide sense, including:

- the text in which the term appears;
- the possible terminological repositories in which the term is classified.
- the historical evolution of the use of the term;
- the spatio-temporal situation in which an agent uses the term; etc.
- the belief space of an agent which uses the term;
- the particular viewpoint an agent chooses on the use of the term;
- the disciplinary domain in which the term mainly occurs;
- the particular viewpoint inside a domain (e.g. morphology, physiology in medical domain);

For example, in the definition of *viral hepatitis*: «inflammation of liver caused by virus», we have to consider that *inflammation* may mean in different — or within a same — terminological source:

- a *physiological function* performing segregation of external agents;
- a *portion of a body part* which embodies that physiological function;
- a specific *abnormal morphology* (texture, color, shape, other abnormalities) of that portion.

One or more of these intended meanings are acceptable in a given context. Context dependency (or "situatedness", cf. [Menzies, 96]) is consequently troublesome when different contexts are to be integrated in the same ontology.

Conceptual relevance How to assess what is conceptually relevant? For example, if we are faced with the task of conceptualizing the term *inflammation*? Should we conceptualize all the possible meanings or only some, and what criterion should guide us?

Within AI, relevance problems in terminology conceptualizations have appeared mainly in the description logics domain, where the classic distinction between *terminologic* and *assertional* knowledge tried to impose a formal criterion on an ontological problem. What is terminological? How to state the border?[1] An interesting study on the difficulty of representing domain (medical) knowledge in such an environment is [Haimowitz, Patil, Szolovits, 88]). Among the others, assessments of the issue are in [Owsnicki-Klewe, 89]. For a general assessment of relevance in information modelling, see [Marjomaa, 93].

Since we claim, with [Davidson, 86], that «linguistic ability [...] is the ability to converge on a passing theory from time to time» (p.173), we can conclude that conceptual relevance itself is context-dependent (Davidson's passing theory is what we may call a conceptualization).

---

[1] Usually, only some rule-of-thumb is suggested, such as "encode in terminologic knowledge only what appears 'obviously' taxonomic".

Detail level A related problem to relevance is the *detail level*: when to stop detailing the explicitation of a conceptualization? If relevance is context-dependent, detail level implies a judgment on the refinability of a conceptualization without loosing the reference to the context (without "overcommitting"). For example, deciding about the conceptualization of the term *ulcer of stomach* as *inflammation of stomach* vs. *inflammation of the wall of stomach* vs. *inflammation of the mucosa of the wall of stomach,* is a detail level issue.

Our responses Although context-dependency is a problem to conceptual integration, it is also a guide as far as relevance and detail are concerned: in fact, we have to conceptualize only what is relevant in a context, and we do stop over there.

This leads to the problem of catching the specifities of different kinds of context and the way to conceptualize them. This is an ongoing research, too complex to present here; we can provide a minimal framework:[2] a *non-symbolic* context (e.g., spatio-temporal regions, situations, time spans, cultural systems) can be opposed to a *symbolic* context (e.g. linguistic texts, formal models), which is a representation of a non-symbolic context (cf. also [Sowa, 96]).

The symbolic context is the target of our method of terminological ontology construction. Namely, we take symbolic contexts to be the direct contexts of terminologies. Consequently, our solution to the relevance question in ontology integration is to integrate terminology repositories which have been developed by experts for given tasks with consequent contextual relevance and significant detail.

Such a solution is performed in our methodology in this way: firstly, after the direct symbolic context of a term is explicitated, we interpret the minimal intended meaning in that context. Thus, we have got a minimal local conceptualization. Secondly, we trigger one or more generic ontologies which can provide the coding of non-symbolic contexts for the local conceptualization: what are its situational, spatio-temporal, and cultural coordinates? Thirdly, we build an ontology library comprising the generic ontologies, as well as the domain ontologies which explicitate our conceptualization. In few words, we place a local conceptualization inside a more global knowledge context.

These responses have a philosophical background. Elsewhere [Steve et al., 96] [Gangemi et al., 97], we have explained in detail a semiologic theory of meaning, based on [Saussure, 80] [Eco, 84] [Peirce, 80], which underlies the way we treat context-dependency. The main features of this theory include the recognition of terms as *expressions* of arbitrary complexity within a *text.* Their interpretation is provided by other expressions, called *interpretants*, which are found in the same or a different text. The set of interpretants for an expression is called the Interpretant Field (IF). When we delimit (with a selection criterion) the set of texts which act as interpretant sources, the resulting set of interpretants is called the Qualified Interpretant Field (QIF). If we trigger some interpretation frameworks made of general knowledge, i.e. the so-called "paradigms", cf. [Kuhn, 62], and put the interpretants within such frameworks, the resulting structured field is called Ontologized Field (OF). OFs are used to construct our ontology libraries; in particular, paradigms are specified in the generic ontologies.

In Figure 1 we give an analogical intuition of this theory. The local analysis and definition of an expression can be represented as an onion section in which the dissected leaves represent the *definientes* i.e. the interpretants actually used in the definition. Different subjects can carry on different analyses and different definitions in the same way one can make different sections of an onion. However, we can figure out that different definitions are equivalent if one can find out correspondences between their definientes. In the onion metaphor this equivalence is analogous to the correspondence between pairs of dissected onion leaves belonging to the same leaf. To catch the core of the concept means to find out all the highly specific interpretants (leaves) necessary to distinguish the concept from all similar concepts: two conceptualizations can be equivalent only if they substantially catch the same leaves.

Another issue should be addressed here: if "paradigms" govern the interpretation framework for QIFs, how to assess the relevance of a paradigm? On this point, we adopt a special "grounding" philosophy which is variously inspired by a bunch of psychologic, linguistic, and philosophic works ([Harnad, 90] [Petitot, Smith, 91] [Talmy, 95] [Varzi, 96], etc.). We do not detail such a philosophy here; we only say that paradigms are constrained on their turn by:

---

[2] The following distinction corresponds to the one adopted in linguistics between "context" and "co-text".

• the actual *structure of the world* that interacts with human behavior: the ordinary, or *common sense structure*. This concerns the *universals*, or *invariants*, of cognitive perception, such as *wholeness* and *parthood* of objects [cf. Simons, 87], *connectedness*, *strata* of reality (material, biologic, psychological, socio-cultural) [cf. Hartmann, 66]; and

• the *cognitive schematization* [Lakoff, 90] [Langacker, 91], which bounds perception and develops to make humans efficiently interact with world; examples are the kinaesthetic image schemata, such as *up/down, front/back, containment, configuration, path, link, force dynamics*.

Therefore, a set of generic ontologies should be provided which contains theories about these two constraints. Such generic ontologies should not be "absolutely" right: they are good as far as ontology integration is successfully performed and intersubjectively acceptable.

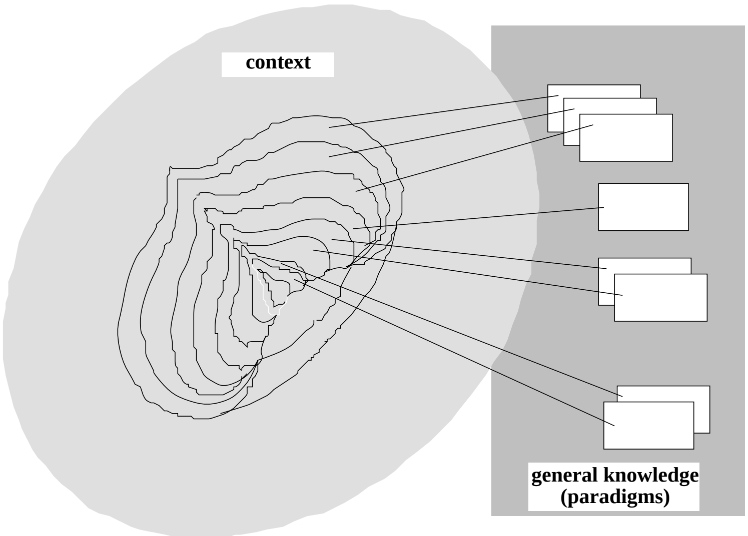

Fig.1: The onion metaphor: leaves represent interpretants in a local definition of an expression, linked to paradigms (rectangle).

2. THE ONIONS METHODOLOGY

2.1 An overview We developed ONIONS (ONtologic Integration On Naïve Sources) in order to analyse and to integrate domain ontologies. We consider ONIONS a methodology that constrains and explicits "intellectual activities", such as conceptual analysis and modeling choices, which cannot be reduced to sequences of "elementary actions" to be accomplished. The goal of our methodology is to transform the potential arbitrariness of such intellectual activity in a trasparent organization according to guidelines coherent with our theory of meaning (§1.); this means that the great amount of subjectivity inherent in this activity is not eliminated nor hidden, but it is made completely explicit step by step, such that it becomes intersubjective if one agrees with it. Partial agreements are possible depending on the modularity of the resulting system. Here we summarize the five main phases of ONIONS:

• M_0: Creating a corpus of validated textual sources of a domain. Sources must be individuated together with an assessment of their diffusion and validation inside the domain community.

• M_1: Taxonomic analysis. If lacking, taxonomies are constructed.

• M_2: Local source analysis. The conceptual analysis of terms in order to locate their free-text descriptions and other constraints (local definitions).

• M_3: Multi-local source analysis. The conceptual analysis of the descriptions allows to link the local definitions with multi-local concepts and general knowledge (paradigms).

• M_4: Building an integrated ontology library. An ontology library covers all the local definitions and the paradigms that have been used in building multi-local, integrated definitions.
• M_5: Implementing and classifying the library. These steps pertain to the diffusion, use, classification, and validation of the model.

Extension, partial acceptance/rejection, refinements, updating, integration with other sources can be carried out iteratively using ONIONS. The ontologies resulting from previous integrations can furtherly play the role of source to be analysed and integrated with its appropriate degree of relevance.

In Figure 2 we introduce in an abstract and schematic form the basics of ONIONS methodology. It describes a methodology with six phases and a set of input and output states in the analysis or in the construction of a terminology system. Such states are described by a set of structural and ontologic properties. We name a property "ontological" if it concerns the principles of *conceptual* organization of a terminology system.

Each ONIONS phase M_i makes a terminology system or repository evolve from a state S_i into a state S_{i+1}. P_i and O_i are respectively the structural and ontological properties of S_i systems. Hence, such properties also allow a classification of existing terminology systems according to their structural and ontological properties.

2.2 Phase M_0: Creating a corpus of validated sources At state S_0 domain knowledge is formally unstructured (P_0) as well as ontologically opaque (O_0: no explicit conceptualization). Phase M_0 aims at collecting a validated terminological corpus for a domain. Such a phase has hooks to corpora formation techniques and textual types definition and acquisition (not examined in this paper).

For example, our primary experiment of medical terminological integration has taken into account five terminology systems: the UMLS Semantic Network (UMLS-SN) [Humphries, 92] (all 215 semantic types and relations, and the ~1400 "templates" defined on them), SNOMED-III [Coté, Rothwell, Brochu, 94] (~600 most general concepts and links) and GMN [Gabrieli, 89] (~700 most general concepts) nomenclatures, ICD10 [WHO, 94] classification (~200 most general concepts), and the CORE model developed by the GALEN project [GALEN, 94] (version 5g, all ~2000 concepts).[3]

Other specialized corpora of medical terms have been conceptualized by the ONIONS methodology (e.g. surgical procedures [Rossi Mori, Gangemi, Steve, et al., 97]).

At state S_1 domain knowledge is represented by a list of valid expressions (P_1), which are meant to be conceptually plausible (O_1). Term lists compiled by experts and standard bodies, or extracted from free text, can be classified as such.

The next phases of ONIONS are designed to build and ontologize qualified interpretant fields (§1.) for such expressions.

A practical example of how qualified interpretant fields are retrieved and ontologized is followed in this overview. It concerns the interpretant field of *viral hepatitis type A*. Several terminology systems in our medical corpus contain interpretants to build a qualified interpretant field concerning *viral hepatitis type A*. We employ here only the UMLS system, which in its 'Metathesaurus' part (a collection of terminology systems) includes "viral hepatitis type A" from SNOMED-II, indicated as synonymous to "hepatitis A" from MeSH.

UMLS system includes the UMLS-SN, which has a hierarchical structure, includes "types" and "relations", and provides free-text definitions and template-like combinations of types and relations. It has a browser but does not allow to create new concepts. Its types are used as "categories" assigned to the UMLS Metathesaurus concepts, which comprise the MeSH thesaurus [NLM, 96] and other nomenclatures (SNOMED-II, ICD-9-CM, etc.).

— Behavior — ... | {MeSH Categories}
— Activity —
— Occupation — ... | {MeSH Categories}
— Event —
— Human-caused — ... | {MeSH Categories}
— Phenomenon — Natural — Biologic Function — Pathologic Fun. — Disease | {MeSH C.}

[3] For a description of the sources used in our experiment or quoted here: [Gangemi, Steve 96; Steve, Gangemi, 96].

or Process — Injury or Poisoning — ... | {MeSH Categories} or Syndrome
T —
— Entity — ...

Frame 1: The UMLS (semantic network) taxonomic context for "Disease or Syndrome".

2.3 Phase M_1: Taxonomic analysis Phase M_1 aims at finding out the main taxonomic structure within term lists. At the beginning, we select the most relevant sets of expressions (*source expressions*).

Given a corpus, the taxonomy of expressions contained in each single source is inferred. The taxonomy is exploited to identify the top-level concepts in the source, and then top-level concepts are used to choose a depth limit (for example, in the medical ontology application, we chose to truncate the *body part* taxonomy — within the *vessel* branching — to kinds of vessel, not including instances of arteries, veins, etc.). Since our main scope was to integrate general medical terminologic knowledge, the most detailed taxa for anatomy were excluded. This seems to be sound to the extent that a specialized microdomain integration could be done in a further research (for example, an extension of the current medical ontology ON9 to *angiology*).

As far as our example is concerned, UMLS-SN makes some distinctions (*Frame 1*). In the Metathesaurus, *viral hepatitis type A* is classified as an instance of UMLS-SN "Disease or Syndrome". In fact, the "Phenomenon or Process" type hierarchy branching in UMLS-SN stops at "Disease or Syndrome". To find the actual taxonomic position of *viral hepatitis type A* we have to search down the UMLS Metathesaurus and look at the MeSH hierarchies and at the other taxonomies referred there (ICD9-CM, SNOMED-II, etc.). The result of such investigation is summarized in *Frame 2*.

UMLS Semantic Network
Event [*see description in* frame 3]
•Phenomenon or process [*see description*]
••Natural phenomenon or process [*see description*]
•••Biologic function [*see description*]
••••Pathologic function [*see description*]
•••••Disease or syndrome [*see description*]
(subsumes:
1) MeSH "Virus diseases" hierarchies *entirely*, without linking to taxonomy
2) ICD9-CM item: "Viral hepatitis A without mention of hepatic coma"
3) SNOMED-II hierarchies *entirely*, without linking to taxonomy)

Entity [*see description*]
•Conceptual entity [*see description*]
••Finding [*see description*]
(subsumes: 1) MeSH "Symptoms and general pathology" hierarchy
2) ICD9-CM item: "Viral hepatitis A without mention of hepatic coma")

MeSH
Diseases
•Virus diseases [*see description*]
••Hepatitis, Viral, Human [*see description*]
•••Hepatitis A [*see description*]
•••RNA virus infections
••••Picornaviridae infections [*see description*]
•••••Enterovirus infections
•••••• ^Hepatitis A
••Digestive system diseases
•••Liver diseases
••••Hepatitis [*see description*]
•••••^Hepatitis A
•Symptoms and general pathology <NON-MeSH>
••Disease [*see description*] <NON-MeSH>

ICD9-CM
Diseases and injuries
•Diseases of the digestive system
••Other diseases of the digestive system
•••Other disorders of liver
••••Hepatitis
•Infectious and parasitic diseases
••Other diseases due to viruses and chlamidiae
•••Viral hepatitis
••••Viral hepatitis A without mention of hepatic coma

SNOMED-II
Disease Axis
•Infectious and communicable diseases
••Diseases caused by viruses
•••Viral hepatitis
••••Viral hepatitis, type A <D-0521>
•Diseases and syndromes of the digestive and urinary tract
••Diseases and syndromes of digestive system
•••Diseases and syndromes of liver and bile ducts
••••Disease of liver

Frame 2: The UMLS (Metathesaurus) taxonomies involving the conceptualization of "viral hepatitis type A".

Therefore, at state S_2 the lists have an order induced by IS_A inclusions, namely, the lists are mono- or multi-hierarchical taxonomies (P_2), as shown in *Frame 2*. From an ontological perspective we could ask how many and which meta-classes further organize the taxonomies: in other words we assess which kind of taxonomy we have got (O_2). For example, the lists in *Frame 2* are all homogeneous taxonomies. Most 'classic' terminology systems (ICD10, SNOMED-II, GMN, etc.) can be classified as such.

2.4 Phase M_2: Local definitions analysis Once a relevant set of concepts from each source is available, we focus on the criteria of classification, namely on the *local definitions* of concepts, in order to create a qualified interpretant field. We have to work out an answer to the "definitional" question: which is the difference within a group of homogeneous concepts from the same source, typically between two children-concepts of the same parent concept?

From a definitional viewpoint, concepts to be defined are "definienda", and defining concepts are "definientes". The problem is that very often sources have informal, or poor definitions, and sometimes they lack at all.

When definitions are lacking, we create a sound explicit definition, exploiting all hints that a terminology system can provide (hierarchy, grouping, free text definitions, boolean combinations, axioms, frames, meta-linguistic modifiers, etc.), as well as additional definitional sources (dictionaries, glossaries, encyclopaedias) and experts.

Additional texts are provided to explicitly construct the qualified interpretant fields of the sources. For instance, ON9 (§3.) exploited dictionaries [Dorland's, 94] [Stedman, 95] and expert physicians from partner institutions. We have also proposed a scale of explicitness for sources [Steve, Gangemi, 96] which is based on the availability of these hints.

Back to our guiding example, UMLS provides a fine environment for comparing various taxonomies from the same or different sources. Only, this environment does not provide explicit conceptualizations, except for a natural language description provided to all UMLS-SN types and to some MeSH items in the UMLS Metathesaurus. We should collect such natural language descriptions to get two results: 1) the explicitation of subsumption criteria and 2) the possible additional constraints that the terminology system has decided to include.

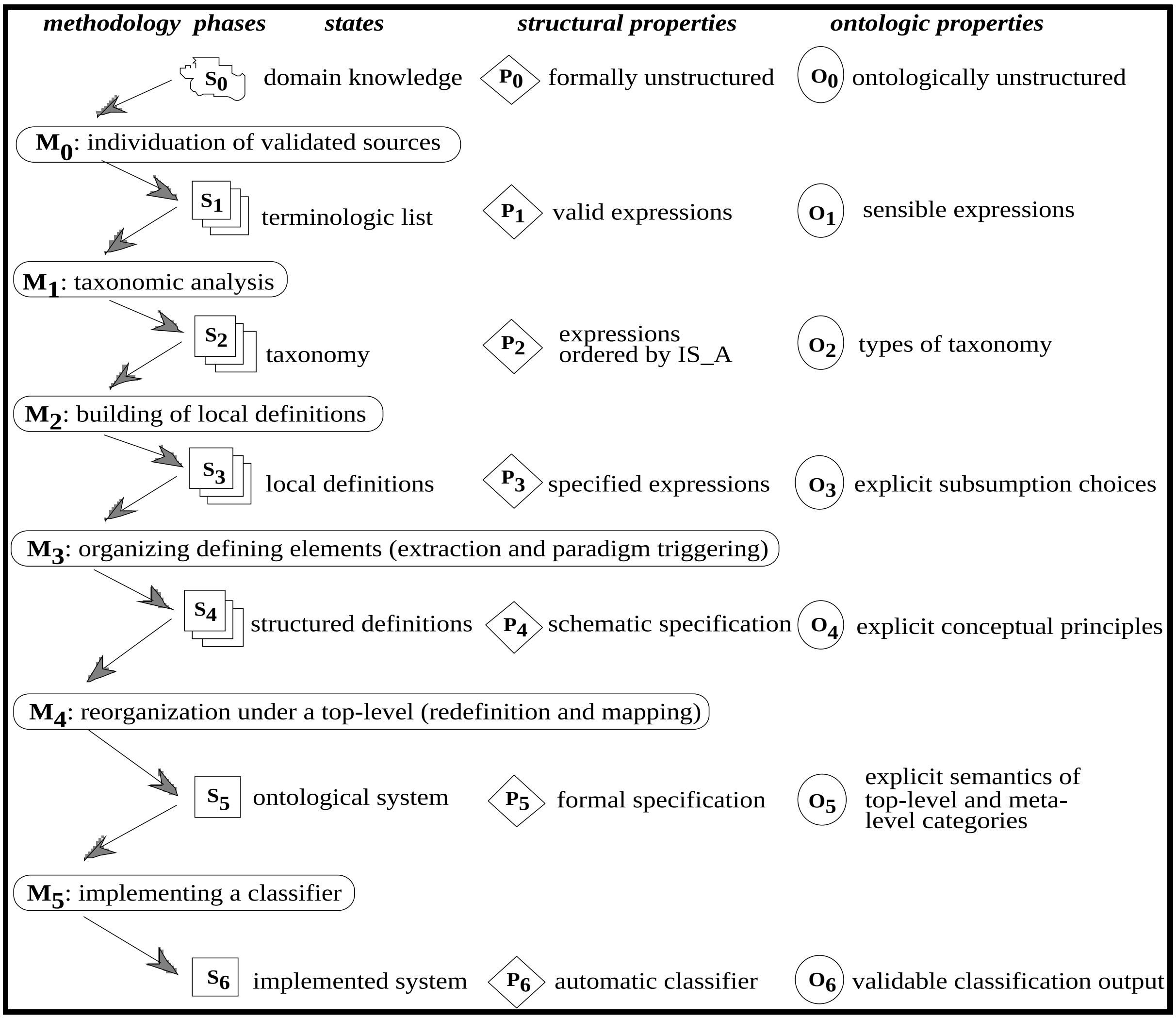

Figure 2. The phases of ONIONS. The output of any phase is a special state of a terminology system, described by both structural and ontological properties. Such states are independently re-usable for a specific purpose.

In *Frame 3* all the UMLS descriptions available for the concepts listed in *Frame 2* are presented. An analysis of such a collection (*Frames 2* and *3*) leads to the following issues:

- top-level concept descriptions are extensionally given, namely they do not provide subsumption criteria, e.g. "Event" is "*a broad type for grouping activities, processes and states*"
- many descriptions show lexical gaps. For example, "Pathologic Function" is described as "*a disordered process, activity or state*", while its child-concept "Disease or Syndrome" is "*a condition*", but "*condition*" is not a UMLS-SN concept: we are obliged to infer that a condition is "*a disordered process, activity or state*" as well. Another example is "*a broad type for grouping activities, processes and state*": neither "*activity*", "*process*", nor "*state*" are concepts in UMLS taxonomies. A more comprehensive example is the following: since "Hepatitis A" is an "Enterovirus Infection" (see taxonomy in *Frame 2*), which is subsumed by "Picornaviridae Infection", which is subsumed by "RNA Virus Infections", we can infer — with the help of additional expertise sources (experts, dictionaries, etc.) — that "Hepatovirus" in the description of "Hepatitis A" is actually the "Hepatitis Virus A", belonging to the genus "Enterovirus", family "Picornaviridae", order "RNA Virus", etc.

  On the other hand, when filling gaps, we should not be tempted to overcommit: for example one might want to include a dictionary definition of "Picornaviridae" [Stedman's, 95]: "a family of very small ether resistant, nonenveloped viruses having a core of single-stranded RNA enclosed in a capsid of icosahedral symmetry with 32 capsomeres". In fact there is no hint that UMLS terminology system has a similar commitment, at least as far as "Hepatitis A" is concerned. If we want to activate the interpretant field of "Hepatitis Virus A", thus we should have to check what the real commitment is. By the way, we would discover that the UMLS commitment for "Picornaviridae" is limited to: "small RNA viruses comprising some important pathogenes of humans and animals". In other words, commitment depends on some variety of non-symbolic context. Non-symbolic contexts are the only guide in decisions concerning a detail level problem (see §1.).

original UMLS-SN type descriptions:
Event $\approx_{df}$ *a broad type for grouping activities, processes and states*
Phenomenon or Process $\approx_{df}$ *a process or state which occurs naturally or as a result of an activity*
Natural Ph. or Pr. $\approx_{df}$ *a phenomenon or process that occurs irrespective of the activities of human beings*
Biologic Function $\approx_{df}$ *a state, activity or process of the body or one of its systems or parts*
Pathologic Function $\approx_{df}$ *a disordered process, activity, or state of the organism as a whole, of a body system or systems, or of multiple organs or tissues. Included here are normal responses to a negative stimulus as well as pathololologic conditions or states that are less specific than a disease. Pathologic functions frequently have systemic effects*
Disease or Syndrome $\approx_{df}$ *a condition which alters or interferes with a normal process, state, or activity of an organism. It is usually characterized by the abnormal functioning of one or more of the host's systems, parts, or organs. Included here is a complex of symptoms descriptive of a disorder*
Entity $\approx_{df}$ *a physical or conceptual entity*
Conceptual Entity $\approx_{df}$ *a broad type for grouping abstract entities or concepts*
Finding $\approx_{df}$ *that which is discovered by direct observation or measurement of an organism attribute or condition, including the clinical history of the patient*
original MeSH categories descriptions:
Virus Diseases $\approx_{df}$ *diseases produced by viruses*
Hepatitis $\approx_{df}$ *infection of the liver and liver disorder involving degenerative or necrotic alterations of hepatocytes*
Hepatitis, Viral, Human $\approx_{df}$ *viral hepatitis in man*
Picornaviridae Infections $\approx_{df}$ *virus diseases caused by the picornaviridae*
Hepatitis A $\approx_{df}$ *hepatitis caused by hepatovirus. It can be transmitted through fecal contamination of food or water*

Frame 3: The available UMLS descriptions for the taxonomies in Frame 2.

- some descriptions include meta-linguistic constraints, eg " *Included here is a complex of symptoms descriptive of a disorder*", which signals that also syndromes are instances of "Disease or Syndrome";
- some descriptions have default or modal expressions, eg " *It is usually characterized by ...*";
- except for UMLS-SN and partly MeSH, taxonomies do not provide a description for each concept in the taxonomy;
- several viewpoints emerge from the taxonomies (see *frame 2*): "Hepatitis A" results to be subsumed in both the "Disease or Syndrome" and "Conceptual Entity" branchings, and in the "Disease or Syndrome" branch it results to be subsumed in both the "Hepatitis, Viral, Human", "RNA Virus Infections", and "Digestive System Diseases" branchings. This is a good feature indeed, since non-multihierarchical taxonomies of some terminology systems (SNOMED, ICD) preclude explicit different views of the same concept. On the other hand, the relationships between the subsumption criteria to different views are not given.

A merging process, which takes into account the taxonomical information and the descriptions we have shown, and applies gap-filling and subsumption criteria to undescribed concepts, leads to a more compact description to *viral hepatitis type A* (*Frame 4*).

Typically, the merging process ignores extensional descriptions and meta-linguistic constraints, and produces some explicit viewpoints.

Once we consider only the characteristic description of *viral hepatitis type A* (the features not inherited from being a human liver disease, a morphology, and a finding), we can assess the "differentia specifica" of *viral hepatitis type A* within UMLS (*Frame 5*).

as human liver disease $\approx_{df}$ it alters or interferes with a normal phenomenon or process of a human organism, usually characterized by an abnormal functioning of the organism as a whole, of a body system or systems (the digestive system), or of multiple organs or tissues (the liver), and which occurs naturally
as aetiology $\approx_{df}$ caused by hepatovirus A, belonging to the genus Hepatovirus, family Picornaviridae, order RNA Virus
as morphology $\approx_{df}$ an infection of the liver involving degenerative or necrotic alterations of hepatocytes

as transmission $\approx_{df}$ it can be transmitted through fecal contamination of food or water *as finding* $\approx_{df}$ an abstract entity which is discovered by direct observation or measurement of an organism feature, including the clinical history of the patient

Frame 4: A concise collection of the UMLS descriptions for the viewpoints to the concept "viral hepatitis type A".

viral hepatitis type A $\approx_{df}$ "*the disease of human liver involving a degenerative or necrotic alteration of hepatocytes; it is caused by hepatitis virus A and is transmitted by fecal contaminated food. It can be viewed as a morphology of liver as well as a finding*"

Frame 5: A concise paraphrase of the conceptual viewpoints in Frame 4.

Frame 5 provides an intuition of what is the minimal situated commitment of the UMLS sources for the QIF of *viral hepatitis type A*.

As the examples show, at state S_3 the taxonomies are coupled with free text descriptions of terms (P_3); conceptually, these definitions explicitate (O_3) taxonomic constraints (the so-called *differentia specifica*). Most dictionaries and glossaries in medicine are classifiable as S_3.

2.5 Phase M_3: Multi-local definition analysis: triggering paradigms The methodological phase M_3 consists in 1) the schematization of the elements produced during the M_2 description, in order to provide some "weak" constraints, and in 2) the "triggering" of the conceptual principles (paradigms) which motivate the description.

This schematization can be performed through the construction of domain ontologies for the QIFs produced in phase M_2. Such domain ontologies do not require an axiomatization, but only a set of "templates". This amounts to say that elements ("fillers") used in the description of an expression have explicit relationships ("slots" or "roles") with the described term (the "frame"). In logical terms, a frame gives constraints on the domain and the range of the relations applicable in a certain context of knowledge. Ontologically, a frame should have explicit conceptual principles. For instance, a conceptual principle may concern *part-whole* relationships, another the *teleology* of processes, still others the *quantities*, the *topology* of objects, the *physical* properties, and so on. In other words, we should make a call for some valuable and practicable generic ontology.

The *viral hepatitis type A* description from M_2 (from *frame 5*) may be schematized by encoding it in some frame language with loose formal constraints. Here we create a local definition written in the Ontolingua "frame-ontology" [Gruber, 93]:

```
(in-theory 'o-umls)                                                          (1)

(define-class viral-hepatitis-a (?vh)                                        (2)
 "the disease of human liver causing a degenerative or necrotic alteration of
  hepatocytes; it is caused by hepatitis virus A and is transmitted by fecal
  contaminated food or water. It can be viewed as a morphology of liver as
  well as a finding"
 :axiom-def (and (subclass-of viral-hepatitis-a human-liver-disease)
                 (subclass-of viral-hepatitis-a finding)
                 (subclass-of viral-hepatitis-a morphology)
                 (caused-by viral-hepatitis-a hepatitis-virus-a)
                 (involves viral-hepatitis-a degenerative-alteration-of-hepatocytes)
                 (involves viral-hepatitis-a necrotic-alteration-of-hepatocytes)
                 (transmitted-by viral-hepatitis-a fecal-contaminated-food)
                 (transmitted-by viral-hepatitis-a fecal-contaminated-water)))
```

Constructs (1) and (2) say that in a theory named `o-umls`, the concept `viral-hepatitis-a` is subsumed by (inherits the constraints from) the concepts: `human-liver-disease`, `finding`, and `morphology`, and features five specific constraints (the syntax is: *template* ::= *slot frame filler*).

This is the schematization part: a minimal structure has been overimposed to the free text description by conjoining all the minimal constraints and unpacking the "or-expressions".

The complete account for the UMLS *viral hepatitis type A* QIF calls for constructs for all the concepts involved: `human-liver-disease`, `finding`, `morphology`, `hepatitis-virus-a`, `degenerative-alteration-of-hepatocytes`, etc. These, on their turn, would require other constructs for their respective QIFs, and so on.

For the ontological part, we have to trigger the paradigms which underly the specific constraints. Many routes can be taken at this point. Our reasoning steps included:

(a) since `_caused-by_`, `_involves_`, and `_transmitted-by_` all concern relationships between objects and processes, and since some valuable cognitive paradigms ("actantial" paradigms) claim for a uniform treatment of all relations involving some object carrying some process and the related dynamics, we make a call for some theory of actants;

(b) moreover, what allows to keep all these constraints together? how does a transmission of contaminated food concern alteration of hepatocytes? these leads to the recognition of the part-whole structure of the objects and processes involved; therefore, we make a call for some part-whole theory;

(c) the explicitation of part-whole structure requires an explicit reference to an organism in which such structure makes sense.

These reasoning steps are encoded as explicit calls for generic theories in the `:issues` field of an Ontolingua construct, which is added to the definition (2):

```
:issues ((:generic-theories  "_caused-by_, _involves_, and _transmitted-by_        (3)
         require a theory of actants and a paradigm of functions (natural
         processes)" "the patient status is not mentioned" "anatomy is not
         mentioned: at least, a part-whole theory is required")))
```

As an output of M_3, at state S_4 definitions are framed (P_4) and have explicit hooks to paradigms (O_4).

Few terminology systems have S_4 properties: the UMLS Semantic Network is one, the GRAIL models in the GALEN project are another example of S_4 (the GRAIL implementation also features an automatic classifier (P_6), see below phase M_5). At present, no terminology systems in medicine fulfil O_4. How to fulfil O_4 as well? For ONIONS-produced ontologies, we have developed a library of generic ontologies (see next phase and §3.).

2.6 Phase M_4: Building an integrated ontology library We have seen that for each source, local definitions have underlying paradigms (see §1.). Our purpose is the enrichment of local definitions by constructing domain ontologies which explicitly include other ontologies specifying such paradigms, for example [Sowa, 95] [Bateman, 90] [Hartmann, 66] [Simons, 87] and many others. Such an enrichment is not an arbitrary choice: it is made in order to connect heterogeneus local definitions by explicitating their paradigms (usually through the application of generic ontologies). Therefore it requires a minimal increase of commitment: only to the extent that we get the raising of definitions from the status of *local* to *multi-local*. Namely to the extent that we have constructed a library of domain and generic ontologies which contains (or allows the construction of) all the heterogenous definitions.

Since generic ontologies are not built with having in mind all their possible specifications in domain ontologies, phase M_4 comprises a lexical gap-filling process. Gap-filling is performed at M_2 to make local taxonomies complete, while in M_4 it is performed to raise the status of taxonomies to multi-local.

A "lexical gap" is better understood when two different languages are compared: for example, where English has *wood*, Italian has *legno* (as *matter*), *bosco* (as *aggregate of trees*), *foresta* (as *wide, heterogeneous aggregate of trees*). Not that English lacks the Italian ontology: it only does not let it emerge in the lexicon of words, in fact English is capable to paraphrase the three Italian meanings through multi-word noun phrases. But to paraphrase the Italian meanings, English needs the explicitation of the paradigms which guide the interpretation of the three meanings (a paradigm for *matter*, another for *parts and aggregates*, another for *heterogeneity*, another for *qualitative sizes*, etc.).

In terminology systems (in the same language), the phenomenon, though less evident, is the lexical correlate of conceptual heterogeneity: for any two formally equivalent expressions (for example, "viral hepatitis type A" in SNOMED and "hepatitis A" in MeSH), the interpretant fields are the same (they are 'synonymous'), but the qualified interpretant fields are different, because the two terminology systems explicitly encode only certain taxonomical constraints. In short, we could define an M_4 *lexical gap* as follows: «when we compare two allied (referrable to the same interpretant field) QIFs gathered from two different lexica, the absence in one QIF of a concept present in the other is a *lexical gap*».

Gap-filling is a finite, decidable work, thus the detail level problem (§1.) is not resurructed in the form: "how much comprehensive a generic ontology should be?". In the *wood* example, a generic ontology concerning parts and aggregates should contain all and only the definitions useful to integrate Italian and English, not a maximally comprehensive ontology.

On the other hand, it is not so easy finding pretty generic ontologies which are ready to be chunked each time for our special purpose of the moment. Above all, could we call such chunks a response to context-dependency? Well, there are two answers to this difficulty: the first is operative: chunking must be made attentively and explicitly; namely, if we want to buy a section of Nelson Goodman's mereology, we must refer explicitly this paradigm and the possible modifications we make. Also, we should include in our generic ontology inspired by that paradigm at least the most general concepts, in order to maintain a minimal soundness to the paradigm formalized in the generic ontology. The second answer is not a short term one: we envisage the creation of 'common good' data bases containing a huge amount of paradigms expressed as (formal) generic ontologies. In this ideal situation, the ontological engineer can buy a piece of a data base and provide the reference to that piece when creates its own ontology.

We summarize the methodological phase M_4 as follows:

(1) the construction (or the reuse, if available) of a library of generic ontologies to account for the `(:issues (:generic-theories))` requirements memorized during M_3 phase; this equals to build a well-grounded *top-level*;

(2) the enrichment of domain ontologies from M_3 phase. This requires the inclusion of the generic ontologies in the domain ontologies;

(3) the assignment of sound meta-level categories to the classes and relations in the library.

As far as the guiding example is concerned, *frame 3* gives us an informal description of the minimal UMLS commitment for *viral hepatitis type A*.

In phase M_3, with construct (3) we have also kept explicit track of the paradigms required to integrate the viewpoints present in *frame 3.* On the other hand, we still lack:

- a formal explicit conceptualization (an ontology) for each viewpoint in the *viral hepatitis type A* (UMLS) QIF;
- an explicit specification of the generic ontologies to be used to specify the QIF (to make it an Ontologized Field, cf. §1.) for *viral hepatitis type A* in the context of an ontology library.

Making an Ontologized Field of the above descriptions for *viral hepatitis type A* would be quite difficult if we had to start from scratch. Luckily, we can define it on the basis of the integrated ontologies already developed by applying ONIONS.

For example, in ON9 (see §3.), we can trace back `viral-hepatitis-type-a` to `viral-hepatitis`, `human-liver-disease`, `disease`, `pathologic-function`, `physiologic-function`, `biologic-function`, `function`, `process` (by the way, this backtracing is an extensive M_4 gap-filling). All these concepts are contained in explicit ontologies, thus `viral-hepatitis-type-a` is automatically inserted in a wide network of ontologized fields (the ON9 ontology library). Consequently, in the multi-local definition (4) only the specific constraints (the "differentia specifica") are listed:

```
(define-class viral-hepatitis-a (?vh)                                    (4)
 "the disease of human liver causing a degenerative or necrotic alteration of
  hepatocytes; it is caused by hepatitis virus A and is transmitted by fecal
  contaminated food or water. It can be viewed as a morphology of liver as
  well as a finding"
 :axiom-def (and (subclass-of viral-hepatitis-a viral-hepatitis)
                 (subclass-of viral-hepatitis-a finding)
                 (!type viral-hepatitis-a ))
 :def (exists (?liv ?pat ?vir ?vec ?pro ?hep ?inf)
             (and (liver ?liv) (*patient ?pat)
                  (part ?liv ?pat) (is-embodied-in ?vh ?liv)
                  (virus-a ?vir) (has-a-cause ?vh ?vir)
                  (or (fecal-contaminated-food ?vec) (fecal-contaminated-water ?vec))
                  (contains ?vec ?vir) (trasmitted-through ?vh ?vec)
                  (or (degeneration-process ?pro) (necrotic-process ?pro))
                  (induces ?vh ?pro)
                  (hepatocyte ?hep) (_plurality ?hep) (element ?hep ?liv)
                  (is-embodied-in ?pro ?hep)
                  (inflammation ?inf) (is-morphology-of ?inf ?liv)
                  (is-a-cause-of ?pro ?inf))))
```

(4) differs deeply from (2). Firstly, the frame definition in (2) is become an axiomatization:
• subclass declarations have been maintained, but with updated subsumptions;
• a !type meta-level assignment has been added (see §4.);
• the existentially quantified first-order sentence gives a clear focus to the vague templates in (2);
• in particular, the compact expressions of (2) have been expanded into well-connected assertions; e.g., `degenerative-alteration-of-hepatocytes` is expanded according to the minimal commitment of ON9: it is now a `degeneration-process` which `is-embodied-in` a `plurality` of `hepatocytes` which are explicitly `element` of any `liver` which `embodies` a `viral-hepatitis-a`. The morphological meaning of `degenerative-alteration-of-hepatocytes` is encoded in the last three conjuncts.

These features are the result of the application of a more formal language as well as they originate from the inclusion of the generic ontologies called for in (3). For example:
• from theory: actants[4] we used the relations: `is-embodied-in`, `has-a-cause`, `induces`. The axiomatizations in theory: actants mainly rely on results obtained in cognitive science, linguistics and narratology investigations [Miller, Johnson-Laird, 76], [Prince, 82], [Fillmore, 71];
• from theory: epidemiology (dependent on theory: actants), we used the relation `trasmitted-through`, which is axiomatized with respect to the object which has the capability of conveying a pathogenic agent to the organism;
• from theory: meronymy[5] we used the relations: `part` and `element`, that, with `plurality` (from theory: quantities), allow to model `hepatocytes`, which are (a plurality of) elements of the liver;

As an output of M_4, at state S_5 definitions are axiomatized (P_5); ontologically, definitions have an *explicit semantics* (O_5) of both the *top-level concepts* (the concepts provided by generic ontologies) and of the *meta-level categories* (the concepts in a representation ontology, §4.).

In Fig. 2 only one system is in the state S_5 (only one square frame), in order to represent that it can integrate the previous ones. Of course, we can get different integrated systems according to the particular generic ontologies that one decides to include or use.

2.7 Phase M_5: Classifying the library The methodological phase M_5 consists in the implementation of a domain ontology in a system which allows automatic classification. Obviously, the generated classification should pass a validation control.

At state S_6, a domain ontology is implemented in an automatic classifier (P_6), e.g. Loom [Mac Gregor, 94]. Some terminology systems are currently implemented with an automatic classifier, for example the GALEN Core Model. Our method is to export a library of ontologies from the Ontolingua [Gruber, 93] or OCML [Motta, 95] form into Loom and to make Loom classify the library (e.g., see the translation of (4) in the following (5):

```
(defconcept viral-hepatitis-a :is-primitive                                                   (5)
    (:and finding viral-hepatitis
     (:some is-embodied-in (:and liver (:some part *patient)))
     (:some has-a-cause virus-a)
     (:some transmitted-through
      (:and (:or fecal-contaminated-food fecal-contaminated-water)
       (:some contains virus-a)))
     (:some induces
      (:and (:or degeneration-process necrotic-process)
       (:some is-embodied-in (:and hepatocyte _plurality (:some element liver)))
       (:some is-a-cause-of (:and inflammation (:some is-morphology-of liver)))))))
```

WWW-available, collaborative modelling tools are currently exploited to support phases M_4 and M_5, and to manage the byproducts of the earlier phases [Gangemi, 97].

---

[4] see the WWW site; http://saussure.irmkant.rm.cnr.it/HOME/ON9/actants/index.html.
[5] see the WWW site: http://saussure.irmkant.rm.cnr.it/HOME/ON9/meronymy/index.html

3. THE ON9 ONTOLOGY LIBRARY

ON9 is a library of ontologies designed by means of the ONIONS methodology (some part of it is available at: http://saussure.irmkant.rm.cnr.it/HOME/ON9/index.html).

ON9 has been defined as the multi-local integration of a corpus of medical sources (§2.2). It includes a detailed representation ontology and it is written in Ontolingua. It has been translated to Loom and it will be soon available through the Ontosaurus HTTP interface [Swartout, Patil, Knight, Russ, 96]. Its top-level thus governs a large set of representation, general and domain ontologies. It can be summarized by the tuple:

$$< s , p , c , r >$$

$s \in S$ **:** the domain of *object sorts*: static, atemporal entities;

$p \in P$ **:** the domain of *process sorts*: kinematic, temporal entities;

$c \in C$ **:** the domain of possible *context sorts* arising as complements of the sum of other intensional entities: regions, domains, time spans, situations, texts;

$r \in R$ **:** the set of intensional *structuring concepts* (the dimensions which give structure to the concepts in an Interpretant Field): $\bigcup (Ar, As, Sc, Ph, La)$ where:

- Ar is the set of *Actantial roles*: it includes relations such as 'agent', 'patient', 'instrument', 'goal', 'cause', etc. This should map the notions of 'scene', 'verb frame', 'script', etc. from Fillmore, Schank, Lagendoen, etc. Such structuring concepts are useful in the definition of contexts and meta-level categories (§4.);
- As is the set of *Assessment*: *relations,* it includes concepts of 'assessment' or 'judging', such as typicality, conventionality, relevance, representation. Such concepts deserve a deeper understanding, for example in terms of more basic cognitive properties;
- Sc is the set of *Schematic relations* of space, time, quantity, etc. (governed by a cognitive schema: such schemata can be taken as 'forms of the intuition', in a kantian fashion, as well as 'a-priori forms of perception', as gestalt psychology proposes). This set is rich: for example spatial ontologies comprise mereological, topological, locative relations;
- Ph is the set of *relations expressing Physical (substantial) concepts*, such as matter features, morphologies, physical states and dimensions, etc.
- La is the set of *ontological Layers*: the realms in which the continuum of the human knowledge about reality can be segmented (material, biologic, psychologic, social, abstract). This set includes granularities as well: molar, molecular, atomic, sub-atomic, etc. (cf. §1.). This is the primary dimension for defining objects.

We could also consider a variable $m \in M$ **:** the set of Meta-level categories, namely the concepts which mediate between the formal and conceptual properties of an ontology. While previous concepts and relations are usually defined within generic ontologies, meta-level categories are defined in representation ontologies, sometimes called 'metaontologies' (see §Introd. and §4.).

Figure 3 shows an inclusion lattice of some ON9 ontologies: the representation ontologies provided by default in Ontolingua are "frame-ontology" and the set of "kif-ontologies". We defined the ontologies: "structuring-concepts", "meta-level-concepts" and "semantic-field-ontology", to link the representation ontologies with the generic ontology library. The sets of "structural ontologies" and of "structuring ontologies" contain generic ontologies. Generic ontologies are variously included in domain ontologies. In particular, the integrated-medical-ontology includes all the generic ontologies which have been used to integrate the terminological ontologies of the five terminology systems.

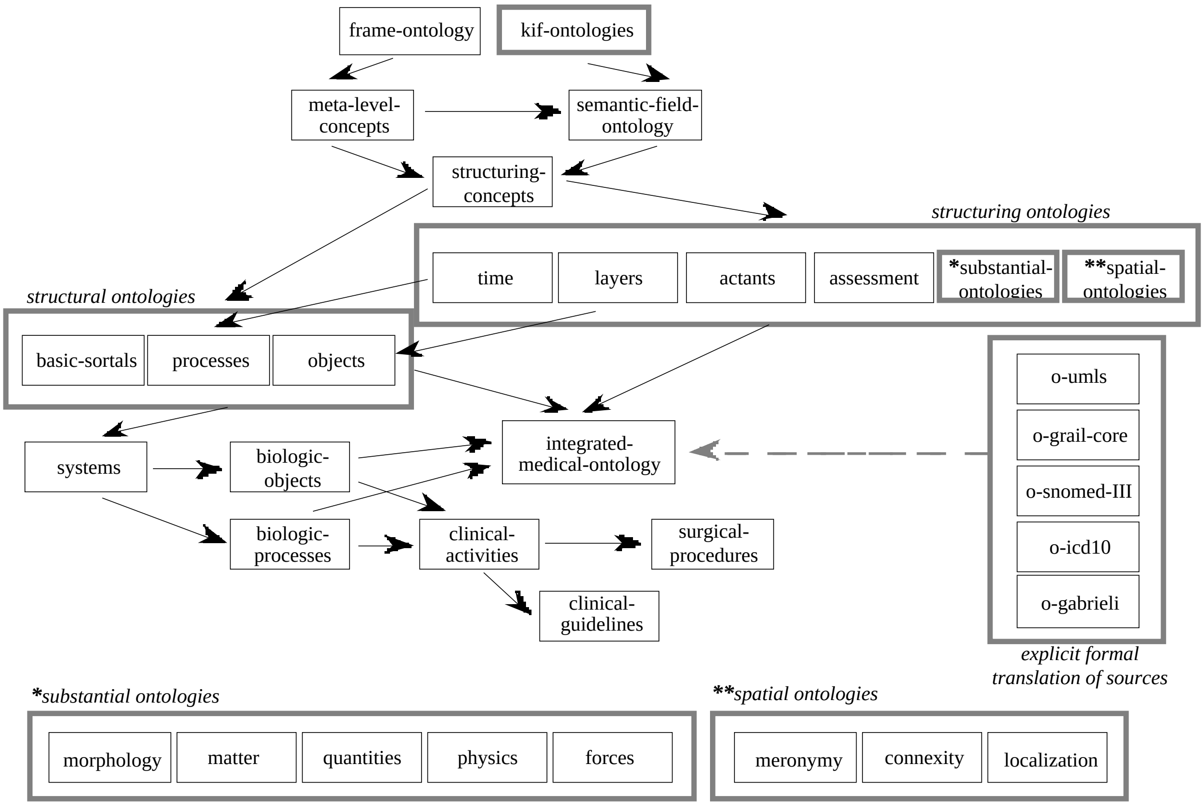

Figure 3. A significant subset of the inclusion lattice of the ON9 library of ontologies. Ontologies are represented by black circles. Thick grey frames or circles are sets of ontologies (some explictly show the elements). The semantics of black arrows is *included-in* (applied according to a system constraints, for example Ontolingua inclusion is looser than Loom's). The dashed grey arrow means *integrated-in* .

The analysis perfomed to build our ON9 ontology library allowed us to explicitly formalize important interpretant fields which have heterogeneous conceptualizations in the sources described in §2.2, for example:

- disease as *pathological function* or as *outcome of an inferential process*, and diagnosis as inferential process or as *outcome of an inferential process*. For example, SNOMED or ICD identify disease and diagnostic outcome, because they are oriented to statistics, records and reports. UMLS-SN is oriented towards scientific knowledge, and disease is considered a pathologic function, while diagnosis is an inferential process; SNOMED or ICD diseases have to be mapped here as diagnostic outcomes.
- morphology as *form* vs. *outcome of a function* vs. *structure in a given condition*, in particular when a structure is the outcome of a surgical procedure or a pathological process;
- regions, spaces, holes as *conceptual arrangements of structures* (e.g. in UMLS-SN) vs. *structures* themselves (*immaterial objects*).

4. METAONTOLOGICAL ISSUES

We have reused terminology systems, also assuming that they contain the *basic concepts* of a domain [Steve et al., 96] [Rosch et al., 78] [Lakoff, 90].

The basic concepts and gap-fillers (cf. §2.2.5) in the resulting ontology library are called the *structural concepts*, while the concepts which provide the relational structure are called the *structuring concepts*, mostly coming from the paradigmatic sources.

If an order-sorted logic is adopted, we suggest to represent structural concepts with *sorts* and structuring concepts with *relations* of various arity.

If a more refined meta-ontology is wanted, one should look for a theory specifying a representation ontology (cf. §Introduction). There is no general agreement on the necessity of a detailed representation ontology; we suggest that it should be part of an ontology library, but should not be implemented in a formal language. In other words, we want to be free to use a fomal language with a minimal meta-level commitment, but we also want the capability of assigning Meta-Level Categories (MLC) to concepts (independently from the formal language used). A MLC (such as slot, concept, sort, relation, etc.) is a predicate which assigns a precise semantics to a (object-level) predicate.

In the following we summarize some issues in the design of MLC.

Guarino and co-workers [Guarino, Carrara, Giaretta, 94] introduce some MLC for unary predicates on the basis of some criteria (*Tab. 1*):

• *type, quality*, *indep-role* and *dep-role* are *sortal* (countable) unary predicates: for example, *book* is countable because a part of an instance of book is not a book (compare to *object*); *roles* must be subordinated to some *type/kind*;
• *category* and *mass-like* are *pseudo-sortal* unary predicates: these are stable and rigid (an instance of such a predicate is always an instance of it, but it is uncountable;
• *property* and *state* are *non-sortal* unary predicates: these are uncountable and non-rigid (predications are not necessary);

MLC	**countable**	**temporally stable**	**ontologically rigid**	**independent from another predicate**	***Examples***
TYPE/KIND	Y	Y	Y	Y	apple, person
QUALITY	Y	Y	Y	N	color, length
INDEP-ROLE	Y	Y	N	Y	fat-person
DEP-ROLE	Y	Y	N	N	son, student
CATEGORY	N	Y	Y	Y	event, object
MASS-LIKE	N	Y	Y	Y	gold, sand
PROPERTY	N	Y	N	Y	red, great
STATE	N	N	N	Y	studies, closed

Table 1: A set of meta-level categories for unary predicates (adaptation from [Guarino, 92] [Guarino et al, 94]).

Criteria for MLC definition of unary (and binary) predicates have been formalized in a modal logic in [Guarino, 92] [Guarino, Boldrin, 93]; for a summary, [Steve et al., 96].

4.1 ON9 meta-ontological commitment Our *sorts* divide into *types*, *categories*, and *reified properties*, which roughly map to Guarino's *sortal* predicates. Our *reified properties* map to Guarino's *roles*, which is non-rigid, say it is an accidental specification of a type. From a converse path, we defined types as motivated by validated basicality in a domain, and our *reified properties* are thus non-basic, although from a local viewpoint.

The uncountable topmost layers of ON9 are the *categories*. A cognitive counterpart of countability is the basicality prototype for common language (for criteria, see [Steve, Gangemi, 96]). In particular, categories are less reidentifiable than types: what's the mental image for "object", or "process"? On the other hand, categories seem to be highly schematizable: imagine a simple diagram for "process" (eg, a path) or "object" (eg, a configuration).

Among reified properties, we can follow the dependent/independent distinction. Its cognitively-based counterpart can be *actantiality*, the property of being able to take a special part in a situation change: agent, patient, instrument, goal, cause, etc. The special status of these categories has been acknowledged in ancient grammars ("complements"), in recent frame grammar ("cases"), in narratology, etc.

Meta Level Categories			**basic**	**field structuring**	**one, accessible**		**actantial**	***Examples***
					image	**schema**		
SORT	TYPE		++	—	+	—	context-dep	liver, disease
	CATEGORY		+	—	—	+	context-dep	activity, object
	REIFIED PROPER -TY	SUBSTANTIV -ATE	—	—	—	—	—	integral-structure
		ROLE	—	—	—	—	+	aetiologic agent
PROPERTY			+	+	—	+	not applied	integral, red
RELATION			+	+	—	+	not applied	part, at right of

Tab. 2: The (cognitive) ontological commitment of meta-level categories in ON9 representation ontology.

Thus, a dependent role is called in this perspective a *role* tout court, which also corresponds to common sense, while an independent role is called a *substantivate*, to stress that it actually is the sortal (nominal) counterpart of a relational (adjectival, property-like) application.*Tab. 2* summarizes the current ON9 representation ontology (formal definitions are retrievable at http://saussure.irmkant.rm.cnr.it/HOME/ON9/metaontology/index.html).

CONCLUSIONS

As a result of our experience in developing medical terminological ontologies through ONIONS, we can conclude that:

(a) as far as validity is concerned, ONIONS methodology appears to be suitable in creating as well as in integrating, updating, extending and maintaining ontologies; however, we need more feedbacks from other users. We hope to get it from current cooperative modeling experiments.

At present, use and integration of ON9 are negotiated or customized on the WWW (through some dedicated tools, cf. [Gangemi et al., 97]).

(b) as far as conceptual integration of terminologies is concerned, ONIONS may support:

- conceptual explicitness and formal upgrading of terminology systems: term classification and definitions are available in a common, expressive formal language;
- mappings to other terminology systems: local term definitions are constructable, even though the source does not include them explicitly;
- conceptual upgrading of terminology systems: 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.

(c) as far as reuse and maintenance are concerned, ONIONS may support:

- a motivated generic ontology library, from the integration of generic and domain sources;
- 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: an integrated medical ontology.

(d) as far as implementation is concerned: representing and situating a terminological ontology requires complex formal specifications involving first-order sentences, some second-order sentences about situation and contextual change, pervasive existential quantification, definition of meta-level categories of the representation language, etc. Our experience suggests that Ontolingua, OCML, and Loom are fitting to these purposes.

(e) as far as representation ontology is concerned, our medical ontology has been built and modeled without any a-priori formal constraint on ontological commitment, through a methodology finalized to knowledge integration. A-posteriori we analyzed it through the ontological commitment costraints proposed by Guarino; the analysis showed that to a great extent those requirements are satisfied in our medical ontology. As the requirements were not widely fulfilled in the sources, this outcome suggests as plausible that one could find out a theory of equivalence between methodological assumptions on integration and a settlement of costraints on MLC commitment.

The necessity of extensive off-line human intervention in the search, choice, and formalization of generic ontologies seems to be an unavoidable bottleneck in ONIONS ontology construction. An appealing alternative would be to adopt a systemic approach in the generic library, which is widely shared and formally available. Our analysis evidentiates that system theory, widely used in engineering domains (the usual configuration of component-state-event-process), does not fit well enough to the medical domain. The basic principles motivating the conceptualization of terminology in medical domains seem to refer also to other fields, often non formalized, such as linguistics, philosophy, and cognitive science.

Acknowledgements

We thank the anonymous referees and Nicola Guarino for their useful suggestions. This work has been partly funded by the Italian National Research Council project SOLMC (Ontologic and Linguistic Tools for Conceptual Modeling).

REFERENCES

Bateman JA. The Theoretical Status of Ontologies in Natural Language Processing. Report cmg-lg/9704010

Borgo S, Guarino N, Masolo C. Stratified Ontologies: The Case of Physical Objects. in Vet (ed.) *Proceedings of ECAI96 Workshop: Ontological Engineering* (1996).

Borst P, Akkermans H, Top J. Engineering Ontologies. *Int. Journal of Human-Computer Studies*, 46 (1997).

Brachman R, McGuinness DL, Patel-Schneider PF, et al. Living with Classic. in JF Sowa (ed.): Principles of Semantic Networks, San Mateo, CA, M Kaufmann (1991).

CEN prENV 12264:1995. Medical Informatics - Categorial structure of systems of concepts - Model for representation of semantics. Brussels: CEN (1995).

Cohn AG, Randell DA, Cui Z. Taxonomies of Logically Defined Qualitative Spatial Relations. *International Journal of Human-Computer Studies*, 43 (1996).

Coté RA, Rothwell DJ, Brochu L (eds). SNOMED International, 3rd ed., 4 vols. Northfield, Ill: College of American Pathologists (1994).

Davidson D. A Nice Derangement of Epitaphs. in R Grandy, R Warner (edd): Philosophical Grounds of Rationality: Intentions, Categories, Ends. Oxford, Clarendon Press (1986).

Eco U. Semiotica e filosofia del linguaggio. Torino, Einaudi (1984).

EPISTOL Core Group. Knowledge Processing for Decision Support in the Health Sector. in Barahona P & Christensen JP (eds.): Knowledge and Decisions in Health Telematics, IOS Press (1994).

Evans DA, Cimino JJ, Huff SM, Bell DS for the CANON Group. Toward a Medical-Concept Representation Language. *Journal of the American Medical Informatics Association* 1:207-17 (1994).

Falasconi S, Stefanelli M. A Library of Medical Ontologies. in Workshop on Comparison of Implemented Ontologies, ECAI 94 (1994).

Falasconi S, Lanzola G, Stefanelli M. Using Ontologies in Multi-Agent Systems. in B Gaines, M Musen (eds), *Proceedings of KAW96* (1996).

Fankhauser P, Kracker M, Neuhold E. Semantic vs. Structural Resemblance of Classes. Special issue: Semantic Issues in Multidatabase Systems, SIGMOD RECORD, Vol. 20, No. 4, December, pp. 59-63 (1991).

Fillmore CJ. Types of Lexical Information. DD Steinberg, LA Jakobovits (eds): Semantics: an Interdisciplinary Reader in Philosophy, Linguistics and Psychology, Cambridge UP (1971).

Gabrieli E. A New Electronic Medical Nomenclature. *Journal of Medical Systems*; 3 (1989).

GALEN Project. Documentation available from the main contractor Rector AL, Medical Informatics Group, Dept. Computer Science, Univ. Manchester, Manchester M13 9 PL, UK, (1992-1994).

Gangemi A, Steve G, Giacomelli F. ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration. in Vet (ed.) *Proceedings of ECAI96 Workshop: Ontological Engineering* (1996).

Gangemi A, Steve G, Pisanelli DM, Giacomelli F. WWW Tools for Ontology Negotiation. CNR-ITBM-RIM Technical Report 0497B (1997).

Genesereth MR, Nilsson NJ. Logical Foundations of Artificial Intelligence. Kauffman, Los Altos (1987).

Gennari JH, Tu SW, Rothenfluh TE, Musen M. Mapping Domains to Methods in Support of Reuse. *International Journalm of Human-Computer Studies*, 41, 399-424 (1994).

Gerstl P, Pribbenow S. Midwinters, Endgames, and Body Parts: A Classification of Part-Whole Relations. *International Journal of Human-Computer Studies*, 43 (1996).

Gruber T. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*; 5 (1993).

Guarino, N. Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases. *Data & Knowledge Engineering*, 8: 249-261 (1992).

Guarino N, Boldrin L - Concepts and Relations - in Guarino N, Poli R: *Pre-Proceedings of the International Workshop on Formal Ontology*, Ladseb, Padova 1-17 (1993).

Guarino, N., Carrara, M., and Giaretta, P. An Ontology of Meta-Level Categories. In J. Doyle, E. Sandewall and P. Torasso (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94)*. Kaufmann, San Mateo (1994).

Guarino N. Formal Ontology, Conceptual Analysis and Knowledge Representation. In N Guarino & R Poli (eds.) Formal Ontology in Conceptual Analysis and Knowledge Representation, special issue of *International Journal of Human-Computer Studies*, 43 (1995).

Haimowitz IJ, Patil RS, Szolovits P. Representing Medical Knowledge in a Terminological Language is Difficult. in Greenes RA (ed.): Proc. of the Twelfth SCAMC. Los Angeles, IEEE Computer Society (1988).

Harnad S. The Symbol Grounding Problem. *Physica D*, 42 (1990).

Hartmann N. Zur Grundlegung der Ontologie. Berlin, de Gruyter, (1966).

Humphreys BL, Lindberg DA.The Unified Medical Language System Project. in Lun KC et al. (eds.) MEDINFO 92. Amsterdam: Elsevier Science Publishers (1992).

Kuhn, T. The Structure of Scientific Revolutions. Cambridge Press (1962).

Lakoff, G. The Invariance Hypothesis: is Abstract Reason Based on Image Schemas? *Cognitive Linguistics*, 1 (1990).

Langacker, RW. Concept, Image, and Symbol. The Cognitive Basis of Grammar - Berlin, De Gruyter (1991).

Laresgoiti I, Anjewierden A, Bernaras A, et al. Ontologies as Vehicles for Reuse: a mini-experiment. in B Gaines, M Musen (eds), *Proceedings of KAW96* (1996)

Lenat DB, Guha RV. Building Large Knowledge-based Systems: Representation and Inference in the CYC Project. Menlo Park, Addison-Wesley (1990).

Mac Gregor RM. A Description Classifier for the Predicate Calculus. in *Proceedings of the Twelfth National Conference on Artificial Intelligence, (AAAI 94)* (1994).

Marjomaa E. Different Senses of 'Relevance' in Information Modelling. in H Kangassalo et al. (edd.): Information Modelling and Knowledge Bases, IV. Amsterdam, IOS Press (1993).

Martil R, Turner T, Terpstra P. Knowledge Reuse in Technical Domains: The KACTUS Project. in *Proc. of The Impact of Ontologies on Reuse, Interoperability and Distributed Processing*, Unicom Seminar London (1995)

McCarthy J, Buvac S. Formalizing Context. Stanford Un. Tech. Note STAN-CS-TN-94-13 (1994).

McGuire JG, Kuokka DR, Weber JC et al. SHADE: Technology for Knowledge-Based Collaborative Engineering. *Journal of Concurrent Engineering*, 1 (1993).
Menzies T. Assessing Responses to Situated Cognition. In B. Gaines, M Musen (eds.): *Proceedings of the Conference KAW96, Track on Sharable and Reusable Ontologies* (1996).
Miller GA, Johnson-Laird PN. Language and Perception. Cambridge UP (1976).
Motta E. KBS Modeling in OCML. Modeling Languages for KBS, VU Amsterdam (1995).
Musen M. Dimensions of Knowledge Sharing and Reuse. *Computers and Biomedical Research*; 25 (1992).
Neches R et al. Enabling Technology for Knowledge Sharing. *AI Magazine*; fall:35-56 (1991).
National Library of Medicine. MeSH Medical Subject Headings. Bethesda Maryland: NLM (1996 edition)
Owsnicki-Klewe B. A General Characterization of Term Description Languages. in Bläsius KH & al. (eds.): Sorts and Types in Artificial Intelligence, Springer Verlag, (1989).
Patel-Schneider PF, Swartout B. Draft of the Description Logic Specification from the KRSS group of the DARPA Knowledge Sharing Effort (1993).
Peirce, Charles Saunders. On Signs and the Categories. In I fondamenti della semiotica cognitiva. Torino (1980).
Petitot, J, Smith, B, New Foundations for Qualitative Physics - in JE Tiles, GJ McKee, GC Dean (eds): Evolving Knowledge in Natural Science and Artificial Intelligence, London, Pitman (1991).
Prince G. Narratology. De Gruyter, Berlin (1982).
Rector A, Gangemi A, Galeazzi E, Glowinski A, et al. The GALEN CORE Model Schemata for Anatomy: Towards a Re-Usable Application-Independent Model of Medical Concepts. in Proc. of MIE94 (1994).
Rosch, E, Mervis CB , Gray WD, Johnson DM, Boyes-Braem P. Basic Objects in Natural Categories, *Cognitive Psychology* 8, pp 382-439 (1976).
Rossi Mori A, Gangemi A, Steve G, et al. An Ontological Analysis of Surgical Deeds. in *Proc. of Artificial Intelligence in Europe, AIME97*, (1997).
de Saussure, F. Cours de linguistique générale. Payot, Lausanne (1906/11) Italian Tr., Bari, Laterza (1970).
Simons, P. Parts: a Study in Ontology. Clarendon Press, Oxford (1987).
Sowa JF. Top-Level Ontological Categories. In N Guarino & R Poli (eds.) Formal Ontology in Conceptual Analysis and Knowledge Representation, *International Journal of Human-Computer Studies*, 43 (1995).
Stedman DL. Stedman's Medical Dictionary. Baltimore, Williams & Wilkins (1995).
Steve G, Gangemi A. Ontological Commitment for Medical Ontologies Reuse in the ONIONS Methodology. In B. Gaines, M Musen (eds.): *Proceedings of KAW96, Track on Sharable and Reusable Ontologies* (1996).
Steve G, Gangemi A. Some Theses on Ontological Engineering in the Context of ONIONS Methodology. in A Farquhar (ed.) *AAAI97 Spring Symposium on Ontological Engineering* (1997).
Sujansky W, Altman R. Bridging the Representational Heterogeneity of Clinical Databases. Stanford University Knowledge Systems Laboratory Report KSL-94-07 (1994).
Swartout B, Patil R, Knight K, Russ T. Toward Distributed Use of Large-Scale Ontologies. in B Gaines, M Musen (eds), *Proceedings of KAW96 (Knowledge Acquisition Workshop)* (1996).
Talmy L. The Cognitive Culture System. *The Monist*, 78 (1995).
Tate A. Towards a Plan Ontology. *Journal of the Italian AI Association* (1996)
Uschold M, King M. Towards a Methodology for Building Ontologies. *IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing* (1995).
Valente A, Breuker J. Towards Principled Core Ontologies. in Gaines, Musen (eds), *Proc. of KAW96* (1996)
van Heijst G, Schreiber ATh, Wielinga BG. Using Explicit Ontologies in KBS Development. *International Journal of Human-Computer Studies*, (1997)
Varzi A. Le strutture dell'ordinario. in Logos, teorie dell'essere, teorie della norma, Milano, Giuffre' (1996).
WHO. International Classification of Diseases 10th revision. Geneva: WHO (1994).