MEDINFO 2007
HISA Ltd Australia
© 2007 The authors. All rights reserved.

An Concern-Oriented and Ontology-Based Approach to Constructing Informational Facets of Clinical Trials

Crengua Bogdan[a], Daniela Luzi[b], Fabrizio L. Ricci[b], Luca D. Serbanati[c]

[a]*Ovidius University, Constanta, Romania,* [b]*CNR-IRPPS, Rome, Italy,* [c] *Politehnica University, Bucharest, Romania*

Abstract

The complexity of the clinical trial (CT) process involves high investments in the research of a new drug or procedure. To reduce costs and enhance the research quality more and more activities in the CT process need an IT support. To develop efficient automated solution first a conceptual modeling step should be carried out. To model the CT research partitioning criteria of the CT conceptual domain are required.

The main objective of our research is to identify effective partitioning criteria for modeling the CT research.

A concern-oriented analysis method for developing information systems is applied. The method uses the concerns of various stakeholders of the CT research for partitioning the CT conceptual domain in stakeholder-oriented subdomains. Mental representations description of stakeholder's beliefs and knowledge related to each concern are identified and on their basis, a concern-oriented ontology of the CT can be created.

Facets of the CT future information system are created based on the CT domain ontology. In this paper, an example of how such facets can be constructed for the subject selection section of the CT protocol is presented.

Keywords:

clinical trial, selection criteria, ontology, modeling, concerns

Introduction

The complexity of the clinical trial (CT) process involves high investments in the research of a new drug or procedure. The great amount of workload required to carry out a CT research as well as the request of high quality in data gathering and processing have led to increase the ICT support. To reduce costs and enhance the research quality many software products are available on the market (for example, CRF (Case Report Form) gathering and processing, electronic data capturing, Clinical Trial Management, Clinical Data Management, etc.). The current trend is to expand the ICT support and automatize as many as possible activities in the CT process. But some CT activities as protocol authoring [1], [2] remain largely manual because are highly creative and intensely knowledge-based.

The paper presents a method to interpret CT protocols which allows us a more systematic approach for automatizing of the protocol writing process.

Our solution is founded on the assumption that any automatization is based on an accurate conceptual model of the reality to be automatized. In the CT case many efforts were done to model the CT universe [3]. Generally, conceptual modeling requires partitioning the reality domain under study in loosely coupled parts and for each part a successive partitioning to be carried out until the resulting parts can be easily described. Finding suitable partitioning criteria is a challenge for any system analyst especially when the system under study is a complex one, as in the case of clinical trial.

Our method claims that the partitioning criteria are tightly related to the stakeholders' concerns in developing, operating or evolution of the information system (IS).

In the paper [4] we identify four perspectives from which an IS can be developed: social, functional, informational and technological. In this paper, we present the result obtained from the informational perspective. In this perspective, an IS is viewed as a distributed repository of data, information and knowledge that is used by the business agents.

Our concern-based approach allows us to divide the conceptual universe of the CT in a systematic manner. In this way an ontology of CTs information system can be constructed [5] and on its base we may construct facets of the IS of the future clinical trial, as we defined and exemplified in this paper.

Ontologies

An ontology is a formal specification of the intent of concepts and the intensional relationships that can exist between concepts. Using logical axioms, it is a declarative model of a domain. As a formal system, an ontology allows us to build proofs of assertions viewed as theorems or conclusions. In respect to other models, ontologies allow accurate expression of meaning of models.

According to Guarino's definition, "An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world" [6]. A conceptualization is

a set of conceptual (intensional) relations defined on a domain space [6].

In our research we use DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [7] with its extension D&S (Descriptions and Situations) [8] as a "foundational"[1] ontology where we "plug in" our ontology. DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted to particulars[2]. We adhered to the DOLCE+D&S ontology because: a) DOLCE captures the ontological categories belonging to natural languages and human commonsense; b) being a formal ontology [6], DOLCE and D&S precisely define its notions and relations used; and c) DOLCE and D&S contain all categories and relations we need for writing our ontology.

Concerns

In [9] we defined the concept of *concern of a stakeholder* (briefly *concern*) as a problem-related care of one or more stakeholders involved in the construction or operation of an information system in its natural environment.

Ontologically, a concern is a state of mind arising from preoccupation or interest in a problem that the stakeholder identified in the real world. In our case the stakeholder's interest derives frequently from a need or from his/her responsibility in the IS evolution process, but can also originate in a desire or another interest.

As it is known, on the one hand, in the process of development of an IS, the stakeholders are people who are directly or indirectly affected by it. On the other hand, they influence the system development and/or use. Examples of such stakeholders are direct users, managers of users, management staff members, customers, suppliers, internal departments that support the project, financial departments, and so on.

The problem specification is the pair: a) initial state description (of the current situation, as the stakeholder perceives it) and b) final state description (of the situation that matches expectations, interests, or desires of the stakeholder). These two elements are respectively considered as hypothesis and conclusion of the problem specification. If grouped with the role of the stakeholder who manifests interest or preoccupation about the problem, this pair becomes the high-level specification of a concern that the stakeholder will try to solve. We consider a problem's initial state containing all data, information, and knowledge necessary to obtain the final state of the problem.

In the case of CTs, applying the informational perspective with the objective of identification of subjects' selection criteria, the writing committee has nine concerns. For instance, the high-level specification of one of these concerns is presented in the Table 1.

Tabel 1 - The high-level specification of a concern

C4	*Name*: Care to find how the clinical manifestation of the disease is related to the objectives attainment and the selection criteria identification	
	Problem	*Hypothesis*: Following information is available: trial objectives, trial design, statistical considerations; previous CT projects and their scientific results *Conclusion*: How the disease in the stage treated in the study will be tested in order to get the diagnostic? What basic characteristics (gender, race, life expectancy, performance status) have the study population affected by the disease? How the clinical manifestation of the disease might affect the safety of the subjects and the attainment of the objectives? What particularities of the disease are important for the identifying the selection criteria? How the pathogenesis of the disease might affect the general clinical state of the subjects and the inferring of the selection criteria?
	Stakeholders: Writing committee	

In our approach, each concern has associated a rationale that we explain it in the next subsection.

Concern rationales

We define the concern's rationale a cluster of concepts and their conceptual relations used for the concern solving. In other words, the concern's rationale describes the semantic domain which the stakeholder should manage in order to solve the concern.

The concepts of such a rationale come from informal descriptions of the mental representations of the beliefs of the stakeholder (-s) during the concern's life cycle. The beliefs represent convictions a stakeholder holds for true in a given situation, independently of the nature of the source

1 A foundational ontology characterizes general terms (e.g. entity, event, process, spatial and temporal location) and basic relations (e.g. part-of, qua-lity-of, participation, dependence).

2 Particulars are entities that cannot have instances that are individuals, whereas universals represent ideas or abstract entities which can have instances.

of the conviction: either from perception or an inference from previous knowledge, or a verbally transmitted knowledge. Of course, a belief may be true for one or more stakeholders (thus a piece of knowledge for them), but false for others. When it is accepted as true by all of us (or in all possible worlds), it is a piece of general knowledge.

Regarding a concern, the beliefs to be considered are those associated to the concern's hypothesis and conclusion, but also those which represent milestones in the concern solving process, mainly the hypothesizes and conclusions of concerns the concern depends of. For instance, due to limited space, a belief and a piece of knowledge of the concern C4 are showed in the Table 2.

Table 2 - Mental representation descriptions of a piece of knowledge and a belief of C4

Code	Mental representation description in natural language
K1	The persons who have the disease in the stage considered by the study may be eligible
B10	The existence of the disease and its stage affect the life expectancy of a person

Methods

To obtain a multi-facetted ontology of an IS a concern-oriented method based on stakeholder analysis was proposed in [4] and [5]. The method has 12 steps: 1) identification of stakeholders; 2) identification of concerns; 3) concern classification; 4) identification of relations between concerns; 5) concern prioritization; 6) identification of rationales in concern specifications; 7) identification of the concepts used in the rationales; 8) ontological analysis of the intension of the concepts; 9) choosing a foundational ontology to be extended by our ontology; 10) classification of the concepts conforming the foundational ontology; 11) identification of conceptual relations in concepts' rationales; and 12) definition of the ontology using a formal logic language. The ontology we obtain with this method is structured according to stakeholders' concerns. Its components are vocabularies. In these vocabularies we find either specific and shared concepts. [5] presents how this method ws used to construct a domain ontology for CT subject selection criteria.

The method also includes guidelines for designing informational views of the IS under study from the obtained ontology. For this, other 3 steps should be added to the basic method: 13) construction of the UML ontological model of each piece of knowledge or belief; 14) construction of facets for each concern rationale; and 15) constructon of the informational view by grouping facets of some related concens.

Results

Views and facets

In our approach, a view is a model of an IS related to a particular, homogeneous from a logical point of view, set of concerns. The particularity is due to the fact that the concerns emerge from a particular perspective of the IS developing process: social, functional, informational, or technological. Therefore, depending on the perspective applied, we obtain social, functional, informational, or technological views. We consider the views as simplified models of the future or existing informational system. In this paper, we consider only the informational views.

In our approach an informational view is the structural model of the system to be modeled, basically a UML class diagram [10]. It contains the categories in the system's conceptual domain, their relations, as well as the constraints regarding the model interpretation. An informational view is a cluster of *facets*. Each facet is a simplified model of the informational view and represents a concern-driven projection of the informational view according to a stakeholder's point of view [11]. This point of view is shaped by the stakeholder's role or responsibilities in the future system, in this case, the CT information system. We can say that a facet describes the rationale of a concern. Technically, it is constructed according to a template that contains the following fields: the codes of the facet and concern, the dependency graph of beliefs and knowledge of the concern and the structure of the facet. The later is a UML class diagram that contains the concepts from CT conceptual domain and their ontological relations. In our approach, the structure of a facet is constructed on the basis of the UML ontological models of the knowledge and beliefs of a concern. In the followings, the UML ontological model is defined and some examples of such models are given.

UML ontological model

A UML ontological model is a class diagram that semi-formally describes the semantics of a piece of knowledge or belief of a concern's rationale. Such a model is constructed from the domain ontology of a CT information system by preserving its semantics. It uses concepts of the UML metamodel [10] like class, datatype, association, and dependence. In our research, we found that the correspondence between these concepts and the categories and conceptual relations of the domain ontology are expressed in the following rules:

- all categories of the domain ontology, excepting the abstracts, are mapped to classes; for instance, in the figures 1 and 2, all the categories from the CT domain ontology are mapped to classes.
- categories subsumed abstract category are eventually mapped to the datatype UML concept. As we know, a

data type is a type whose values have no identity [10]. For instance, in the Figure 3 the class AgeValue is a subclass of the AbstractRegion marked by the stereotype datatype, because the property doesn't carry an identity criterion, as it was defined by the methodology OntoClean [12]. In other words, knowing two occurrence of the same value of age they cannot be differentiated. But the time intervals can be differentiated after their duration and the time in which they occurred. That is why the TimeInterval class isn't marked with the stereotype datatype. In addition, TimeInterval is an association class is due to the fact that the corresponding ontological category participates in temporal ontological relations like *participateIn, sequences*, etc.

- all ontological relations, excepting parthood and constitution, are mapped in associations in UML ontological models. As it is known, an association is a relation that describes semantic connections between individuals that are instances of the given classes [10]. For instance, the *refer* association (Figure 1) between the classes DiseaseDescription and DiseaseState maps the ontological relation with the same name which denotes that the DiseaseDescription category, which is subsumed by the Description category, uses the course DiseaseStage, which sequences the referred perdurant DiseaseState.
- temporal and temporary parthood, also constitution relations are mapped in UML aggregation relations. For instance, the instances of the DiseaseDescription class contains as temporary part an instance of the DiseaseName class, but the later can appear in the CT objectives and also in the selection criteria of subject. As it is known, in the case in which the aggregate class has the responsibilities to manage its parts, the aggregation relation becomes a composition one [10].

In order to construct an UML ontological model of a piece of knowledge or belief, our approach proposes the applying of the abstraction mechanism and the following rules on its mental representation description in the natural language:

- if a concept from the mental representation corresponds to a category from the CTs domain ontology, we map this category and the category or categories from the foundational ontology which subsume it into the classes or datatypes. For instance, in order to construct the UML ontological model of the K1 (Figure 1), we applied this rule in the case of the categories Person, Disease, DiseaseStage, NewAdmissibleDiseaseStage and we add to the model the classes AgentivePhysicalObject, Process, Value.
- if a concept corresponds to a quality from foundational or CTs domain ontology, the model contains corresponding class and in addition the class or classes that map the category or categories in which the quality inheres in, because according to DOLCE, each quality is specifically and constantly dependent on the entity it inheres in [7]. For instance, in order to construct the UML ontological model of the B10 (Figure 2), we applied this rule in the case of quality LifeExpectancy and its quale, a value of Abstract Region category.
- in the case of a relation between two concepts, we check if it is an ontological relation. If so, we transform the relation into an UML one, according to the rules above enumerated.
- if the relation between two concepts, excepting the causality one, is not an ontological one, the CT domain ontology is traversed on the basis of the subsumption relation of the corresponding categories of concepts and, on the basis of the reasoning supplied by the ontology, we search the ontological relation that has the same meaning with the initial relation. For instance, in the K1 the relation *have* between Person and Disease is ontologically interpreted by the DOLCE relation of participation ("*x participates in y*" [7]) between those categories. That is why, in the UML ontological model of K1 the name of association between two classes is *participateIn*.
- the causal relation between two concepts is described in model by the dependence relation. For instance, the belief B10 supplies the meaning that the value of the life expectancy of a person depends on the stage of his/her disease. We show this semantics in the UML ontological model from the Figure 2.

In addition, in the UML ontological models of K1 and B10 we observe that these may contain incomplete information, namely those that are particular to the future CT. This information is described by subclasses whose names are prefixed by "New" or map categories from foundational or domain ontology, like AbstractRegion, Activity, and so on.

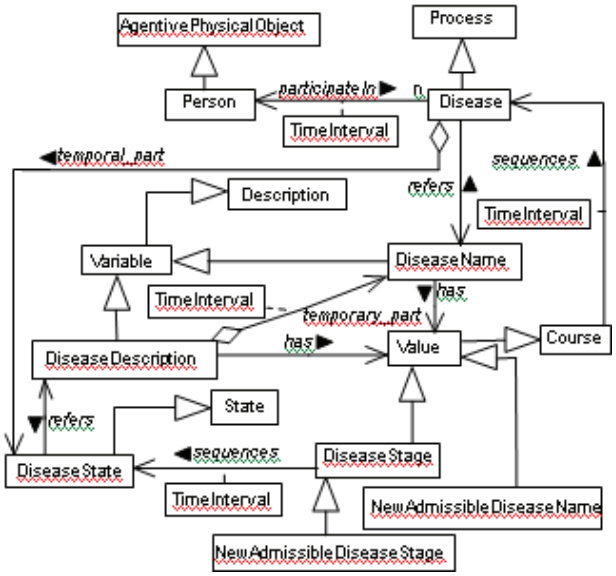

Figure 1 - The UML ontological model of K1

After we constructed the UML ontological models of beliefs and knowledge of a concern, we may construct its facet. In order to construct the facet structure, we again apply the abstraction mechanism on the UML ontological models and we take only the classes that will be part from the future informational system and the ontological relations between them. That is why the facet structure of the concern C4 (Figure 3) doesn't contain the classes that map the categories from the foundational ontology or other like Variable, Value, and so on, that is those have no direct correspondent in the future IS.

Discussion

In this paper, we showed the application in the case of CTs of a concern-oriented analysis method for developing information systems. The method uses the concerns of the writing committee for partitioning the CT conceptual domain in stakeholder-oriented sub-domains. Mental representations of stakeholder's beliefs and knowledge related to each concern are identified and on their basis, a concern-oriented ontology of the CT can be created.

Having the CT domain ontology, we showed how facets of the CT future information system could be created. In this paper, we give examples for a concern emerged from the need of identification of the subject selection criteria.

Some of the advantages supplied of our approach are: a) being a concern-oriented method, it allows to the system analyst to step by step resolve himself or with help of other stakeholders the concerns manifested by the stakeholders and identified from the perspectives of the IS developing. These facts also imply that the concerns of each stakeholder can be resolved in an independent way; b) by definition, the CT domain ontology can be reused in automatized constructing of the CT protocol; c) being constructed based on domain ontology, we think that facets can be automatically constructed using a software tool. In this way, the facets could be used as a verifying and validation tool of the beliefs and knowledge declared by the stakeholders of the future IS.

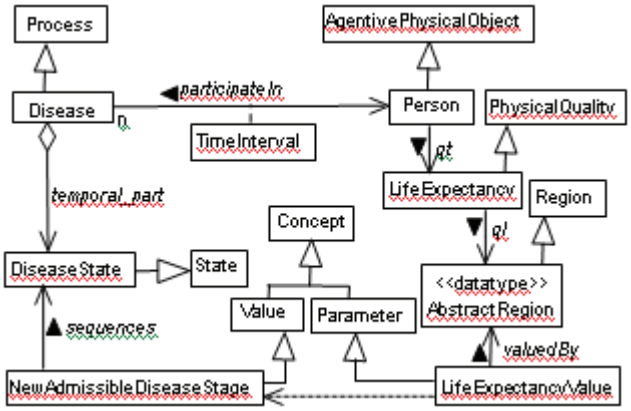

Figure 2 - The UML ontological model of B10

As future research we intend to identify the needed composition rules to construct the views of the information system, and in particular, of clinical trials.

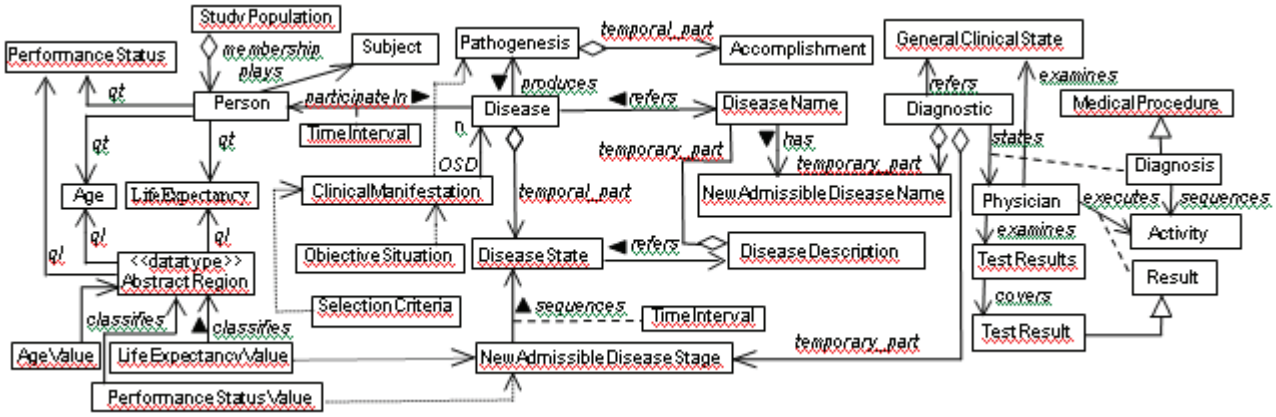

Figure 3 - The facet structure of the concern C4

References

[1] Fazi P, Luzi D, Manco M, Ricci FL, and Vignetti M, Supporting Writing and Diffusion of Clinical Trials using XML and RDBMS. Proc. AMIA Symp. 2002, 240-244.

[2] Modgil S, Hammond P, Wyatt JC, and Potts H, Developing a Knowledge-based Tool for Authoring Clinical Trials Protocols, Proc. EWGLP, 2000

[3] Fazi P, Collada A, Luzi D, Ricci FL, Serbanati LD, and Vignetti M, A Proposed Clinical Trial Model: Analysing the CT Process. *Applied Clinical Trials*, Vol. 15, 1, 2006

[4] Bogdan C and Serbanati LD, Toward a Concern-Oriented Analysis Method for Enterprise Information Systems, IEEE International Multi-Conference on Computing in the Global Information Technology (ICCGI 2006), Bucharest, Romania, 2006

[5] Bogdan C, Luzi D, Ricci FL, and erbnai LD, Towards an Ontology using a Concern-Oriented Approach for Information Systems Analysis, submitted to the Conference on Interoperability for Enterprise Software and Applications, Portugal, 2007

[6] Guarino N, Formal Ontology and Information Systems, Proceedings of Formal Ontology in Information Systems (FOIS 1998) IOS Press, Trento, Italy, 1998

[7] Masolo C, Borgo S, Gangemi A, Guarino N, and Oltramari A, WonderWeb Deliverable D18. Ontology Library. IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web, 2003

[8] Gangemi A and Mika P, Understanding the Semantic Web through Descriptions and Situations, International Conference ODBASE03, Italy, Springer, 2003

[9] Bogdan C and Serbanati LD, Understanding Concerns with Ontologies, submitted to the Conference on Interoperability for Enterprise Software and Applications, Portugal, 2007

[10] OMG, Unified Modeling Language Superstructure, version 2.0, ptc/03-0802, 2003

[11] Serbanati LD, Integrating Tools for Software Development, Yourdon Press Computing Series, Prentice Hall, 1992.

[12] Guarino N and Welty CA, An overview of OntoClean, in S. Staab and R. Studer, editors, Handbook on Ontologies in Inf. Sys., Springer, 2004

Address for correspondence

Crengua Bogdan, Ovidius University, Constanta, Romania;
e-mail: <cbogdan@univ-ovidius.ro>