A Correlation Verb (CorrVerb) Ontology for Object Properties Specification within Knowledge Domain Representations

Christos Pierrakeas
Educational Content, Methodology and Technology Laboratory (e-CoMeT Lab)
Hellenic Open University
Patras, Greece
(0030)2610367730
pierrakeas@eap.gr

Ioannis Panagiotopoulos
Educational Content, Methodology and Technology Laboratory (e-CoMeT Lab)
Hellenic Open University
Patras, Greece
(0030)2610367964
gpanagiotopoulos@ecomet.eap.gr

Achilles Kameas
Educational Content, Methodology and Technology Laboratory (e-CoMeT Lab)
Hellenic Open University
Patras, Greece
(0030)2610367735
kameas@eap.gr

ABSTRACT
An important task when designing educational material for distance learning is the modeling of the teaching domain. However this representation should be formal and semantically standardized in order to be reusable. In this work we propose an ontology-based model for the formal specification of the relations which have been used for the representation of a cognitive domain. This ontology has been used for the representation of the educational material developed for Hellenic Open University (HOU).

Categories and Subject Descriptors
K.3.1 [**Computers and Education**]: Computer Uses in Education – *Computer-managed instruction.*

General Terms
Documentation, Design, Theory, Standardization.

Keywords
Formal Ontology, SUMO, Domain Knowledge Representation, e-Learning, Distance Learning, WordNet.

1. INTRODUCTION
Knowledge modeling plays an important role in the design of educational material for open distance learning systems [1]. The formal representation of the cognitive domain allows tutors to define the learning strategies and learning outcomes for every domain [2]. Additionally, the representation of the concepts facilitates the selection of the appropriate educational material. However, designing a domain of knowledge is a difficult and complicated task. For example, standardizing the semantics of the terms used in such a representation is a complex task. There are many ways to represent a domain of knowledge with different degrees of expressivity. Several programming languages oriented to knowledge representation have been developed, such as Prolog [3]. With the evolution of the Semantic Web other languages and standards have emerged, such as Concept Maps and Ontologies.

Ontologies is a widely-used technique for facilitating understanding and communication between human and software agents as they permit the clear definition and explicit specification of all the basic terms of a specific field [4]. Moreover, ontologies offer a machine-readable representation of concepts and relations between these concepts within a domain, permit the common understanding of the information between humans and software agents and enable the reuse of domain knowledge [5]. Finally, within an ontology all the terms which are included can be described in a semantically rich way with metadata and attributes. The latter is very important in the context of our work as we wish to describe the relations (object properties in terms of ontologies) with as many attributes and characteristics as possible (in the following sections we list these characteristics).

So, we have chosen the notion of the ontology as the most suitable tool to model a cognitive domain's representation. In more details we propose the correlation verb ontology (CorrVerb) in order to standardize the relations (verbs) used in domain knowledge representation.

In HOU we apply a domain modeling methodology (see section 3) based on ontologies, in order to produce representation models for the cognitive domains of the HOU's courses. In this phase we have developed the ontology models that correspond to the cognitive domains for a specific course module (in the near future we will extend the methodology to a large number of course modules).

The need for the development of a formal ontology which describes the relations within a knowledge representation arises from the study of the representation for the course module we mentioned above. This study showed us the incorrect use of relations and thus the modeling of the domain in the next level could lead knowledge engineers into confusion.

The rest of the paper is structured in the following way. In section 2 we present some existing representative upper level ontological models that attempt to define formal relations. In the following section, the study on which we were based in order to build the proposed ontology is presented and we apply the basic steps of the domain modeling methodology. In section 4, we present in details the proposed ontology and also the methodology we have followed. Finally, in section 5 we discuss future work and summarize our conclusions.

2. RELATED WORK

There is a number of top-level ontologies that have been already presented in the literature which attempt to provide a formal representation of the universe of discourse. Below there is a list of the most representative ontologies in this field.

WordNet [6] is a lexical database of English. Initially it was designed as a semantic network based on psycholinguistic principles, while it was further extended with the addition of definitions and it can be used as a dictionary. It can be considered as an upper ontology which includes abstract and more specific concepts, which are connected to each other not only with hierarchy relations but also with other semantic relations like "*part-of*". The General Formal Ontology (GFO) [7] is a top-level ontology that includes elaborations of categories like processes, objects, relations, time, etc. It is primarily designed for applications related to biology, medicine and sociology. Basic Formal Ontology (BFO) [8] focuses on the task to provide an upper ontology which can be used in support with domain ontologies, developed for scientific research. It contains only abstract terms (classes) which could be combined with object properties by other special domains. DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [9] is the first module of the WonderWeb Foundational Ontologies Library (WFOL)[1]. According to the authors DOLCE is not intended as a candidate for a "universal" standard ontology. Instead it is intended to act as a starting point to clarify the hidden assumptions of other existing ontologies. SUMO (Suggested Upper Merged Ontology) [10] and its domain ontologies form one of the largest formal public ontologies nowadays. SUMO is the only ontology that has been mapped to the WordNet lexicon. The OpenCyc platform [11] provides the largest knowledge base with more than 200.000 terms organized in an ontology.

In the context of our work we have chosen not to use directly one of the above mentioned ontologies, because they incorporate a large number of terms (classes and properties) as general knowledge bases. From the above approaches we have used the WordNet lexicon in order to define the synonyms of the verbs (relations) we have included in our model and also to provide their definition. For the definition of the classes we have used in the proposed ontology we have used the class hierarchy presented in the SUMO ontology because as mentioned before, is the only formal ontology where its terms have been mapped to all the WordNet lexicon.

3. DOMAIN KNOWLEDGE REPRESENTION FOR INFORMATICS

In this section we present the basic steps of the methodology we apply for the modeling of a specific cognitive domain and we also discuss the results from the application of this methodology.

[1] http://cordis.europa.eu/projects/rcn/60325_en.html

3.1 Domain Modeling Methodology

We present at this point the basic steps of the methodology we use for the modeling of the cognitive domain. The methodology is described in details in [12]. This methodology is based on the collaboration between domain experts and knowledge engineers and consists of four phases. During the first phase (specification phase) the knowledge engineers discuss about the requirements of the ontology model that is going to be built. Here, face-to-face meetings with the domain experts may take place.

In the second phase (conceptualization phase) the domain experts (tutors) design the cognitive domain (i.e. the basic concepts and the relations between these concepts). During this phase the domain experts make use of drawing tools in order to design the domain, which is then stored automatically in any type of image or XML format.

During the third phase (implementation phase) the ontology engineers based on the outcome of the previous phase, develop two or more ontologies (depending on the number of the tutors participating in the process). The development of the ontologies is followed by the process of merging them into one unified model, which is accomplished in a fully automated way.

The final phase of the methodology includes the evaluation of the final merged ontology from the side of domain experts. A questionnaire that measures the competence of the final model is used in this phase. This questionnaire helps the evaluator (domain expert) to check if the ontology meets a number of predefined criteria such as functionality (if the ontology can be used for a particular task) and completeness (if the ontology includes the minimum necessary number of terms).

The process and development of domain ontologies require not only the knowledge of ontology engineers, but also the knowledge and experience of the domain experts. The above is achieved with the implementation of the aforementioned methodology. Although such a collaborative methodology is not fully automated (apart from the merging process) may require time and effort, however provides more accurate and precise domain knowledge representation.

3.2 Results

After having seen the methodological framework, here we present the results of the application of this methodology for a specific cognitive domain. This domain is the HOU's first year's course module "PLI10 – Introduction to Informatics"[2]. This particular course module covers the fundamentals about programming languages, data structures and programming techniques. It consists of four different domains and for each domain two different domain experts worked. So we had to study eight representations in total.

Below is the list with the most frequently used relations (verbs) for the domain representation of the course module PLI10:

1) *is_a*, 2) *has*, 3) *uses*, 4) *implemented*, 5) *contains*, 6) *represented*, 7) *consists_of*, 8) *applied_to*, 9) *symbolized*, 10) *used_for*.

From the above relations, the most frequently used is the hierarchy relation "*is-a*". However it was noted from the study of the representations the misuse of the relation "*is-a*". Generally the hierarchy relation does not define an instance (or individual). In

[2] http://www.eap.gr/view_en.php?artid=2175

terms of knowledge representation an individual is a specific example of a concept. For that reason we have defined a new relation "*instance_of*" for the connection of an individual with the class (concept) that instantiates it.

The next relation with the highest percentage of usage is the relation "*has*". However, it is a very semantically abstract relation and could cause confusion to the knowledge engineers who are going to model the representation through ontologies. That is why we have introduced two relations: (a) *part_of* and (b) *contains*.

The first one denotes composition while the second one denotes aggregation. Some other general problems that have shown up after the study of the domain experts' representations are: (a) the use of (semantically) different relations in order to relate the same concepts and (b) the use of the same relation in order to connect two different concepts (i.e. different semantics for the same relation). These problems can be solved by: (a) the formal definition of the relations that are used in a domain knowledge representation and (b) with the existence of a list with synonyms, from which list the domain experts will be able to choose the appropriate term (relation) but this relation will be a synonym with a clearly defined term.

From the above brief discussion becomes clear the standardization of the relations that will be available for the domain experts during the representation of a specific knowledge domain is essential. These relations will be completely defined (semantically) and also further characterized with attributes like transitivity, functionality, etc.

In Table 1 we can see the usage percentage of the most frequently used relations we mentioned above, as well as their synonyms according to the WordNet lexicon (grouped by sense of use).

Table 1. Relations and their synonyms

Relation	***Synonyms***	***Usage Percentage***
uses	1) utilizes, applies 2) consumes 3) exploits	30%
is a	-	81%
has	1) holds 2) features 3) owns, possesses	50%
implemented	1) used, utilized 2) applied	20%
represented	1) typified, symbolized 2) expressed 3) embodied	15%
contains	1) includes, incorporates, comprises 2) holds, carries	11%
includes	1) has part 2) adds	11%
symbolized	1) represented 2) intended	11%
used	1) utilized 2) exploited 3) applied	10%
applied	1) used, utilized 2) referred, related	13%

The methodology for the domain knowledge modeling and management we have used is described in details in [12].

4. THE CorrVerb ONTOLOGY

In this section we will briefly describe the proposed CorrVerb Ontology and also the methodology we have followed.

4.1 Development Methodology

In order to develop the ontology we have followed a widely-used methodology proposed in [13]. As far as the ontology's representation is concerned, we have adopted the Web Ontology Language (OWL) [14], which is a W3C [15] standard. More specifically the sublanguage we have used is OWL DL which provides the maximum expressiveness. The development tool we have used is Protégé [16].

The implementation was based on two main axes: (a) the research and study of existing upper ontologies and (b) the study of the representations for the cognitive domains of the first year's module "PLI10 – Introduction to Informatics" as they were given by the domain experts who were responsible for the specific module. In more details, firstly we studied the concepts and the relations between these concepts of the existing upper ontologies, and then we compared all these terms with the terms that were used by the domain experts. In this way we have chosen the basic concepts in the CorrVerb ontology which include the corresponding concepts that the domain experts have used in their own representations.

4.2 The CorrVerb Ontology

The design of proposed ontology that aims to explicitly define the relations between the concepts in a domain knowledge representation has been mainly based on the SUMO ontology which is one of the candidate ontologies for standardization.

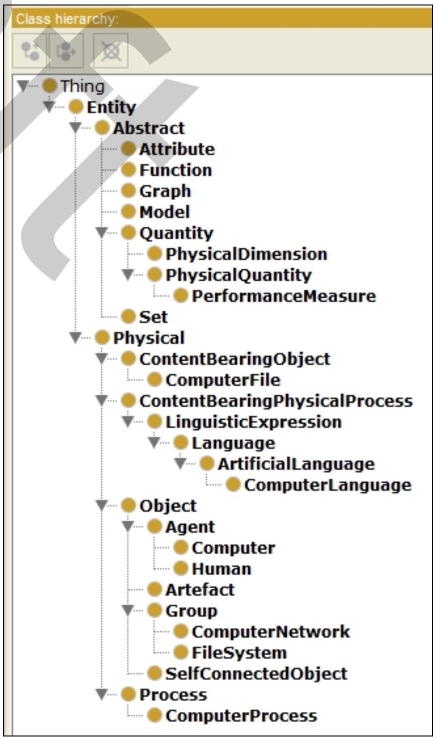

Figure 1. Class hierarchy of the CorrVerb ontology.

In the proposed CorrVerb ontology we define two main upper classes, namely *Abstract* and *Physical*. The first represents all the entities that have a location in space-time. The second includes instances that cannot exist at a particular place without some physical encoding. The whole class hierarchy of the ontology as displayed in Protégé is depicted in Figure 1. It is worth noting here that the hierarchy is not exhaustive as the purpose of the proposed ontology is not to fully describe and standardize all the concepts but to formally define the relations used in the representation of a cognitive domain.

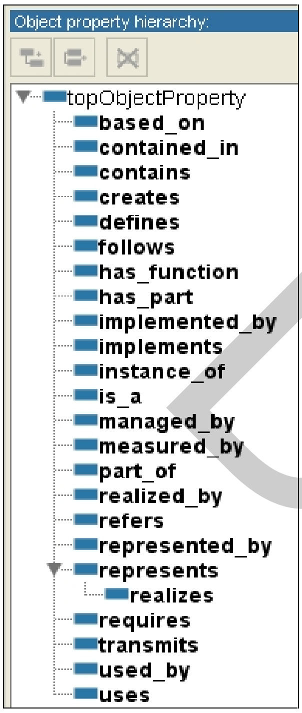

Figure 2. Object properties in the CorrVerb ontology.

Concerning the object properties of the CorrVerb ontology, we propose the characterization of the properties by the domain experts with attributes that the Protégé tool provides, such as:

- *transitivity*
- *equivalence (with other relations)*
- *functionality*
- *inverse (with other relations)*

In Figure 2 are depicted the object properties of the CorrVerb ontology, as displayed in Protégé.

In Table 2 are listed some of the basic relations of the CorrVerb ontology, according to the usage percentage (Table 1).

Table 2. Relations' definitions in the CorrVerb ontology

RELATION	*DESCRIPTION*
is_a	The basic hierarchical relation. If concept A is a sub-concept of concept B, it means that instances of A typically are parts of instances of B.
uses	(uses ?OBJ1 ?OBJ2) means that ?OBJ1 is used by ?OBJ2 as an instrument in a Process.
part_of	The basic mereological relation. (part ?PART ?WHOLE) simply means that the Object (Concept) ?PART is part of the Object ?WHOLE.
contains	This property should be used in the case that the two objects (concepts) are separable. In other case (e.g. a car and its tires) the property of part should be used. (contains ?object1 ?object2) means that ?object1 has a space that is partially filled by ?object2.
represents	(represents ?ENT1 ?ENT2) means that entity ?ENT1 expresses indirectly entity ?ENT2 through an image, form or model. Entity ?ENT1 can be considered as a symbol.

For every object property (relation) defined in the CorrVerb ontology, we have used a number of metadata such as alternative terms (synonyms) and comments (definition, examples, etc). In Figure 3 we can see the metadata for the relation "*represents*".

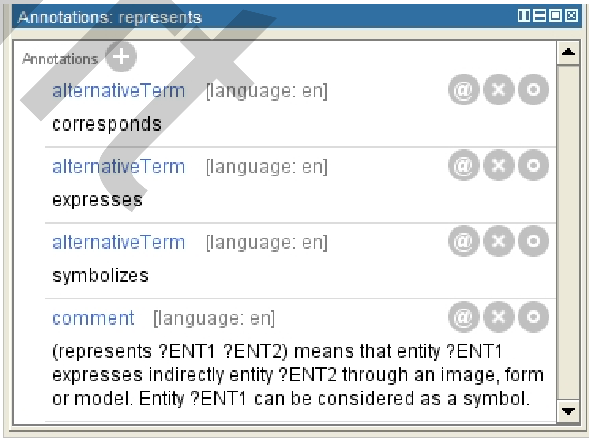

Figure 3. Object property metadata in the CorrVerb ontology.

After having seen the classes and the set of the relations included in the CorrVerb ontology, in the following Table 2 we present the synonyms of the relations we have defined (Figure 2). As in Table 1 the synonyms are grouped by sense of use, according to the WordNet lexicon.

Table 3. Relations and their synonyms in the CorrVerb ontology

RELATION	SYNONYM
based_on	1) grounded, founded 2) depends on
contains	1) incorporates, comprises 2) includes
contained_in	1) included in 2) carried in
creates	1) makes, does 2) designs
defines	1) specify, determine 2) characterize
follows	1) results, comes after 2) complies 3) adopts
has_function	1) has utility 2) operates, performs
has_part	1) consists of
implemented_by	1) applied 2) utilized
implements	1) uses, utilizes 2) applies 3) completes, finishes
instance_of	1) example, case 2) member
is_a	1) exists 2) belongs to 3) represents, symbolizes
manages	1) handles, controls 2) succeeds
managed_by	1) controlled by
measured_by	1) quantified by 2) evaluated by
part_of	1) component, portion 2) region 3) object
refers_to	1) mentions 2) classifies
represents	1) corresponds 2) symbolizes 3) expresses
represented_by	1) symbolized by 2) depicted, pictured by 3) described by 4) expressed by
realizes	1) makes, creates
requires	1) needs, demands, involves 2) depends on 3) expects
transmits	1) communicates, transfers 2) connects 3) moves
uses	1) utilizes 2) exploits
used_by	1) utilized 2) exploited

5. CONCLUSIONS AND FUTURE WORK

In this paper we propose an ontology for describing in a formal way the relations (object properties) which are used within a domain knowledge representation. After careful study of the domain representation for the course module "PLI10 – Introduction to Informatics" and the results described in section 3, it is essential to have a model which explicitly defines the relations which are used in a specific domain representation but also enrich them with additional characteristics (equivalence, transitivity, etc.). This model can be used as a tool by the domain experts while designing a cognitive domain by choosing the right relations.

The current version of the CorrVerb ontology contains 24 basic relations with their synonyms (Table 3). As far as the classes are concerned, we have followed the approach described by the SUMO ontology starting from more abstract concepts and resulting to more specific ones.

In the future, we plan to enrich the proposed ontology with terms (classes and object properties) from cognitive domains other than Informatics. We also work on mapping in an automated way the terms of the proposed ontology with the WordNet lexicon (relations, definitions, synonyms, etc). It is worth noting here that CorrVerb is not a stand-alone ontology, but it is intended to be used as a core part within an application. In more details, we plan to integrate CorrVerb ontology into a web application (tool) where domain experts will be able to see for each relation its definition and also the alternative terms, so that they can easily choose which is the appropriate according to their approach when designing a domain representation. In addition, domain experts will have the ability to propose new relations in the case where the existing ones do not cover the representation of their domain of interest. Along with the new relations, domain experts will also have to denote the metadata for each new property they add (see section 4.2). In this way, we achieve the dynamic maintenance and development of the ontology, which (in our point of view) is an advantage over other static approaches.

6. ACKNOWLEDGMENTS

This research described in this paper has been co-financed by the European Union (European Social Fund - ESF) and Greek National Funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) (Funding Program: "HOU").

7. REFERENCES

[1] Gaeta, M., Orciuoli, F., and Ritrovato, P. 2009. Advanced ontology management system for personalised e-Learning. *Knowledge-Based Systems, Vol. 22*, 292-301.

[2] Kalou, A., Solomou, G., Pierrakeas C., and Kameas, A. 2012. An Ontology Model for Building, Classifying and Using Learning Outcomes. In *Proceedings of the 12th International Conference on Advanced Learning Technologies*, 61-65.

[3] Clocksin, W., Mellish, C. 2003. Programming in Prolog. Berlin ; New York: Springer-Verlag.

[4] Panagiotopoulos, I., Seremeti, L., and Kameas, A. 2010. PROACT: An Ontology-Based Model of Privacy Policies in Ambient Intelligence Environments. *In Proceedings of the 14th Panhellenic Conference on Informatics*, 124-129.

[5] Malik, K. S., Prakash, N., and Rizvi, S. 2011. Ontology Creation towards an Intelligent Web: Some Key Issues Revisited. *International Journal of Engineering and Technology*, Vol. 3, No. 1.

[6] Miller, A., G. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, Vol. 38, No. 11: 39-41.

[7] Herre, H., Heller, B., Burek, P., Hoehndorf, R., Loebe F., and Michalek, H. 2006. General Formal Ontology (GFO). *Part I: Basic Principles, Version 1.0. Deliverable No. 8*. http://www.onto-med.de/en/publications/scientific-reports/om-report-no8.pdf.

[8] Grenon, P., Smith, B., and Goldberg, L. 2003. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In *Proceedings of the Workshop on Medical Ontologies*, Rome. IOS Press, Amsterdam.

[9] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. 2002. Sweetening Ontologies with DOLCE. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, 166-181.

[10] Niles, I., and Pease, A. 2001. Toward a Standard Upper Ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems*.

[11] C. Matuszek, J. Cabral, M. Witbrock, and J. DeOliveira. 2006. An introduction to the syntax and content of Cyc. In *AAAI Spring Symposium.*

[12] Panagiotopoulos, I., Kalou A., Pierrakeas, C. and Kameas A. 2012. An Ontological Approach for Domain Knowledge Modeling and Management in E-Learning Systems. In *Proceedings of the 1st Artificial Intelligence in Education Workshop: Innovations and Applications*, 95-104.

[13] Noy N., and McGuiness D. 2001. Ontology Development 101: A Guide to Creating Your First Ontology, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMi-2001-0880.

[14] OWL: Ontology Web Language. URL: http://www.w3.org/TR/owl-features/

[15] W3C: World Wide Web Consortium. URL: http://www.w3.org/

[16] Knublauch, H., Fergerson, R., Noy, N., and Musen, M. 2004. The Protege OWL plugin: An open development environment for semantic web applications. *In Proceedings of ISWC 2004*, 229–243.