

The Universal Ontology: A Vision for Conceptual Modeling and the Semantic Web (Invited Paper)

Antoni Olivé^(✉)

Department of Service and Information System Engineering,
Universitat Politècnica de Catalunya – Barcelona Tech, Barcelona, Catalonia
antoni.olive@upc.edu

Abstract. This paper puts forward a vision of a universal ontology (UO) aiming at solving, or at least greatly alleviating, the semantic integration problem in the field of conceptual modeling and the understandability problem in the field of the semantic web. So far it has been assumed that the UO is not feasible in practice, but we think that it is time to revisit that assumption in the light of the current state-of-the-art. This paper aims to be a step in this direction. We try to make an initial proposal of a feasible UO. We present the scope of the UO, the kinds of its concepts, and the elements that could comprise the specification of each concept. We propose a modular structure for the UO consisting of four levels. We argue that the UO needs a complete set of concept composition operators, and we sketch three of them. We also tackle a few issues related to the feasibility of the UO, which we think that they could be surmountable. Finally, we discuss the desirability of the UO, and we explain why we conjecture that there are already organizations that have the knowledge and resources needed to develop it, and that might have an interest in its development in the near future.

Keywords: Conceptual modeling · Semantic web · Conceptual schemas · Ontologies · Universal ontology

1 Introduction

This paper puts forward a vision of a universal ontology aiming at solving, or at least greatly alleviating, two important problems in the fields of conceptual modeling and the semantic web.

As a first approximation, by universal ontology (UO) we mean the formal specification of all the concepts that we use and share. This includes the concepts of general use, those that are particular to the existing disciplines, and those specific to any kind of human or organizational activity. The UO specifies the concepts that apply to objects, to their relationships, and to the actions or events involving those objects.

The UO could be a radical solution to at least two important problems. One of them is the problem of the semantic integration of information in the fields of information systems and databases. The problem arises when two or more systems whose

conceptual schemas¹ have been developed independently need to exchange messages or share information. This poses some problems at the physical and syntactic levels, but the most difficult ones are at the semantic level, in which the systems must agree on the meaning of the messages and the data. For some authors, despite its pervasiveness and importance, semantic integration remains an open and extremely difficult problem [1–5]. If all systems were developed using a global schema, such as the UO, that problem would not exist [6].

The other problem is more recent, and it is related to the semantic web. Sometimes it has been called the “tower of Babel” problem [7], but perhaps a more precise name is the “understandability” problem. The root of the problem is that in the web people can build their own web page and say whatever they want on it. This feature has been nicely captured by the AAA slogan: “Anyone can publish Anything on Any topic” [8]. In the case of the semantic web, which is our focus here, we could rephrase the slogan as “Anyone can publish Any fact instance of Any concept”. However, in order to be useful to its target audience (people and machines), the facts must be understood.

There is an understandability problem when a publisher cannot publish a fact in a way that is understood by its full target audience. This happens when the corresponding concept is not included in any ontology shared by the publisher and that audience. Sometimes this is due to the absence of ontologies. In other cases, however, the concept is included in several ontologies, but none of them is shared by all interested parties. To some authors, for the semantic web to be a success it is almost necessary that there is a single, comprehensive ontology [9, p. 476], that is, the UO.

Many authors would agree that the UO could be a solution to the above mentioned problems, but so far it has been assumed that the UO is not feasible in practice. We believe that this was certainly true in the past, but we think that it is time to revisit that assumption in the light of the current state-of-the-art. This paper aims to be a step in this direction.

We try to make an initial, practical proposal of a feasible UO. We describe its scope, the kinds of concepts that could be defined in it, and the specification of each concept. We propose a modular structure of the UO consisting of four levels, and we describe the contents of each level. We argue that the UO requires a powerful mechanism for concept composition, and we sketch a few composition operators. We then tackle a few issues related to the feasibility of the UO and show that currently there are solid reasons to think that they could be surmountable. Finally, we show that there are already organizations that might have in the near future an interest in the UO and the resources needed to develop it.

The rest of the paper is organized as follows. Section 2 presents the scope of the UO, the kinds of concepts that could be defined, and the specification we propose of each concept. Section 3 describes the modularization of the UO and the contents of each module. Section 4 deals with concept composition and explains why it is needed in the UO. Section 5 tackles the issues of the feasibility and desirability of the UO. Finally, Sect. 6 presents the conclusions.

¹ In the context of this paper, we will consider ontologies and conceptual schemas as synonymous.

2 Concept Specification in the UO

In this section we outline the scope, the kinds of the concepts, and the elements that could comprise the specification of each concept of the UO. We will not suggest any particular language construct for the specification of those concepts, focusing instead on their characteristics.

In general, the scope of an ontology depends on its intended objective [10, 11], and the UO is not an exception. The initial objective we propose for the UO is to allow the publication, search, and reading by people and machines of any fact of any domain, using an integrated set of all existing concepts. We make the usual assumption that domains consist of entities and relationships between them. The facts of a domain are the classifications of entities into entity types, and that of relationships into relationship types. Note that the proposed UO does not aim at developing new concepts, but at integrating the existing ones.

Initially, the UO could comprise three kinds of concepts: entity types, n-ary relationship types, and datatypes. Given that n-ary relationship types can be transformed into a set of binary ones [12], which are simpler, we propose to adopt in the UO the two relationship concepts (properties) used in the semantic web: entity properties, which link entities to entities, and datatype properties, which link entities to data values. We will assume that properties have a direction, from subject (domain) to object (range).

The specification of each concept (entity type, entity property, datatype property and datatype) should include at least:

- The kind of the concept.
- The concept identifier. Each concept should have a natural language-independent, unique, and immutable identifier [13]. Among other uses, identifiers are used for defining facts.
- The name and synonym(s) of the concept in each natural language spoken by the UO users. The names of the concepts need not be unique. In general, the name of an entity or data type must be a noun phrase, while the name of a property must be a verb phrase or a noun phrase [14]. When the name of a property is a noun phrase, the property is seen as an attribute (characteristic, feature, ...) of the subject. For example, the properties whose names in English are the nouns *seller*, *product*, *date*, *unit cost*, *total amount*, etc. are attributes. We will see that attributes have a special relevance in the UO.
- The definition of the concept. It may be a natural language description (possibly in each language) or a derivation rule in some formal language.
- The supertypes of the concept (*IsA* relationships).
- The analytical constraints that the instances of the concept must satisfy to be considered universally valid [12]. Such constraints are useful for understanding the meaning of the constrained concepts. Among these constraints there are the allowed domain and range of properties, and the disjointness constraints of concepts.
- The (meta-)entity types of which an entity type is an instance (*InstanceOf* relationships). In general, the UO would not include instances of entity types. The exceptions may be immutable instances of general interest like, for example, *meter InstanceOf Unit*.

As an example, consider the concept whose name in English is *dog*. Its identifier could be *Q144*. It would be defined as an entity type. The concept has one or more names in each natural language. The definition would be an expression in each natural language. *Q144* is involved in several *IsA* relationships, such as (in English) *Dog IsA Animal*. An analytical constraint could be that the sets of instances of *Dog* and *Cat* are disjoint. The concept *Dog* may be defined as an instance of *Species* (*Dog InstanceOf Species*), where *Species* is a meta entity type.

There is a large set of similar properties that are instantiated in many facts. These are the attributes whose name is the name of an entity or data type. Almost all names of entity and data types can be in some context the name of an attribute of some subject. This observation leads us to propose to make in the UO the “*assumption of implicit attributes*”, by which we mean that for each entity or data type defined in the UO there is an implicit attribute property with the same name. The domain of the property is the top-level entity type (such as *Entity*) and its range is the corresponding entity or data type. For example, if *Seller* is the name (in some language) of an entity type, then we assume that there is an implicit attribute whose name (in the same language) is *seller*, domain *Entity* and range *Seller*.

The assumption of the implicit properties is of practical importance because it may save a great deal of effort in the definition of the UO. As an example, the great majority of the properties defined in schema.org² would be implicit in the UO, because their name would coincide with those of entity and data types, or could be composed from them, as will be explained in Sect. 4.

We note that in schema.org (and in many other ontologies) the definition of a property includes its domain and range, which may be different from that assigned by the above assumption. However, this might not be important in the proposed UO, since its envisaged objective does not include the control of the validity of facts nor the reasoning from facts.

3 The Structure of the UO

The UO is huge; therefore, in order to be manageable, it must be modularized [15]. We distinguish four levels of concepts, and we group the concepts at the same level into a module [16]. The levels are: Conceptual Model, Foundational, General, and Domain. Each of these levels is briefly described below. If we arrange vertically the levels, and populate each level, the result can be seen as a pyramid, which we call the UO pyramid (Fig. 1).

There must be an organization that has an overall responsibility for the UO. We call UO regulator to this organization. Other regulators would have a responsibility for specific parts of the UO. In the description that follows we will indicate the role of each regulator in each level.

² <http://schema.org/>.

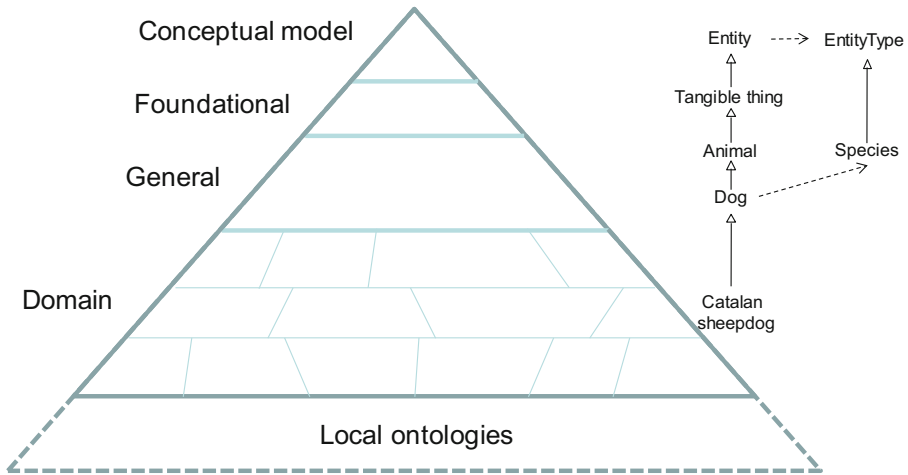


Fig. 1. The levels of the UO pyramid

3.1 The Conceptual Model Level

The conceptual model (or ontology model) level comprises the meta types and the direct or indirect supertypes of all the concepts in the UO [17, 18]. The concepts at this level are used to define the rest of the UO. Figure 1 shows two example concepts at this level, that we have called *Entity* and *EntityType*, with an *InstanceOf* relationship between them. *Entity* would be the supertype of all entity types, while *EntityType* would be the supertype of all meta entity types.

Each conceptual/ontology model (such as, for example, UML, ER, RDFS or OWL) includes the concepts needed for this level of the UO. For example, in RDFS the concepts would be *Class*, *Resource*, *Datatype*, *Literal*, *Property*, etc. [19]. In any case, the number of concepts at this level is very small.

Based on the existing conceptual/ontology models, it should not be difficult to reach an agreement on the concepts to be included in the proposed UO. The concepts at this level should be under the responsibility of the UO regulator.

3.2 The Foundational Level

The foundational level, which is also small in size, includes abstract concepts that have been proposed in the foundational ontologies [20, 21], such as DOLCE [22] or UFO [23, 24]. The concepts at this level cannot be directly instantiated to publish facts and, therefore, they are not essential in the proposed UO. However, they may be useful for clarifying the semantics of other concepts, for defining only once knowledge that is common to several concepts, and for reasoning purposes. As an example, UFO makes a fundamental distinction between individuals that are *Endurant* and *Event*. Another example is the concept *TangibleThing*, a subtype of *Entity*, shown in Fig. 1. In general, the definitions and constraints of these concepts are “inherited” by all the concepts that are defined as their direct or indirect subtypes.

There are several foundational ontologies, each of them useful in some contexts. A practical approach to the inclusion of such ontologies in the UO could then be the one proposed in the Wonderweb vision [20]. The basic idea would be that the foundational level consists of a library of selected foundational ontologies. The library would include the specification of the links between the ontologies and the mapping (mainly the *IsA* relationships) of each ontology with the concepts in the general level.

The management of the library would be the responsibility of the UO regulator. However, each foundational ontology should have a specific regulator, with the responsibility for proposing the links with the other ontologies and the mappings with the general level.

3.3 The General Level

Most linguists make a basic, informal distinction between language for general purposes (LGP) and language for special purposes (LSP) [25] and also between their corresponding dictionaries. General dictionaries contain those words of the language which are of general use, representing various spheres of life and presenting a complete picture of the general language. They are meant for the general user of the language. Special dictionaries either cover a specific part of the vocabulary or are prepared with some definite purpose [26].

Going from words to concepts, it seems natural to establish a similar basic, informal distinction between concepts for general purposes and concepts for special purposes. We can then include the former in the general level of the UO pyramid and the latter in the domain level. In the example of Fig. 1, there are three concepts in the general level: *Animal*, *Dog* and *Species*, with *Dog IsA Animal* and *Dog InstanceOf Species*.

The concepts at the general level are subtypes or instance of concepts at the conceptual model level, and, possibly, of concepts at the foundational level.

There are several ontologies that could provide an excellent basis from which to build the general level of the UO. Among them, we mention here WordNet [27], SUMO [7], CYC [28], and BabelNet [29]. For the purposes of illustration, in the following we will assume WordNet.

WordNet defines noun, verb and adjective synsets that may be the source of the entity types and properties of the UO. WordNet 3.0 comprises over 80,000 noun synsets (concepts), which include most (if not all) entity types that have a name in the English LGP. By the application of the assumption of implicit attributes, there would also be an implicit attribute for each entity type. There are already “wordnets” in many languages [30], which include links to the English WordNet.

WordNet comprises also over 13,000 verb synsets, which include most (if not all) properties that have a name in verb form in the English LGP.

Finally, WordNet comprises also over 18,000 adjective synsets, most of which can be considered as Boolean properties. For example, the adjective synset *local#2*, with the gloss “of or belonging to or characteristic of a particular locality or neighborhood”, could be considered as a Boolean datatype property, with its own identifier and English name “is local”.

In order to guarantee the consistency of the general level, its creation and evolution should be under the responsibility of the UO regulator.

3.4 The Domain Level

In the UO pyramid, the domain level contains the concepts for special purposes corresponding to the LSP. Therefore, this level contains all existing domain ontologies. Since there are many domain ontologies, some of which very big, the domain level includes in total several millions of concepts. Achieving a satisfactory arrangement of these ontologies is the main technical challenge of the UO. Figure 1 shows an example of concept at this level, *CatalanSheepdog*, which is a subtype of the concept *Dog* at the general level.

The concepts at the domain level are subtypes or instance of concepts at the general level, and, possibly, of concepts at the foundational level.

An ontology can be a part of the domain level of the UO if its mappings with the rest of the UO are defined. There are two kinds of mappings: vertical and horizontal. The vertical mappings define the correspondences between that ontology and the concepts at the general level. The horizontal mappings define the correspondences between that ontology and the other ontologies at the domain level.

In both mappings, a correspondence is a relationship between two concepts. In general, it can be an equivalence (the concepts are the same), an *IsA* (a concept is a subtype of the other) or a disjointness (no entity -or property- can be an instance of both concepts) [31]. Equivalent concepts are considered the same and, therefore, the equivalence correspondences are ignored.

The mappings must preserve the completeness of the UO [12, 32, 33]. In our context, this implies that the two following conditions are satisfied at any time:

- “1. Let C_1 and C_2 be two concepts in the UO. If in the real-world the instances of C_1 must necessarily be also instances of C_2 , then in the UO there must be a direct or indirect subtype correspondence between C_1 and C_2 .”
- “2. Let C_1 and C_2 be two concepts in the UO. If in the real-world the instances of C_1 cannot be also instances of C_2 , then in the UO there must be a direct or indirect disjointness correspondence between C_1 and C_2 .”

Satisfaction of the first condition guarantees, among other things, that users querying the instances of C_2 will get also the instances of C_1 even if users are unaware of the existence of C_1 in the UO. Satisfaction of the second condition is mandatory in an open world assumption of the UO.

Each domain ontology should be under the control of a specific regulator. To include a new ontology in the domain level, its regulator should provide the vertical mappings with the concepts at the general level. The UO regulator should review and approve those mappings before the “official” adoption of the new ontology within the UO.

The new ontology may overlap with one or more ontologies already existing in the domain level. Therefore, it is necessary to discover those ontologies and to define the corresponding mappings. The discovery of the set of potentially overlapping ontologies can be automated to a great extent by using the previous vertical mappings [34]. For each potentially overlapping ontology, it will be necessary to define the correspondences with the new ontology. Ideally, this should be done by the regulators of both ontologies. Existing or future matching systems should be of great help in determining

both the potentially overlapping ontologies and their correspondences [31, 35]. A recent example of the use of matching systems for automatically determining the mappings between ontologies is described in [36].

3.5 Local Concepts

Local concepts are specializations of concepts defined in the UO, but that are not part of it. Local concepts are not intended to be generally shared. Through time, a local concept may evolve and become part of the appropriate level of the UO.

There will always be a strong need of local concepts. However, many of them could be defined as a composition of other concepts already defined in the UO. If there were a mechanism for defining and using compound concepts that would not require their inclusion in the UO, the need of local concepts could be significantly decreased. We deal with this in the next section.

4 Concept Composition

The UO described above specifies only a limited (even if very large) number of concepts. However, it is a fact that using an appropriate set of composition operators we could compose a limitless number of concepts from them. We call *core* UO the explicitly defined ontology, and *extended* UO the set of concepts that could be composed from the core. The *full* UO would then be the union of the core and extended parts.

The concepts of the extended UO could be used in the publication and query of facts, markup of web pages, conceptual schemas, database schemas, and similar places, like those of the core UO. The crucial point is that such use would be done without the explicit inclusion of the composed concepts in the core UO. Composed concepts are defined when and where used.

There is an insightful parallelism between the UO and a human language. Human languages are usually described as consisting of two parts: a *lexicon*, a catalogue of a limited number of words, and a *grammar*, a system of rules which allow for the combination of those words into a limitless number of sentences. Applying this parallelism, the lexicon would be the core UO, the grammar the set of composition operators, and the sentences the full UO.

It is surprising that concept composition, as indicated above, has been used so little in the conceptual modeling and semantic web fields, especially if one takes into account that some of the languages used in those fields (such as OCL and OWL) allow the definition of compound concepts.

One of the few exceptions is SNOMED CT [37], which is a controlled vocabulary for the clinical domain. SNOMED CT provides a mechanism that enables clinical phrases (facts) to be represented, even when a single SNOMED CT concept does not capture the required level of detail. This is important as it enables a wide range of clinical meanings to be captured in a record, without requiring the terminology to include a separate concept for every detailed combination of ideas that may potentially need to be recorded.

We believe that the UO could achieve its intended objective only if there is a powerful set of composition operators that allows defining and using the concepts in the extended UO. In rest of this section, we sketch only three of these operators: two inspired in compound nouns (Sect. 4.1) and one based on aggregate functions (Sect. 4.2).

4.1 Compound Nouns

Word compounding is a mechanism we use to generate a limitless number of words from an existing, limited, lexicon. Word compounding has been widely studied in linguistics [38]. Similarly, concept combination is a mechanism we use to generate a limitless number of concepts from a limited number of existing ones. Concept combination has been studied in cognitive psychology and cognitive science [39].

In linguistics, a compound consists of the concatenation of two or more words. A compound may be of any syntactic category, but in this paper we will only deal with compounds that are nouns that correspond to entity types (such as *lodging business*). From a morphological point of view, English noun compounds can be open (as in *lodging business*), hyphenated (as in *world-beater*) and closed (as in *sheepdog*) [40].

There exist several different classes of noun compounds. In this section we will focus on the endocentric compounds, which are the most frequent in English [41]. An endocentric compound noun W consists of a head H , which is a noun, and a modifier M . The noun W is more specific than H , and therefore it holds that $W \text{ IsA } H$ [36]. In English, the modifier M is normally a noun, an adjective or a verb, as illustrated by the following examples from schema.org:

- Noun: *Flight reservation, Government organization, Tourist attraction.*
- Adjective: *Financial product, Local business, Medical organization.*
- Verb: *Sell action, Send action, Receive action.*

Based on this, in the following we propose two concept composition operators of entity types.

Entity-Property Composition. The entity-property compound is analogous to the above adjective-noun compound. Let E_i be an entity type and let P_j be a datatype property whose range is *Boolean*, and such that E_i is in the domain of P_j . Then we denote by $EP(E_i, P_j)$ the compound entity type whose instances are the instances of E_i for which P_j is true.

As an example, consider the entity type that corresponds to a local business. This type is not defined in WordNet, but we could define it as an *EP* composition from two concepts defined in it: the noun synset *business#1* and the adjective synset *local#2*. The noun synset *business#1* corresponds to an entity type, while the adjective synset *local#2* would corresponds to a *Boolean* property. Then, $EP(\text{business\#1}, \text{local\#2})$ would be the compound entity type whose instances are the instances of *business#1* for which *local#2* is true.

Note that the *EP* operator is language independent. We do not suggest here a user-friendly notation for compound concepts. The point that we want to make here is

that the expression $EP(E_i, P_j)$ (or some equivalent notation or name, see below) can be used like any other entity type of the core UO, even if it is not explicitly defined in it.

The expression $EP(E_i, P_j)$ would be the identifier and the default name of the compound concept. In general, however, these names are not user friendly. A better option could be the use of naming functions. There could be a naming function F_C for each composition operator C , such that $F_C(C_{exp}, L)$ gives a name of the concept obtained by the expression C_{exp} using the operator C in language L . For example, $F_{EP}(EP(\text{business}\#1, \text{local}\#2), \text{English})$ could give the name “Local business”.

From the definition it follows that $EP(E_i, P_j) \text{ IsA } E_i$. On the other hand, if P_j and P_k are two properties such that $P_j \text{ IsA } P_k$ then it follows that $EP(E_i, P_j) \text{ IsA } EP(E_i, P_k)$.

In large ontologies there are many compound concepts that could be defined using EP compositions from noun and adjective synsets included in WordNet. A prominent example may be Microsoft Concept Graph (MCG), which contains above five million concepts, most of which named with a compound [42]. In MCG there are over 1.3 million concepts that could be defined by means of the EP operator, using over ten thousand adjective synsets.

Entity-Property-Entity Composition. Let E_1 and E_2 be entity types, and let P_j be an entity property such that E_1 is in the domain of P_j and E_2 is in the range of P_j . Then, we denote by $EPE(E_1, P_j, E_2)$ the entity type whose instances are the instances of E_1 for which the value of P_j includes instances of E_2 [39, 41].

As an illustration, consider the following examples, involving noun and verb synsets of WordNet:

- *Toy store* can be defined as $EPE(\text{store}\#1, \text{sell}\#1, \text{toy}\#1)$. Then, an instance of that compound concept is a store that sells toys.
- *Dog magazine* can be defined as $EPE(\text{magazine}\#1, \text{deal}\#1, \text{dog}\#1)$. Then, an instance of that compound is a magazine that deals with dogs.
- *Flu virus* can be defined as $EPE(\text{virus}\#1, \text{cause}\#1, \text{flu}\#1)$. Then, an instance of that compound is a virus that causes flu.

From this definition it follows that $EPE(E_1, P_j, E_2) \text{ IsA } E_1$. Furthermore, if P_j and P_k are two properties such that $P_j \text{ IsA } P_k$ then $EPE(E_1, P_j, E_2) \text{ IsA } EPE(E_1, P_k, E_2)$. Finally, if $E_2 \text{ IsA } E_3$, then it follows that $EPE(E_1, P_j, E_2) \text{ IsA } EPE(E_1, P_j, E_3)$. For example, $EPE(\text{magazine}\#1, \text{deal}\#1, \text{dog}\#1) \text{ IsA } EPE(\text{magazine}\#1, \text{deal}\#1, \text{domestic animal}\#1)$.

The analogous construct in linguistics is the noun-noun compound. However, there is an important difference: in natural language, the property that connects the two nouns of a noun-noun compound is not specified. This fact leads to ambiguities in some cases. We do not suggest here any example of naming function for this operator.

In MCG, there are over three million compound concepts that could be defined using EPE compositions from noun synsets defined in WordNet. The number of noun synsets that would be used is over thirty thousand.

4.2 Count Composition

A very large set of frequently used properties give the result of aggregate functions [43]. For example, the datatype property that gives the number of employees of a

company. It is practically impossible to define all those properties in the core UO, but they can be easily defined when needed by means of composition operators. In the following we sketch the operator corresponding to the count function. Others could be defined similarly.

Let P_j be a property with domain E_1 and range E_2 . Then we denote by $Count(P_j)$ the datatype property with domain E_1 and range *Integer* that gives the number of instances of type E_2 that are related through P_j to an instance of E_1 .

For example, assuming the implicit attribute corresponding to the WordNet noun synset *employee#1*, the operator $Count(employee\#1)$ is the datatype property that gives the number of employees of a given instance of its domain. The domain of $Count(employee\#1)$ would be *Entity*, and the range *Integer*.

$Count(employee\#1)$ (or an equivalent notation) would be the identifier and the default name of the datatype property. A better name could be obtained by the corresponding naming function F_{Count} , which in this case could give, for example, $F_{Count}(Count(employee\#1), English) = \text{“number of employees”}$.

5 Feasibility and Desirability of the UO

Once we have analyzed a possible basic structure of the UO and shown how it could be extended by means of the composition operators, in this section we tackle the issues of the feasibility and desirability of the UO.

5.1 Feasibility

Terminology. Some authors point out that “different communities of practice use the same terms with quite different meanings” [13], which can be a problem for the references to the concepts of the UO. This is the well-known problem of homonymy and/or polysemy in natural languages. In the proposed UO, each concept has a unique identifier, which is used in the publication of facts. Each concept has also a name and a set of synonyms in each language, which can be used in the external references to the concept. As in, for example, WordNet, the name and synonyms may not be unique, but people should be able to solve ambiguities by means of the definition of the concept or its composition expression.

Agreement. Some authors think that it would be very difficult to reach an agreement on the UO because it is very large and diverse. The following excerpts are representative of these views:

- “Although some may think the solution is to come up with a single context for the whole world... in reality this is extremely difficult for any complex organization” [44]
- “... people will always disagree about what terms to use and how to define them, a global ontology will always be seen as flawed” [13]
- “It is of course unrealistic to hope that there will be an agreement on one or even a small set of ontologies” [45]

- “Enforcing one centralized global ontology... is impractical to develop an ontology with consent from the user community at large” [54]
- “A single huge ontology of everything is difficult to accomplish, as the effort of getting consensus on it becomes unimaginable” [46].

Our general response to these views is that the proposed UO would not be built from scratch, but it would integrate existing concepts and ontologies. The concepts to be included in the UO are not new; they have been already agreed, defined, and are currently used by people and organizations. In the following, we detail this response for each UO level.

Technically, it should not be difficult to agree on the conceptual model level. The ambition of the proposed UO is quite limited, and therefore a subset of the existing conceptual/ontology models would suffice.

In the foundational level, several foundational ontologies could coexist. We have shown that there is no need to select only one of them. The only requirement is that each ontology includes the mappings to the other ontologies in the same level, and with the general level. The addition of foundational ontologies can be done incrementally.

The concepts to be included in the general level have already been specified in several places, notably in WordNet and in similar ontologies. The names of these concepts in many languages are known. There are satisfactory definitions of most of them, and their *IsA* relationships are known in most cases. It is safe to say that there exists already a substantial agreement on the concepts of the general level among their users.

The domain level would include all public domain ontologies. There are many, but the concepts of each of them have been already defined and agreed by their regulators. The problem may be the definition of the mappings of these ontologies with the general level and with the other ones in the same level. As we have mentioned in Sect. 3.4, this is the main technical challenge of the UO. The addition of domain ontologies can be done incrementally.

Management. Some authors argue that the management of the UO would be very difficult: “A huge, central ontology would be unmanageable” [47]; “Even if initial agreement were reached, there are many maintenance issues to be faced” [13].

The management of the UO may be difficult, but some of the management approaches that have been applied to successfully build similar artifacts could be appropriate for a UO with a modular structure. Besides the large ontologies, examples of such artifacts may be the open source projects [48], the Oxford English Dictionary (over 600,000 words³), the Encyclopedia Britannica, or UMLS (over 3.4 million concepts⁴). Of particular interest could be the approach taken in the development of schema.org [49].

Redundancy and Usability. There are a few problems in the UO that are also present in the natural language field. In principle, techniques developed in this field for dealing with those problems could be adapted in the UO context. Among those problems, we

³ <http://www.oed.com/>.

⁴ https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html.

mention here construct redundancy [50] and usability [51]. Ideally, there should be little redundancy in the core UO, but it is likely to have more in the extended UO, because sometimes the same concept can be expressed by means of several combinations of composition operators. A similar problem in natural language is that a concept may have several compound names.

On the other hand, there may be usability problems, because in a large ontology it may be difficult to find the most appropriate concept for a particular situation. A similar problem occurs with the lexicons.

5.2 Desirability

Some authors have expressed in the past the view that a global ontology would be desirable. The following excerpts are representative of these views:

- “some may think the solution is to come up with a single context for the whole world” [44]
- “In theory, a good solution to this problem would be to adopt a single global vocabulary that is widely accepted and embraced by everyone in the organization” [13]
- “one centralized global ontology prevents semantic heterogeneity since no more ontology exists and everyone is using the same ontology” [54]
- “for the semantic web to be a success, would it be nice, or almost necessary, that we could have just one single ontology, which actually covers all the common things in life?” [9].

We certainly agree that the UO is (highly) desirable but, of course, we must take into account its cost. Some authors have indicated that “the creation and maintenance of such an ontology is usually prohibitively expensive” [54]. Socio-economic factors dictate reality, and therefore the obvious question is “Who will develop the common ontologies and will they invest the effort and then allow them to be used for free?” [52].

It is not possible to give a precise answer to that question. However, we believe that there are currently already organizations that might have in the near future an interest in the UO and the resources needed to develop it. We mention two of them here. The first are standard organizations that have experience in the development of large standards, and have a means to get the resources needed. One of these organizations could be the World Wide Web Consortium (W3C), which is the main international standards organization for the web. For W3C, the UO would be a natural continuation of the many standard ontologies and languages that have developed so far. On the other hand, it could not be difficult for W3C to get the resources from its member organizations.

Search engine companies, which already have the knowledge and experience in building large knowledge graphs, might also be interested in the development of the UO. A clear indication of this is schema.org. It was built with the collaboration of the major search engines Bing, Google, and Yahoo (later joined by Yandex). Schema.org was launched in 2011 with 297 classes and 187 relations, and since then its size and adoption level have been increasing continuously [49]. Both webmasters and search engines have a strong interest in the schema.org markup. The former can publish the contents of their websites in a way that is understood also by the search engines. The

latter can provide much better results to the search requests. The interest might be so high that “the question often comes up whether schema.org is an end-all solution for defining terminology for the Semantic Web” [53].

6 Conclusions

We have put forward a vision of a universal ontology (UO) aiming at solving, or at least greatly alleviating, the semantic integration problem in the field of conceptual modeling, and the understandability problem in the field of the semantic web. The semantic integration problem arises when two or more systems, whose conceptual schemas have been developed independently, need to exchange messages or share information. The understandability problem arises when the structured data published in datasets or in webpages cannot be understood by its full target audience (people and machines).

So far, it has been widely accepted that the UO would be a solution for those problems, but, at the same time, it has been assumed that it is not feasible in practice. In this paper, we have challenged that assumption. We have argued that in the current state-of-the-art it could be feasible to build a UO that solves those problems to a great extent. We have made an initial proposal of a UO able to achieve a limited objective, but useful for the (big) problems intended to solve.

We have explained the kinds of concepts that could be defined in the UO, and the minimum specification we propose of each concept. We have proposed also a modular structure for the UO, with four levels. We have also shown that the UO needs a powerful mechanism for concept composition, which we have sketched.

We have tackled a few issues related to the feasibility of the UO, such as terminology, agreement and management, and we have shown that although they are important, there are solid reasons to think that they are currently surmountable. Finally, we have discussed the desirability of the UO, and we have shown that there are already organizations that might have in the near future an interest in the UO, and the knowledge and resources needed to develop it.

Acknowledgments. The author is greatly indebted to his colleagues Albert Abelló, Jordi Cabot, Ernest Teniente, and Toni Urpí for their comments to earlier drafts of this paper. This work has been partially supported by the Ministerio de Economía y Competitividad, under project TIN2014-52938-C2-2-R.

References

1. Batini, C., Ceri, S., Navathe, S.: *Conceptual Database Design: An Entity-Relationship Approach*. Benjamin-Cummings Publishing Company Inc., Redwood City (1992)
2. Doan, A., Halevy, A., Ives, Z.: *Principles of Data Integration*. Morgan Kaufmann, Burlington (2012)
3. van Harmelen, F.: *Ontology mapping: a way out of the medical tower of babel?* In: Miksch, S., Hunter, J., Keravnou, E.T. (eds.) *AIME 2005*. LNCS, vol. 3581, pp. 3–6. Springer, Heidelberg (2005). doi:[10.1007/11527770_1](https://doi.org/10.1007/11527770_1)

4. Heath, T., Bizer, C.: *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology, 1st edn., vol. 1, no. 1, pp. 1–136. Morgan & Claypool (2011)
5. Park, J., Ram, S.: Information systems interoperability: what lies beneath? *ACM Trans. Inf. Syst.* **22**(4), 595–632 (2004)
6. Uschold, M., Grüninger, M.: Ontologies and semantics for seamless connectivity. *SIGMOD Rec.* **33**(4), 58–64 (2004)
7. Pease, A., Niles, I., Li, J.: The suggested upper merged ontology: a large ontology for the semantic web. In: *Proceedings of AAAI-2002 Workshop on Ontologies and the Semantic Web* (2002)
8. Allemang, D., Hendler, J.A.: *Semantic Web for the Working Ontologist - Effective Modeling in RDFS and OWL*, 2nd edn. Morgan Kaufmann, Burlington (2011)
9. Yu, L.: *A Developer's Guide to the Semantic Web*. Springer, Heidelberg (2014). doi:[10.1007/978-3-662-43796-4](https://doi.org/10.1007/978-3-662-43796-4)
10. Gruninger, M., Fox, M.S.: Methodology for the design and evaluation of ontologies. In: *Proceedings of Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI 1995* (1995)
11. Noy, N.F., McGuinness, D.L.: *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford University (2001)
12. Olivé, A.: *Conceptual Modeling of Information Systems*. Springer, Berlin (2007). doi:[10.1007/978-3-540-39390-0](https://doi.org/10.1007/978-3-540-39390-0)
13. Uschold, M.: Creating, integrating and maintaining local and global ontologies. In: *Proceedings of First Workshop on Ontology Learning OL 2000 in Conjunction with the 14th European Conference on Artificial Intelligence ECAI 2000* (2000)
14. Aguilera, D., Gómez, C., Olivé, A.: A complete set of guidelines for naming UML conceptual schema elements. *Data Knowl. Eng.* **88**, 60–74 (2013)
15. Spaccapietra, S. (coordinator): Report on Modularization of Ontologies. Technical report, Knowledge Web Deliverable D2.1.3.1 (2005)
16. Roussey, C., Pinet, F., Kang, M.A., Corcho, O.: An introduction to ontologies and ontology engineering. In: Falquet, G., et al. (eds.) *Ontologies in Urban Development Projects*. Advanced Information and Knowledge Processing, vol. 1, pp. 9–38. Springer, London (2011). doi:[10.1007/978-0-85729-724-2_2](https://doi.org/10.1007/978-0-85729-724-2_2)
17. Mylopoulos, J., Borgida, A., Jarke, M., Koubarakis, M.: Telos: representing knowledge about information systems. *ACM Trans. Inf. Syst.* **8**(4), 325–362 (1990)
18. de Carvalho, V.A., Almeida, J.P.A., Fonseca, C.M., Guizzardi, G.: Multi-level ontology-based conceptual modeling. *Data Knowl. Eng.* **109**, 3–24 (2017)
19. Brickley, D., Guha, R.V.: RDF Schema 1.1. W3C Recommendation (2014). <http://www.w3.org/TR/rdf-schema/>
20. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: *WonderWeb Deliverable D18, Ontology Library (final)*. ICT Project (2003)
21. Pastor, O.: Conceptual modeling of life: beyond the homo sapiens. In: Comyn-Wattiau, I., Tanaka, K., Song, I.-Y., Yamamoto, S., Saeki, M. (eds.) *ER 2016*. LNCS, vol. 9974, pp. 18–31. Springer, Cham (2016). doi:[10.1007/978-3-319-46397-1_2](https://doi.org/10.1007/978-3-319-46397-1_2)
22. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A.: Sweetening WORDNET with DOLCE. *AI Mag.* **24**(3), 13–24 (2003)
23. Guizzardi, G., Wagner, G., de Almeida Falbo, R., Guizzardi, R.S.S., Almeida, J.P.A.: Towards ontological foundations for the conceptual modeling of events. In: Ng, W., Storey, V.C., Trujillo, J.C. (eds.) *ER 2013*. LNCS, vol. 8217, pp. 327–341. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-41924-9_27](https://doi.org/10.1007/978-3-642-41924-9_27)

24. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.: Towards ontological foundations for conceptual modeling: the unified foundational ontology (UFO) story. *Appl. Ontol.* **10**(3–4), 259–271 (2015)
25. Bergenholtz, H., Tarp, S. (eds.): *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries.* John Benjamins Publishing Company, Amsterdam (1995)
26. Singh, R.A.: *An introduction to lexicography.* Central Institute of Indian Languages (1982)
27. Fellbaum, C. (ed.): *WordNet: An Electronic Lexical Database.* MIT Press, Cambridge (1998)
28. Lenat, D.B.: CYC: a large-scale investment in knowledge infrastructure. *Commun. ACM* **38** (11), 32–38 (1995)
29. Navigli, R., Ponzetto, S.P.: BabelNet: the automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.* **193**, 217–250 (2012)
30. Global Wordnet Association. <http://globalwordnet.org/wordnets-in-the-world/>
31. Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. *IEEE Trans. Knowl. Data Eng.* **25**(1), 158–176 (2013)
32. Lindland, O.I., Sindre, G., Sølvsberg, A.: Understanding quality in conceptual modeling. *IEEE Softw.* **11**, 42–49 (1994)
33. Gómez-Pérez, A.: Evaluation of ontologies. *Int. J. Intell. Syst.* **16**(3), 391–409 (2001)
34. Noy, N.F.: Ontology mapping. In: *Handbook on Ontologies*, pp. 573–590 (2009)
35. Choi, N., Han, H., Song, I.: A survey on ontology mapping. *SIGMOD Rec.* **35**, 34–41 (2006)
36. Arnold, P., Rahm, E.: Enriching ontology mappings with semantic relations. *Data Knowl. Eng.* **93**, 1–18 (2014)
37. SNOMED International. *SNOMED CT Starter Guide* (2017). <http://snomed.org/sg>
38. Lieber, R., Stekauer, P. (eds.): *The Oxford Handbook of Compounding.* Oxford University Press, Oxford (2011)
39. Murphy, G.L.: *The Big Book of Concepts.* MIT Press, Cambridge (2004)
40. Wikipedia. https://en.wikipedia.org/wiki/English_compound
41. Nakov, P., Hearst, M.A.: Semantic interpretation of noun compounds using verbal and other paraphrases. *TSLP* **10**(3), 13:1–13:51 (2013)
42. Microsoft. <https://concept.research.microsoft.com/>
43. Cabot, J., Mazón, J.-N., Pardillo, J., Trujillo, J.: Specifying aggregation functions in multidimensional models with OCL. In: *ER 2010*, pp. 419–432 (2010)
44. Madnick, S.E.: Are we moving toward an information superhighway or a tower of babel? The challenge of large-scale semantic heterogeneity. In: *ICDE 1996*, pp. 2–8 (1996)
45. Noy, N.F.: Semantic integration: a survey of ontology-based approaches. *SIGMOD Rec.* **33** (4), 65–70 (2004)
46. Berners-Lee, T., Kagal, L.: The fractal nature of the semantic web. *AI Mag.* **29**(3), 29–34 (2008)
47. Herman, I.: State of the Semantic Web. Norway (2007). <https://www.w3.org/2007/Talks/0424-Stavanger-IH/Slides.pdf>
48. Mockus, A., Fielding, R.T., Herbsleb, J.D.: Two case studies of open source software development: apache and Mozilla. *ACM Trans. Softw. Eng. Methodol.* **11**(3), 309–346 (2002)
49. Guha, R.V., Brickley, D., Macbeth, S.: Schema.org: evolution of structured data on the web. *Comm. ACM* **59**(2), 44–51 (2016)
50. Burton-Jones, A., Wand, Y., Weber, R.: Guidelines for empirical evaluations of conceptual modeling grammars. *J. AIS* **10**, 1 (2009)
51. Conesa, J., Storey, V.C., Sugumaran, V.: Usability of upper level ontologies: the case of ResearchCyc. *Data Knowl. Eng.* **69**(4), 343–356 (2010)

52. Kashyap, V., Bussler, C., Moran, M.: The Semantic Web. Semantics for Data and Services on the Web. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-76452-6](https://doi.org/10.1007/978-3-540-76452-6)
53. Mika, P.: On Schema.org and why it matters for the web. *IEEE Internet Comput.* **19**(4), 52–55 (2015)
54. Ding, L., Kolari, P., Ding, Z., Avancha, S., Finin, T., Anupam Joshi, A.: Using ontologies in the semantic web: a survey. In: Sharman, R., Kishore, R., Ramesh, R. (eds.) *Ontologies*. Springer, Boston (2007). doi:[10.1007/978-0-387-37022-4_4](https://doi.org/10.1007/978-0-387-37022-4_4)