Metrol. Meas. Syst., Vol. XIX (2012), No. 2, pp. 343-354.

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

MEASURES FOR EVALUATION OF STRUCTURE AND SEMANTICS OF ONTOLOGIES

Wojciech Waloszek

Gdansk University of Technology, Faculty of Electronics, Telecommunications, and Informatics, Department of Software Engineering, Narutowicza 11/12, 80-233 Gdansk. Poland (✉ wowal@eti.pg.gda.pl, +48 58 347 2475)

Abstract

Recently, the topic of ontologies has growing attention from the IT community. Various processes of ontology creation, integration, and deployment have been proposed. As a consequence there appeared an urgent need for evaluating the resulting ontologies in a quantitative way. A number of metrics has been defined along with different approaches to measuring the properties of ontologies. In the first part of this paper we review the state of the art in this domain. Special attention is devoted to discussing differences between syntactic measures (referring to various properties of graphs that represent ontologies) and semantic measures (reflecting the properties of the space of ontology models). In the second part we propose an alternative approach to quantification of semantics of an ontology. The original proposal presented here exploits specific methods of representing the space of semantic models used for optimization of reasoning. We argue that this approach enables us to capture different kinds of relations among ontology terms and offers possibilities of devising new useful measures.

Keywords: ontology, description logic, knowledge cartography.

© 2012 Polish Academy of Sciences. All rights reserved

1. Introduction

With the development of the Internet and the dawn of the Semantic Web [1] the interest in ontologies skyrocketed. Scientists and practitioners continued to look for a method of describing knowledge distributed among many web sites in order to make such knowledge easier to retrieve and integrate. Even now, with many concerns and doubts about the future of the Semantic Web [2], the interest in ontologies seems to blossom.

The notion of an ontology is deeply rooted in philosophy. The word itself is of Greek origin and is a composition of two terms: *ontos*: being, something that exists, and *logos*: science or description. In philosophy, ontology is most frequently used in one of the two meanings: (1) a theory describing the construction of the world, or (2) a theory describing the properties of basic beings of which the world is constructed.

Discussion in this paper refers to a different understanding of an ontology: a technical ontology which is used in computer science in the field of artificial intelligence and knowledge management. A first attempt to define such an ontology is attributed to Thomas Gruber. Gruber, in his paper [3], formulated a short definition: "An *ontology* as an explicit specification of a conceptualization". This definition was widely adapted, sometimes enriched by more details, most notably by a somewhat superfluous requirement that the specification should be formal (Gruber treated all specifications as formal) and that a conceptualization should be shared or at least potentially be able to be shared.

Ontologies are then formal descriptions of specific domains of interest and from the very beginning they were destined to be combined into larger descriptions and reused by the community. At first many different languages, though almost all based on first-order logics,

Article history: received on Jan. 31, 2012; accepted on Apr. 23, 2012; available online on May 18, 2012.

have been used to specify the ontology contents: examples are: KIF, DAML, OIL, or even Datalog. With the emergence of the Semantic Web, standards proposed by W3C (World Wide Web Consortium) started to dominate. OWL [4], Web Ontology Language, became a frequent language of choice for ontology creation, also for the reason that the language, being based on other Internet standards, explicitly addresses many minor but repeatedly bothersome technicalities, like e.g. use of namespaces.

With the Semantic Web gaining momentum, the number of ontologies publicly available in the Web has grown enormously. In 2008 Watson search engine [5] indexed over 100,000 ontologies. With such a number and with growing popularity, the issue of evaluating ontologies in many dimensions quickly became crucial.

In this paper we briefly review the current state of research on ontology evaluation (Section 3). We introduce to the reader prominent approaches to ontology evaluation, most importantly OntoClean and O^2 framework. At the basis of both of the methods lies the assessment of ontology contents: primarily relationships between terms.

Specificity of ontologies results in the fact that the relationships may be perceived from two perspectives: syntactic (focusing on axioms included in the ontology) and semantic (which also consider conclusions coming from explicitly defined axioms). This important distinction has been made by Vrandečić and Sure in [6], where they discuss possible misconceptions coming from assuming only a syntactic perspective. Vrandečić and Sure also propose a series of transformations that can be applied to an ontology and consists in augmenting the set of axioms by new sentences explicitly expressing the most important conclusions.

While we very strongly agree with [6] that there is a clear need to introduce semantic structural measures, we propose a different approach to this problem. Apart from syntax (sentential) and semantic representation of an ontology, we distinguish a third kind: an *intermediate representation* of model space. We base this kind of representation on *Knowledge Cartography*—a method of indexing conclusions originally used to describe a wide range of entailments resulting from axioms expressed in the ontology. In the next part of this paper (Sections 4 and 5) we argue that this approach can lead to capturing a broader range of ontological relations (e.g. disjointness), however at the cost of the necessity of redefinition of some of existing measures.

In the discussion presented in the paper, we illustrate the main ideas using Description Logics (DLs; [7]), which constitute a formalism underlying the OWL 2 language. DLs are also often considered as a good choice for presentation of logic-related ideas as they have been designed to be human-readable [7]. We briefly review the preliminaries of DL in the following section.

2. Preliminaries

Description Logics is a family of formalisms for describing knowledge. All the formalisms share similar characteristics: they are (almost always) subsets of first-order logics (in the terms of expressiveness), and they exploit a simplified approach to defining predicates.

The domain of interest (a universe) is described in DL ontologies with use of three kinds of terms: *individuals* (denoting single objects from the domain), *concepts* (unary predicates, denoting categories of objects), and *roles* (binary predicates, denoting relationships between pairs of objects). Concepts and roles can be *atomic names* like *Student*, *Lecturer*, *teaches*, *likes*, *dislikes* (concept names are, by convention, written with first letter capitalized, while roles are written with lower-case letters only). From atomic names one can also build *complex* terms, with the use of *constructors* (operators), like e.g. $Student \sqcap Lecturer$ denoting objects that are students and lecturers at the same time.

Each description logic (a single specific formalism from the DL family) establishes a set of valid constructors. The precise meaning of each constructor is defined *semantically*. DL relies on Tarski-style semantics: names can be interpreted, and each *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$

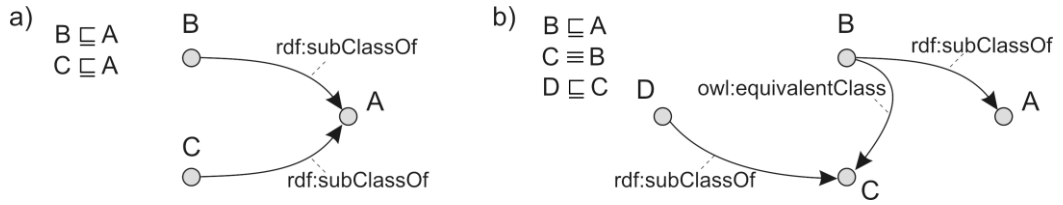

Fig. 1. Contents of two simple ontologies in sentential form, and in the form of an ontology graph.

consists of a non-empty interpretation domain $\Delta^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$ which assigns each individual an element of $\Delta^{\mathcal{I}}$, each concept a subset of $\Delta^{\mathcal{I}}$, and each role a subset of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. Such an interpretation is called *base interpretation* as it interprets only base terms. A base interpretation can be extended, so that it interprets also complex terms in the way defined by a specific logic. The extension of a certain base interpretation is explicit and unambiguous: Table 1 shows the interpretations of complex concepts used in $\mathcal{ALCQ}$ description logic.

Ontologies in DL are built from sentences. The four kinds of sentences are present in every description logic: *inclusion axioms* $C \sqsubseteq D$, *equivalence axioms* $C \equiv D$, *unary assertions* $C(a)$, and *binary assertions* $R(a, b)$ (where C and D denote, possibly complex, concepts, a and b are individuals, and R is any role). We say that an interpretation $\mathcal{I}$ *satisfies* a sentence α (denoted $\mathcal{I} \vDash \alpha$) when: $\mathcal{I} \vDash C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, $\mathcal{I} \vDash C \equiv D$ iff $C^{\mathcal{I}} = D^{\mathcal{I}}$, $\mathcal{I} \vDash C(a)$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$, $\mathcal{I} \vDash R(a, b)$ iff $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$.

The interpretations that satisfy all the sentences in an ontology are said to be its *models*. By specifying each sentence, a creator of the ontology in fact constrains its model space. The question how accurately the space captures the intention of the creator is actually an origin of important ontology measures (we elaborate on this topic in Section 3.2).

Description Logics ontologies are usually encoded in OWL 2. Since OWL 2 can be represented in RDF ([8]), OWL 2 ontology may be treated as a graph with labelled nodes representing terms, and directed labelled edges representing relationships between the terms. Fig. 1 presents two sets of sentences expressed with the use of such graphs.

Table 1. Complex concepts in $\mathcal{ALCQ}$ (C, D denotes any concept, R any role, n is a natural number).

Concept	Name	Interpretation
$\top$	*universal concept*	$\Delta^{\mathcal{I}}$
$\bot$	*empty concept*	$\varnothing$
$C \sqcap D$	*intersection*	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
$C \sqcup D$	*union*	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
$\neg C$	*complement*	$\Delta^{\mathcal{I}} - C^{\mathcal{I}}$
$\exists R.C$	*existential restriction*	$\{e: \exists f: (e, f) \in R^{\mathcal{I}}\}$
$\forall R.C$	*universal restriction*	$\{e: \forall f: (e, f) \in R^{\mathcal{I}} \rightarrow f \in C^{\mathcal{I}}\}$
$\geqslant n\, R.C$	*min. cardinality restriction*	$\{e: \lvert\{f: (e, f) \in R^{\mathcal{I}}\}\rvert \geqslant n\}$
$\leqslant n\, R.C$	*max. cardinality restriction*	$\{e: \lvert\{f: (e, f) \in R^{\mathcal{I}}\}\rvert \leqslant n\}$

3. Ontology evaluation

Gómez-Pérez in his work [9] distinguishes between two ontology evaluation tasks: *verification* and *validation*. Ontology verification consists in answering the question whether

it is properly built in accordance with requirements. Ontology validation in turn refers to the question whether the ontology captures the specificity of the chosen domain of interest.

Both notions refer to assessing ontology contents by themselves or with a limited context of expertise on the domain of interest. In contrast to many other fields, like software engineering, relatively less stress is put on assessment from a perspective of an individual user inclined to use an ontology for a specific application or task (though there exist notable works in the field, especially covering ontology search and selection, like the already mentioned Watson search engine [5] or Cupboard system [10]). One of the reasons behind it is that ontologies are intended to represent a *shared* conceptualization and their success lies in potential to be used in various settings. Another reason, pointed out in [11], is the commonly assumed point of view that formal assessment of fitness to the needs of users should be founded on preceding assessment of an ontology by itself.

In this section we present two prominent methods for ontology verification and validation: OntoClean [12] (Sec. 3.1), and O^2 framework (Sec. 3.2). Though other approaches exist (like the one described by Gómez-Pérez in [9]), the two chosen methods proved themselves to be viable, seminal for other research, and may be considered representative for the ontology evaluation state-of-the-art.

3.1. OntoClean

OntoClean [12] is one of the most widely known approaches to ontology evaluation. Originally proposed by Guarino and Welty in [13], OntoClean focuses primarily on assessment of the hierarchy of concepts (the taxonomy). Within the approach, each concept is assigned *metaproperties*, which are properties of meanings of specific terms.

OntoClean defines a set of basic metaproperties, and its first stage consists in tagging concepts with special tags indicating whether, and to what extent, a specific concept carries a specific property (an exemplary tag may look like +R+U–D+I). After the tagging, the taxonomy is verified against a set of rules. The fact that some of the rules are broken may indicate a serious error within the ontology.

An example of a metaproperty is *rigidity* (marked with R), which refers to a notion of *essence*. A concept is *essential* for an individual if the individual has to be its instance in all possible worlds, at every moment in its lifespan. If a concept is essential for all of its instance, we say it is *rigid* (and mark it with +R); if it is essential for some of its instances, it is *non-rigid* (–R); finally, if it is essential for none, we call it *anti-rigid* (~R). For instance, the concept *Person* is rigid as one cannot start or cease being a person, while the concept *Lecturer* is *anti-rigid* as no lecturer was always a lecturer, and a lecturer can always stop giving lectures.

An example of a rule which is checked during the second stage of the method is: *no rigid concept C* (+R) *can be subsumed by an anti-rigid concept D* (~R). The rationale for this rule is quite simple: *D* cannot be anti-rigid, as for at least some of its instances, namely instances of *C*, it is essential to be instances of *C* and thus *D*. A classical example of breaking this rule would be inclusion of the axiom *Person* $\sqsubseteq$ *Lecturer* in the ontology.

OntoClean gained much popularity both as a method of evaluating existing ontologies and as a method for supporting the ontology creation process [14]. Due to necessity of assigning metaproperties, OntoClean cannot be performed fully automatically and requires participation of a human evaluator. However, there are successful attempts to support this process and make it semi-automated (like AEON [15]).

3.2. O^2 framework

The O^2 approach [16] defines a framework whose aim is to capture possibly the broadest range of ontology evaluation aspects. A key element of the framework is a metaontology O^2. A metaontology is an ontology which describes other ontologies (one of the first successful use of such description was *Ontology Metadata Vocabulary* [17]; recognition of metaontologies has been growing ever since).

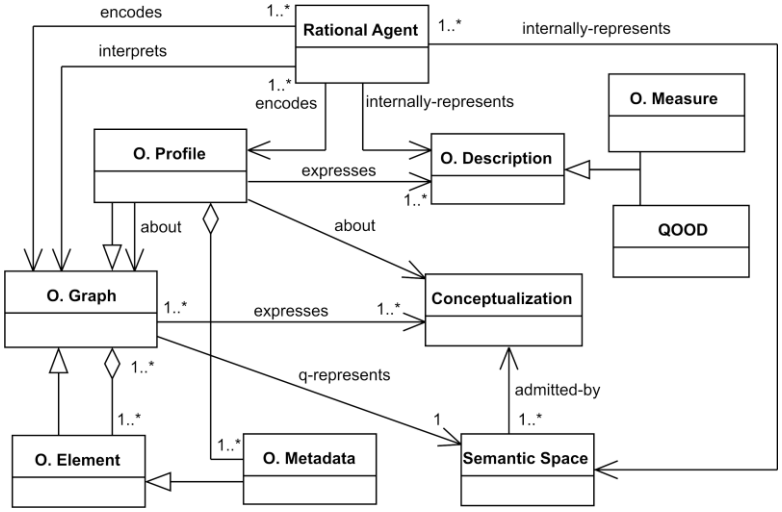

Fig. 2. The main concepts of O^2 metaontology (basing on [16]).

The main terms defined in O^2 metaontology are depicted in Fig. 2. *Rational agent* is any entity able to process ontological information (be it an ontology creator or a software application). *Ontology graph* is an encoded set of sentences forming the contents of the ontology (cf. Fig. 1). The graph contains various *ontology elements*, which are subgraphs encoding definitions and sentences, and also additional information called *metadata*. In accordance with the definition of ontology, the ontology graph expresses *conceptualization*, and establishes (orig. represents) some space of models (*semantic space*).

Any agent can provide an *ontology profile*. A profile is basically a set of metadata *descriptions* attached to a given ontology. From the point of view of the framework the most important descriptions are quality-related, among them we can distinguish *measures* and *QOODs* (*Quality-Oriented Ontology Descriptions*).

O^2 framework distinguishes three types of measures: *structural*, *functional* and *usability-profiling*. The first category embraces measures applied to an ontology graph, treated as a mathematical or topological structure. Among measured parameters are graph *depth* (number of levels), *breadth* (cardinality of paths), *tangledness* (related to the number of nodes participating in multiple hierarchies), and *fan-outness* (related to "dispersion" of nodes). All the measures are adapted versions of their counterparts in the graph theory.

Functional measures are related to the issue whether the ontology correctly captures the intended conceptualization of the chosen domain of interest (and are therefore very strongly connected with ontology validation). The authors of the framework propose to rely here on the approach proposed by Guarino in [18].

Guarino considered two parameters, known from the information processing field, namely *precision* and *coverage* (or *recall*), and adapted them to evaluation of the relationship between ontology and conceptualization. He assumed that the conceptualization determines the space of intended models (i.e. models that are indeed in accordance with intentions of the author). Due to necessary simplifications and deficiencies of the used language the space of ontology

models differs from the former one (see Fig. 3). Using the symbols from Fig. 3, and assuming a finite and fixed domain, we can define coverage as $|I_C \cap I_O| / |I_C|$ (how many intended models are covered) and precision as $|I_C \cap I_O| / |I_O|$ (how many of ontology models are indeed intended). In fact, Fig. 3 presents a very desirable situation with maximal coverage and (relatively) high precision.

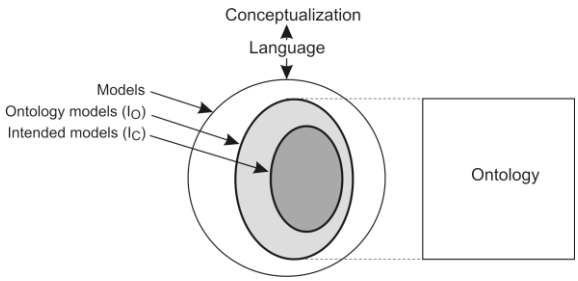

Fig. 3. The relationship between an ontology and a conceptualization (basing on [18]).

The definitions of precision and coverage from the previous paragraph should be treated as purely illustrative. They are impossible to be used in practice, as in the vast majority of cases the domain being described is (at least potentially) infinite. The practical methods utilized for estimating functional measures rely on intellectual judgement and expertise [16].

The last category of measures, *usability-profiling*, focuses on assessment of available ontology profiles, i.e. provided descriptions and annotations. The annotations are divided into several groups, like recognition annotations (including provenance and versioning information). The examples of related measures are *presence* and *amount* of annotations from a specific group.

Apart from measures, the second very important quality-related annotations in O^2 are QOODs. A QOOD is a formalized description of conformance of the ontology to a specific requirement, expressed with use of specific measures. A requirement can be as specific as the ability to determine instances of a certain class, or as general as a statement that the ontology should be flexible.

The O^2 framework provides descriptions of several elementary (most general) QOODs, called principles. The examples of distinguished principles are: *cognitive ergonomics* (if an ontology can be easily understood), *transparency* (explicitness of organization), *computational efficiency*, *flexibility*. Within each principle (QOOD) there are defined value spaces for chosen measures, meaning that a value of the measure within the space supports conformance to the principle (for instance, low values of depth and presence of annotations support cognitive ergonomics, while low tangledness supports computational efficiency).

The O^2 framework has been very influential within the community involved in ontology evaluation. Many newly created methods and approaches (like e.g. [11]) treat O^2 as a basis or at least a reference point. Also within the further part of the paper we will follow the assumptions and the terminology of the O^2 framework.

4. The problem: semantic and syntactic structural measures

Structural measures are very important tools in ontology evaluation, and, according to [19], most widely explored. [19] mentions four main reasons of this state of affairs. Specifically, structural measures can be calculated effectively from the graph, they yield simple numbers, their results can be automatically checked against constraints, and they can be easily visualized and reported.

Unfortunately, structural measures, when applied directly to an ontology graph, may produce results that are misleading. One of the reasons is that the same knowledge (understood as the same model space) may be expressed with different ontology graphs.

Consider a measure for depth from O^2 framework, called there M3. Adapted to RDF, it can be defined as the length of the longest path through nodes representing atomic concepts along directed edges with label `rdf:subClassOf`. Literal use of such a definition would result in obtaining the depth of 2 for the graph in Fig. 1a (which is correct), but also in obtaining the same depth for the graph in Fig. 1b, which seems incorrect and counterintuitive.

The reason behind the results stems from focusing on sentences explicitly expressed in the ontology graph and neglecting conclusions. This effect has been originally described in [6] by Vrandečić and Sure. We briefly recall their motivating example in a slightly modified form:

$$ProudFather \equiv \geq 1\ hasChild.\top \tag{1}$$

$$HappyFather \equiv \geq 2\ hasChild.\top \tag{2}$$

$$VeryHappyFather \equiv \geq 3\ hasChild.\top \tag{3}$$

It is obvious that every *HappyFather* is also a *ProudFather*, and every *VeryHappyFather* is a *HappyFather*. But naturally, applying the measure M3 to the ontology would result in obtaining the value of 1 for the depth.

The course of actions that may be taken to circumvent this problem is to explicitly express relevant conclusions in the form of additional sentences included in the ontology. This is the essence of the proposal of *normalization* described in [6]. The normalization consists in giving atomic names to all relevant complex concepts, and in augmenting the ontology with subsumption axioms between atomic terms (for details, refer to [6]). Applying the normalization to the ontology from Fig. 1b would result in addition of two axioms: $C \sqsubseteq A$, $D \sqsubseteq B$, and, consequently, in obtaining the value of 3 for the depth. Similarly, the ontology (1)-(3) would be augmented by axioms $HappyFather \sqsubseteq ProudFather$ and $VeryHappyFather \sqsubseteq HappyFather$, which "repairs" the measure M3, as it assumes the value of 3.

The process of normalization is effective for simple ontologies and allows for using the existing measures to better capture the semantic properties (in other words, to convert them into semantic structural measures). Nevertheless, the definition of normalization focuses primarily on subsumption and neglects other kinds of semantic relationships. This can be dangerous and may lead to undesirable results, as shown in the following example.

Consider an ontology consisting of a single axiom:

$$Person \sqsubseteq \neg Animal \tag{4}$$

The axiom (4) simply states that *Person* and *Animal* concepts are disjoint (i.e. no *Person* can be an *Animal*, and no *Animal* can be a *Person*). Application of the normalization in the exact form would result in obtaining the following ontology:

$$Person \sqsubseteq X \tag{5}$$

$$X \equiv \neg Animal \tag{6}$$

The shape of the normalized ontology graph is misleading as there is no subsumption between the basic terms. Therefore the new result for measure M3 (depth of 2) is incorrect.

As illustrated in the above example, normalization is not a remedy for the problems of measuring the space of models. We argue that in order to measure semantic properties of ontologies in more detailed fashion we need to take into consideration a different kind of relations between concepts, like disjointess. While it seems possible to extend the process of normalization so that it would produce other sets of axioms capturing this kind of relation, this way of extension may easily become infeasible. For example, answering a question:

"Which atomic concepts can be expressed as a union of other atomic concepts" may require to perform non-standard reasoning and highly complicate the process of normalization.

As a solution to this problem we propose to use a different (than an RDF graph or a simple list of sentences) form of representing the space of models of an ontology. The representation we propose is called the "map of concepts" and stems from the technique of Knowledge Cartography (KC) [20]. Maps of concepts aim at seizing the largest possible number of practically useful conclusions that can be drawn from the original form of the ontology. Consequently, measuring properties of the map of concepts can respond to many important questions concerning ontology semantics. This solution has one main drawback when compared to normalization: one cannot reuse the existing measures in their exact form. However, we show in the following parts of the paper that existing measures can be relatively easily redefined in the terms of KC. By the use of maps of concepts we are able to seize more semantic properties of an ontology, which is very important for its evaluation.

5. Cartographic approach to measuring properties of ontologies

This section is devoted to presenting Knowledge Cartography (KC) and its potential for ontology evaluation. Section 5.1 presents an outline of KC, Section 5.2 shows how the existing structural measures may be adopted to the use of KC, Sections 5.3 and 5.4 discuss how specific features of KC may be useful in ontology evaluation for analysis of stability of measures and for assessing additional characteristics of ontologies.

5.1. Knowledge Cartography

The technique of Knowledge Cartography (KC) [20] has been created within the Knowledge Management Group at Gdańsk University of Technology. As it has been previously mentioned, KC focuses on creating a representation capturing the largest number of possibly useful conclusions, and therefore it is also useful for characterizing the space of semantic models.

Representations in KC base on a *decomposition* of the universe. A decomposition **D** is a set of pairwise disjoint concepts (called *regions*) which cover the whole domain of interest. A *map of concepts* $\mathcal{M}$ (the key notion of KC) for the ontology O is a pair (M, m) where M is a set containing all the concepts from **D** that are satisfiable (i.e. can have instances) in O and m is a function assigning each concept a subset of M, such that $m(C) = C$ for $C \in M$, $m(C \sqcap D) = m(C) \cap m(D)$, $m(C \sqcup D) = m(C) \cup m(D)$, and $m(\neg C) = M - m(C)$.

Although the formal notion of a map of concepts may be difficult to understand, such a map has a very intuitive graphical representation in the form similar to a Venn diagram (see Fig. 4). Decomposition determines a "resolution" of such a diagram, while the mapping m assigns each concepts its "location" (i.e. a set of regions; in Fig. 4 regions are concepts I_1 to I_4).

A usual choice of decomposition is similar to this depicted in Fig. 4. Such a decomposition embraces the intersection of selected concepts or their negations: $\mathbf{D} = \{E_1 \sqcap E_2 \sqcap \ldots \sqcap E_n\colon E_i \in \{L_i, \neg L_i\}$ for $i \in [1..n]\}$. The list of selected concepts $\{L_i\}_{i \in [1..n]}$ (called the *base*) contains, like in Fig. 4, all the atomic concepts from a given ontology O, and also all the concepts of the form $\exists R.C$, $\forall R.C$, $\geqslant n\ R.C$, $\leqslant n\ R.C$ used in O. Such a choice of a base guarantees that for a given map $\mathcal{M}$ and all the pairs of atomic concepts A, B, $m(A) = m(B) \Leftrightarrow A \equiv B$ and $m(A) \subseteq m(B) \Leftrightarrow A \sqsubseteq B$. This is very important as it means that the relationships between sets of regions assigned to atomic concepts reflect all the semantic relationships (e.g. disjointness, overlapping, equivalence) between them.

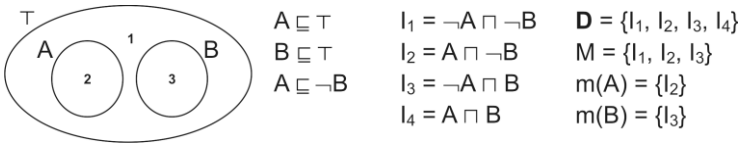

Fig. 4. A decomposition and a map of an exemplary ontology.

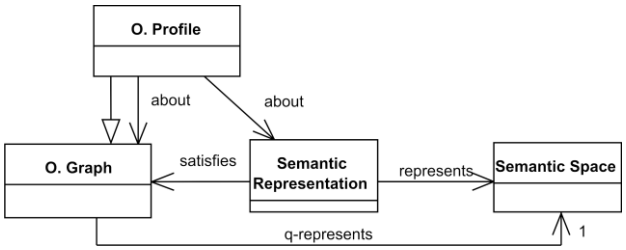

Fig. 5. The proposed extension of the O^2 framework.

5.2. Semantic structural measures for Knowledge Cartography

A systematic approach to using KC in ontology evaluation requires to situate it within one of the known frameworks. Fig. 5 depicts the proposed extension of O^2 ontology. We propose to introduce a new element of the framework: the *semantic representation*, which is an intermediate form between an ontology graph and a space of models. While the properties of a semantic space cannot be directly measured, these of the representation can, and therefore the representations may have associated profiles containing appropriate descriptions.

A map of concepts is a kind of semantic representation and as such it expresses relationships between terms in an ontology. A map might be then a subject to all the measures, especially structural ones, applied to an ontology graph. However, the structural measures need redefinition in order to be used for maps of concepts.

We show that the process of adaptation is rather simple and straightforward. The first example is the aforementioned measure M3. In its original form it measures the length of the longest path along edges labelled `rdf:subClassOf`. If we denote the set of atomic concepts from an ontology O as $A(O)$, we may change the definition to the length n of the longest sequence of concepts from $A(O)$: $A_1, A_2, \ldots A_n$, such that $i < j$, $i \in [1..n]$, $j \in [1..n] \Rightarrow m(A_i) \subset m(A_j)$.

The second measure we consider as an example is the tangledness. Originally, the measure M7 in O^2 is defined as a ratio of the number of all atomic concepts to the number of atomic concepts having at least two direct descendants. Once again we have to redefine a notion of "direct descendants". The set of direct descendants $D(A)$ of an atomic concept A can be defined as the set of atomic concepts A' such that $m(A') \subset m(A)$ and there does not exist any atomic concept B such that $m(A') \subset m(B)$ and $m(B) \subset m(A)$. Consequently, we can redefine the measure M7 in the following way:

$$t = |A(O)| \,/\, |\{A \in A(O)\colon D(A) \geqslant 2\}| \tag{7}$$

The last example concerns one of the measures for fan-outness. The measure M9 is defined as a ratio of all atomic concepts to leaf concepts. As usual, we have to redefine the auxiliary notion of the set of leaf concepts as $LEA = \{A \in A(O)\colon D(A) = \varnothing\}$ (because the leaf concepts are the concepts with no descendants). The new form of the measure M9 is presented below:

$$f = |A(O)| \,/\, |LEA| \tag{8}$$

5.3. Defining new measures in the framework of KC

In the previous subsection we have shown the approach for adapting KC to the existing measures. However, we believe that the main strength of the use of KC is the flexibility in defining new measures, which take into consideration different relationships between concepts. In this subsection we give several examples of such measures.

One of the factors often exploited in ontology evaluation is exhaustiveness of the decomposition in the concept hierarchy. While frequently used in various frameworks (e.g. [9]), it has no associated structural numerical measure (assessment bases only on presence, i.e. finding appropriate patterns in an ontology graph). The measure we propose relates the number of concepts having the exhaustive decomposition to the number of all non-leaf atomic concepts. Since the decomposition of a given atomic concept A is exhaustive if the $m(D(A)) = m(A)$, the proposed measure (*exhaustiveness*) assumes the following form:

$$e = |\{A \in (A(O) - LEA): m(D(A)) = m(A)\}| \,/\, |A(O) - LEA| \tag{9}$$

The second measure considers disjointness. Proper use of disjoint concepts is very important in ontology development (as argued in [21]), nonetheless this relationship is often neglected by ontology evaluation frameworks. As the measure for this, we propose an arithmetic average of the number of overlapping leaf atomic concepts (*overlapping*):

$$o = \Sigma_{A \in |LEA|}(|\{A' \in LEA: A' \neq A, m(A') \cap m(A) \neq \emptyset\}| \,/\, |LEA|) \tag{10}$$

Finally, we propose a measure more strongly connected with the notion of the map. In that way we answer the question of how much we know about an individual that is an instance of a leaf concept. We propose to measure this property with the average number of regions assigned to an atomic concept (*vagueness*):

$$v = \Sigma_{A \in |LEA|}(|m(A)| \,/\, |LEA|) \tag{11}$$

The measure (11) is a bit similar to the one defined for disjointness (10), because the number grows with a growing number of overlapping leaf concepts. However, the correspondence is not straightforward as the number of regions also indicate our knowledge about presence or absence of certain role instances (as the concepts like $\exists R.C$, $\forall R.C$, $\geqslant n\ R.C$, $\leqslant n\ R.C$ are also present in the map).

Fig. 6 illustrates a possible use of introduced measures. In Fig. 6a a relatively low value of exhaustiveness and high values of overlapping and vagueness indicate possible deficiencies in the ontology (it might not always be the case, but for this particular ontology we assume that it was the intention of its creator to design a proper—i.e. exhaustive and disjoint—taxonomy; with the new measures (9)-(11) such a requirement can be described with an appropriate QOOD, and its violation may be automatically detected). In Fig. 6b we can see the corrected version of the ontology: $e = 1$ indicates that the hierarchy of concepts is fully exhaustive, while $o = 0$ and $v = 1$ that all the leaf concepts are pairwise disjoint.

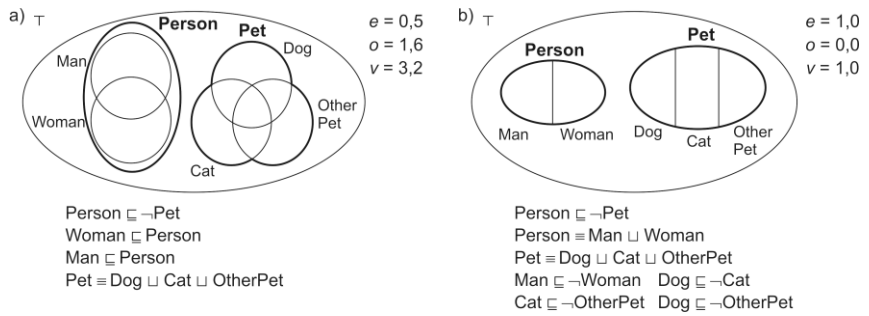

Fig. 6. Changes of a map caused by introduction of new axioms.

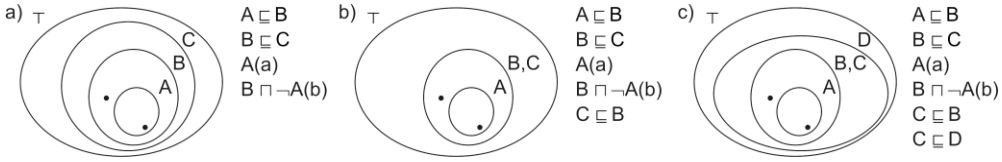

Fig. 7. Changes of a map caused by introduction of new axioms.

5.4. Assessing stability of measures within KC

The problem of stability of measures (signaled in [6]) is connected with the answer to the following question: "How much can the ontology (and its measured properties) change if the new axioms are introduced?" The issue is quite fundamental, because ontologies are designed to be reused in various settings, and reusing means also introducing new axioms.

The answer to the question of stability can be easily facilitated within the framework of KC. The map of concepts can change in two ways, when new axioms are added to a terminology:

- a region may be removed from the map (see Fig. 7b), which results in its removal from all the sets $m(C)$,
- a region (or regions) may be "split" into several regions as a result of introducing a new concept (see Fig. 7c) and consequent changes in the decomposition. This also requires addition of new regions (being the effect of the split) to appropriate sets $m(C)$.

As a result of repetitive introduction of the changes described above, values of practically all measures may change dramatically. For example, the depth (measure M3) can be raised to any value in consequence of region splitting, and reduced to 1 (or even 0, if we allow for unsatisfiable atomic concepts) as the outcome of region removal.

The situation changes, if we take individuals into account. In order for the knowledge base to remain consistent, we cannot remove from the map regions which have at least one instance (such regions in Fig. 7 contain a dot inside). This allows us to set at least one bound on the depth. The illustration of such a situation is presented in Fig. 7. The existence of individuals prevents from removing regions, so that the depth cannot be decreased below 2.

6. Summary

In the paper we focused on the topic of ontology evaluation. The problem is complicated and important not only because of the recent growth of popularity of ontologies, but also because it involves so many factors that are hard to capture. Questions like: "How good is my conceptualization?", "How broad is it?", "Can it be shared?", "Is its description accurate and precise?" are very difficult to answer, or even to be expressed in more formalized forms.

Despite the complexity of the task many attempts in this area have been undertaken and they resulted in the development of notable frameworks addressing the majority of aforementioned questions. However, as pointed out in [6], the methods very often focus on syntactic aspects of ontology structure, neglecting the properties of the space of semantic models.

We propose to use an ontology representation for which there exists a guarantee of capturing basic relationships between terms. We extend the O^2 framework with the notion of such semantic representation and suggest using maps of concepts from Knowledge Cartography in this role. This allows us to cope with ambiguity of definitions of structural measures, to propose new measures for other kinds of semantic relationships, and, finally, to obtain a helpful tool in assessment of stability of ontology measures.

Future development would consist in inventing new types of semantic representations, especially directed towards modular ontologies. Modularity of ontologies introduces new

levels of difficulty in measuring their properties, yet it seems to be unavoidable in the rapidly evolving environment of a more and more semantic World Wide Web.

References

[1] Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43.

[2] Hendler, J. (2011). Why the Semantic Web will Never Work. *http://www.slideshare.net/jahendler/why-the-semantic-web-will-never-work*.

[3] Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220.

[4] Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Ulrike, S. (2008). OWL 2: The next step for OWL. In *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4), 309-322.

[5] d'Aquin, M., Baldassarre, C., Gridinoc, L., Angeletou, S., Sabou, M., Motta, E. (2007). Characterizing knowledge on the semantic web with Watson. In *Proceedings of the 5th Workshop on Evaluation of Ontologies and Ontology-based tools*, 1-10.

[6] Vrandečić, D., Sure, Y. (2007). How to Design Better Ontology Metrics. In The Semantic Web: Research and Applications, Franconi E., Kifer M., May W. (eds.), LNCS 4519, Springer-Verlag, 311-325.

[7] Baader, F.A., McGuiness, D.L., Nardi, D., Patel-Schneider, P.F. (2003). *The Description Logic Handbook: Theory, implementation, and applications*, Cambridge University Press.

[8] Beckett, D. (2004). RDF/XML syntax specification (revised). W3C Recommendation, *http://www.w3.org/ TR/REC-rdf-syntax/*.

[9] Gómez-Pérez, A. (2004). Ontology Evaluation. In *Handbook on Ontologies*, Staab S., Studer R. (eds.), Springer-Verlag, 251-273.

[10] d'Aquin, M., Holger, L. (2009). Cupboard – a place to expose your ontologies to applications and the community. In *The Semantic Web: Research and Applications, 6th European Semantic Web Conference*, LNCS 5554, Springer-Verlag, 913-918.

[11] Vrandečić, D. (2010). Ontology Evaluation. *PhD dissertation*, Karlsruher Institut für Technologie.

[12] Guarino, N., Welty, C.A. (2004). An Overview of OntoClean. In *Handbook on Ontologies*, Staab S., Studer R. (eds.), Springer-Verlag, 151-171.

[13] Guarino, N., Welty, C.A. (2000). Ontological Analysis of Taxonomic Relationships. In *Proceedings of ER-2000: The 19th International Conference on Conceptual Modeling*. Springer-Verlag, LNCS 1920, 210-224.

[14] Gómez-Pérez, A. (2004). Evaluating Ontology Evaluation. IEEE *Intelligent Systems*, 19(4), 74-76.

[15] Völker, J., Vrandečić, D., Hotho, A., Sure, Y. (2008). AEON – an approach to the automatic evaluation of ontologies. *Applied Ontology*, 3(1-2), 41-62.

[16] Gangemi, A., Catenaccia, C., Ciaramita, M., Lehmann, J. (2006). Modelling ontology evaluation and validation. In *Proceedings of the 3rd European Semantic Web Conference*, Sure Y., Domingue J. (eds.), LNCS 4011, Springer-Verlag.

[17] Hartmann, J., Sure, Y., Haase, P., Palma, R., Suárez-Figueroa, M. (2005). OMV – Ontology Metadata Vocabulary. In *Ontology Patterns for the Semantic Web Workshop*, Welty Ch. (ed.), Galway, Ireland.

[18] Guarino, N. (2004). Toward a Formal Evaluation of Ontology Quality. *IEEE Intelligent Systems*, 19(4), 78-79.

[19] Vrandečić, D. (2009). Ontology Evaluation. *Handbook on Ontologies*, second edition, Staab S., Studer R. (eds.), Springer-Verlag, 293-313.

[20] Goczyła, K., Grabowska, T., Waloszek, W., Zawadzki, M. (2006). The Knowledge Cartography A New Approach to Reasoning over Description Logics Ontologies. In: *SOFSEM 2006: Theory and Practice of Computer Science*, LNCS 3831, Springer-Verlag, 293-302.

[21] Völker, J., Vrandečić, D., Sure, Y., Hotho, A. Learning Disjointness. In *The Semantic Web: Research and Applications*, Franconi E., Kifer M., May W. (Eds.), LNCS 4519, Springer-Verlag, 175-189.