Towards a Semiotic Metrics Suite for Product Ontology Evaluation

Joerg Leukel, University of Hohenheim, Germany
Vijayan Sugumaran, Oakland University, USA

Abstract
In recent years, product ontology has been proposed for solving integration problems in product-related information systems such as e-commerce and supply chain management applications. A product ontology provides consensual definitions of concepts and inter-relationships being relevant in a product domain of interest. Adopting such an ontology requires means for assessing their suitability and selecting the "right" product ontology. In this article, the authors (1) propose a metrics suite for product ontology evaluation based on semiotic theory, and (2) demonstrate the feasibility and usefulness of the metrics suite using a supply chain model. The contribution of our research is the comprehensive metrics suite that takes into account the various quality dimensions of product ontology.

Keywords: Electronic Commerce, Metrics, Ontologies, Ontology Quality, Supply Chain Management

INTRODUCTION

Product-related information is of paramount importance in many interorganizational applications, since it concerns goods and services being procured, manufactured and sold to customers. Due to the involvement of multiple organizations, there is a need for integrating product-related information, e.g., by standardization or mediation. In the past years, *product ontology* has attracted both industry and academia because of its potential contribution to solving integration problems (Shim & Shim 2006). A product ontology provides, at least to some extent, consensual definitions of concepts and inter-relationships between these concepts in a product domain of interest. Most product ontologies define a hierarchy of product classes and respective properties for describing product instances. Such ontologies may support finding and comparing products being offered by multiple suppliers and described in distributed data sources, or allow for benchmarking the procurement activities of organizational units (Doring et al., 2006). Ontology users are required to annotate their product instance data accordingly.

Product ontologies have already emerged in diverse industries and for various tasks (Park et al., 2008). However, assessing the quality and suitability of a given product ontology, i.e., to what degree it actually meets user requirements, remains a critical question for potential ontology adopters. This question is the focus of *ontology evaluation*, which aims at providing metrics reflecting the ontology's quality and suitability. There is great difficulty in determining what elements of quality to evaluate. In other words, what factors should be considered in evaluating product ontology quality? Current research yields a number of approaches, metrics, and tools for automatically evaluating ontologies (Garcia-Castro et al., 2007; Hartmann, 2005). However, most of this research originates from the Semantic Web arena, and therefore relies mainly on the expressiveness of ontology languages such as DAML (DARPA Agent Markup Language) and OWL (Ontology Web Language); hence their scope is constrained by these languages and does not take the specific setting of product ontology into account.

Very often, an ontology is regarded as an artifact used by a community as a common vocabulary without considering the organizational properties of the respective community and thus the inter-relations within the community (Zhdanova et al., 2007). For example, a community that often uses product ontologies is made of entities belonging to a supply chain. A supply chain is a system of entities participating in producing, transforming, and distributing goods and services from supply to demand. A single product ontology is thus used within supply chains and determining its quality and suitability has to consider the supply chain characteristics, e.g., by distinguishing different roles such as manufacturer and distributor. A major trend affecting supply chains is individualization, caused by customers demanding individualized products, which are tailored to their specific needs (e.g., custom-made products) (Coates, 1995) (Kirn, 2008). For instance, enabling customers to order custom-made shoes via an e-commerce application does not only concern the e-commerce firm but also the stakeholders in the respective supply chain (e.g., manufacturer and its suppliers). Here, a product ontology may help provide a common terminology and means of describing products along the entire supply chain.

In the context of supply chain and individualization, a product ontology should emphasize the importance of quality metrics that allow the assessment of product complexity in terms of richness of product description and product structure, and how the final product is composed of individual parts. Current evaluation metrics do not take these factors into account: Domain-independent metrics are not able to exploit the domain characteristics (e.g., Yao et al. 2005), whereas domain-specific evaluation metrics regard products as single and atomic items without considering existing inter-relations that arise due to supply chain structures and customer requirements (e.g., Hepp et al. 2007). To overcome this limitation, we address product ontology evaluation on a broader scale by taking a semiotic perspective. Semiotics studies the properties of signs; for our purposes, it can provide a theoretical basis for distinguishing generic categories of quality. We define evaluation metrics based on Stamper's et al. (2000) semiotic framework and adopt the domain-independent semiotic metrics suite proposed by Burton-Jones' et al. (2005).

The objectives of this research are to: (1) develop a semiotic set of metrics that allow for assessing the quality of product ontologies, and (2) apply the metrics to a commonly available product ontology to demonstrate the feasibility and usefulness of the metrics suite. The contribution of this research is the comprehensive metrics suite that takes into account the various quality dimensions of product ontologies. A preliminary study of semiotic metrics for product ontology evaluation can be found in (Leukel & Sugumaran, 2007). The contribution of this research is that the current work adapts the metric suite developed by Burton-Jones et al. (2005) to the product domain in the context of supply chain management to determine which metrics are applicable and how they relate to the existing work in the product domain ontology. We map the metrics developed in both streams of research and develop a unified set of metrics for the product ontology domain.

The rest of the article is organized as follows. The next section defines the basic model of supply chain and product ontology. After that, we present our semiotic metrics suite. The subsequent section provides a preliminary validation of the metrics. In the section that follows, we review the related work. Finally, we draw some conclusions and outline avenues of future research.

BASIC MODEL OF SUPPLY CHAIN AND PRODUCT ONTOLOGY

A supply chain model is a representation of entities participating in producing, transforming, and distributing goods and services from supply to demand (Supply-Chain Council, 2007). Such a model thus concerns both products and organizations which we call actors. The inter-relations between actors are constituted by the flow of products. We define the supply chain model as a directed graph S = (A, F) using the following notions.

A: set of actors
a: element of A
F: set of flows of products with $F \subseteq A \times A$-1
f: element of F connecting two actors, $f = (a_1, a_2)$

To allow for distinguishing the specific role of each actor within a supply chain, we define five generic roles: 1) OEM, 2) N-tier supplier, 3) distributor, 4) retailer, and 5) customer. These roles represent the generic types of participants in a supply chain. A specific supply chain may include several instances of each of these roles. Typically, a supply chain is made of two major parts being separated by the original equipment manufacturer (OEM). The left hand side role of n-tier suppliers describe how the product is made from parts (thus it focuses on procurement and manufacturing) with n denoting the number of supply chain stages. For instance, by assigning the 1st tier supplier role to different actors each supplying a different part of the final product, one can represent the product structure in the supply chain model. 2nd tier suppliers provide product to the 1st tier supplier etc. The right hand side roles participate in distributing the product from the OEM across one or more stages which are: distributor and retailer. Respective actors do not apply manufacturing technology but change the product regarding location, time, and quantity (by means of logistics, e.g., bundling) until the product reaches the final customer. The sequence of roles can be regarded as the reference structure of many real-world supply chains, as shown in Figure 1. Note that this simple model contains only one actor for each role whereas in reality multiple actors exist for almost any supply chain stage.

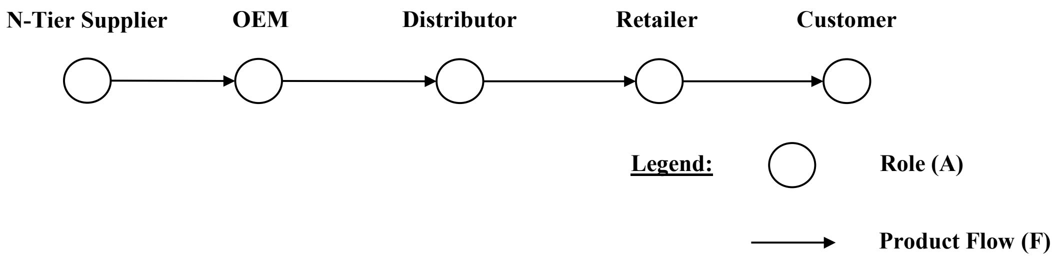

Figure 1. Roles within a supply chain

Product ontology evaluation focuses on the ontology constructs that are used in this type of ontology. Product ontology relies essentially on providing an often broad and deep hierarchy of product classes based on is-a relationships, while other relationship types play a minor role. Having reviewed current product ontologies and respective conceptual models, we define product ontology PRO as a 6-tuple PRO = (PC, RC, PP, RP, PV, RV) as defined in Table 1.

Table 1. Elements of product ontology

Element	Definition	
PC	set of product classes	A product class pc represents the product concept and often consists of, besides the class name, natural language definition and some data management information (e.g., identifiers, version etc.).
RC	set of relations between product classes	Is-a relations build a hierarchy of product classes with $RC \subseteq PC \times PC$; often a taxonomy with top-level classes $CT \subseteq PC$ for separating domains (e.g., automotive, chemical, textile etc.).
PP	set of product properties	A product property pp is a template for describing product instances; consists of property name, a natural language definition, data type, and unit of measurement.
RP	set of class-property relations	Maps properties of PP to classes PC with $RP \subseteq PC \times PP$; the semantics of such relations can range from loose recommendation to mandatory.
PV	set of property values	Values v for properties besides standard data types such as integer, float, and string. Used for expressing a narrower domain, e.g., for colors, shapes, materials etc.
RV	set of property-value relations	Maps values of PV to properties PP with $RV \subseteq PV \times PP$.

The relationship between supply chain and product ontology model is two-fold. First, since the arrows (F) connecting two roles in the supply chain model represent flows of one or more products, each such arrow (f) has to be mapped to at least one respective product class "pc" in the ontology; otherwise the ontology would not be able to cover the respective part of the supply chain model. Figure 2 shows an example how product flows can be annotated with references to product classes (pc) to describe the product structure along the supply chain stages.

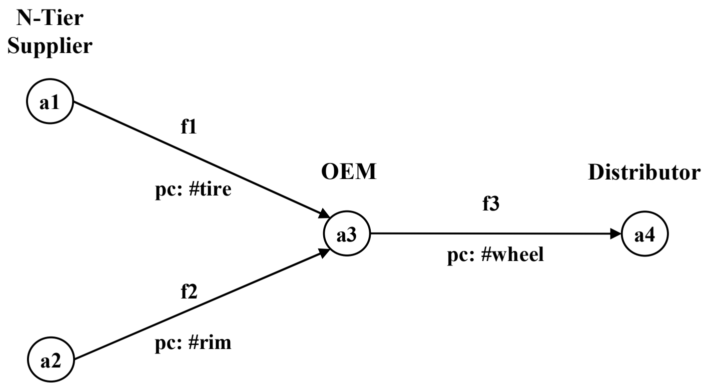

Figure 2. Example: Relationship between supply chain and product ontology

However, this type of relationship does not reflect how well the product class represents the semantics of the product flow. For instance, choosing rather abstract product classes would allow for describing many different product flows, but lose the actual semantics. Therefore, a second relationship has to be considered, which represents the richness of the product class. Thus, additional information from the product ontology is used to specify a product flow in greater detail. In the example shown in Figure 2, it is not specified whether the wheels are for cars, trucks, or both. If the product ontology contains more specific classes, one could, for instance, describe the supply chain of a specialized OEM which produces wheels for trucks only more precisely.

PROPOSED METRICS SUITE

In this section, we propose a metrics suite for product ontology evaluation based on semiotic theory (Stamper et al. 2000) and earlier work by Burton-Jones et al. (2005). First, we introduce the semiotic approach to ontology evaluation and define the basic metrics. Then we define five categories of semiotic metrics.

Semiotics

Stamper et al. (2000) present a general theoretical semiotic framework derived from linguistics that includes general elements of quality for evaluating signs. It includes pragmatic issues to develop a metrics suite that is widely applicable yet can be tailored to the needs of specific applications. They provide a 6-level semiotic framework to support the analysis of signs: a) physical, b) empirical, c) syntactic, d) semantic, e) pragmatic, and f) social. There is a strong relationship between the successive levels, i.e., each level contributes to the next level. The physical level deals with representation of signs in hardware, components, etc. and the empirical level is concerned with communication properties of signs including channel capacity, noise, and entropy. Since these levels are very implementation specific, they may not be highly relevant to quality assessment (Burton-Jones et al. 2005). The remaining levels are considered in developing the metrics suite for quality assessment.

Based on these levels, Burton-Jones et al. (2005) have developed a metric suite that consists of syntactic, semantic, pragmatic, and social qualities. As mentioned above, they do not include physical and empirical quality since they correspond to implementation aspects. With regard to product ontology, these two stages can also be omitted because signs do exist (physical in terms of explicit concepts) and can be seen (empirical in terms of a formal representation of the concepts). For our purposes, their metric suite can provide a theoretical basis for developing metrics, because (1) the general semiotic framework takes the various dimensions of the meaning of signs into account, and (2) it has been proven to be applicable and valuable for ontology evaluation.

Metrics

The metrics suite originally proposed by Burton-Jones et al. (2005) is adapted for the product ontology domain. While the metrics themselves and their constituent parts are applicable to product ontologies, additional constructs specific to the product ontology domain are identified and added to the overall metric suite. In particular, we consider the work of Hepp et al. (2007) in identifying additional constructs and create a unified set of metrics by explicating the similarities and differences between the two sets of metrics discussed in the literature.

The overall quality (Q) of an ontology is computed using a weighted function of its syntactic (S), semantic (E), pragmatic (P), and social (O) qualities (Burton-Jones et al. 2005):

$$Q = b_1 \times S + b_2 \times E + b_3 \times P + b_4 \times O$$

The weights sum to unity. In the absence of pre-specified weights, the weights are assumed to be equal. Since the numerical values of these relative scores could exceed one for any given ontology, the scores for these metrics are normalized so that the values of all metrics vary between zero and one prior to calculating the overall ontological quality. According to Burton-Jones et al. (2005), the values for a given ontology will depend on external benchmarks such as the metric's average value across all ontologies in the ontology library. With respect to product ontology, the small size of the respective ontology library may prevent using such a benchmark.

Syntactic Quality

Syntactic Quality (S) measures the quality of the product ontology according to the way it is written. It consists of *Lawfulness,* the degree to which an ontology language's rules have been complied, and *Richness,* the proportion of features in the ontology language that have been used in an ontology. Lawfulness is entirely domain-independent because one would require that any given product ontology complies with the syntax of the ontology language used. Otherwise the ontology cannot be processed. The interpretation of richness, however, must consider the requirements of the product domain with respect to expressiveness. For instance, if a highly expressive language based on description logic is used but the domain does not require much reasoning support, the richness metric would have quite a low value.

Semantic Quality

Semantic Quality (E) evaluates the meaning of terms in the product ontology. Three attributes are used in this metric: interpretability, consistency, and clarity.

Interpretability refers to the meaning of terms in the ontology (e.g., names of product classes and product properties). Meaning could be determined by checking whether the terms used can be found in the product domain of interest, for example, by searching the standards and references. *Consistency* is whether terms have a consistent meaning in the ontology. The appearance of the same term in more than one concept could indicate inconsistency. *Clarity* is whether the context of terms is clear. For example, if a product ontology claims that class "Chair" has the property "Material", an application must know that this describes furniture, and not academics.

Pragmatic Quality

Pragmatic Quality (P) refers to the product ontology's usefulness for users, irrespective of syntax or semantics. Three criteria are used for this metric, namely, accuracy, comprehensiveness, and relevance.

Accuracy is whether the claims about products an ontology makes are "true." In general, it can only be determined by domain experts, or by reasoning if the product ontology defines respective relationships between concepts. *Comprehensiveness* is a measure of the size of the ontology. Larger product ontologies are more likely to be complete representations of their domains, and provide more knowledge to the user. Size could indicate both how well the product domain is covered and to which degree of detail it is represented by explicit concepts. However, one has to be careful since sometimes, simply considering the total number of concepts and relationships may lead to false assessments. *Relevance* is whether the ontology satisfies the user's specific requirements. It could be calculated by checking against a set of explicit requirements articulated for a particular scenario.

Social Quality

Social quality (O) reflects the fact that users and product ontologies exist in communities. It consists of two attributes, namely, authority and history. The *authority* is the number of other ontologies that link to it. More authoritative product ontologies signal that the knowledge they provide is accurate or useful. The *history* is the number of times the ontology is accessed, and more precisely how often its concepts are used in actual product-related information systems. It is assumed that ontologies with longer histories are more dependable.

For the purpose of evaluating product ontologies in the context of supply chain management, comprehensiveness can be used as an indicator for both coverage of supply chains and richness of the product concepts (as defined in the basic model section). Therefore, we extend the comprehensiveness metric at this level by integrating the work of Hepp el al. (2007), which proposes an elaborate set of pragmatic metrics for product ontology evaluation. We select the relevant metrics and transform their definition into our notation of product ontology. The result is shown in Table 2, which depicts the following four aspects of product ontology quality: (1) amount of ontology content, (2) hierarchical order and balance of scope, (3) class-specific property sets, and (4) ontology growth and maintenance.

Table 2. Determination of pragmatic metrics of Hepp et al. (2007)

Aspect	Metric	Determination
Amount of content	M11: Number of classes	Number of elements in PC
	M12: Number of properties	Number of elements in PP
	M13: Number of enumerative data types	Number of elements in PV
Hierarchical order and balance of scope	M21: Number of classes per top-level class $CT_i \subseteq PC$	Number of elements in PC which are subclass of CT_i
	M22: Services ratio	(Number of elements in PC representing services) / (number of elements in PC)
	M23: Distribution properties of metric M21	Minimal value, maximal value, mean, median, first quartile, third quartile, interquartile range, standard deviation, and coefficient of variation of M21
	M24: Percentage of content in the three biggest top-level classes CT	(Number of subclasses of the three biggest classes CT) / M11
	M25: Size of the biggest top level class vs. median of M21	(Number of subclasses of the biggest class CT) / median of M21
	M26: Number of descendents per subordinate class CO_n	Number of elements in RC with CO_n superclass of CO_m
Class-specific property lists	M31: Specific property lists ratio	(Number of elements in PC with a specific property list in RP) / M11
	M32: Distribution of specific property lists per top-level class CT_i	(Number of elements in PC which are subclass of CT_i and have a specific property list in RP) / M21
	M33: Property usage in property lists	Minimal value, maximal value, mean, median, standard deviation, and coefficient of variation of {number of elements in RP with PP_i}
	M34: Semantic weight of property PP_i	1 / (number of property lists in RP including PP_i)
	M35: Semantic value of property lists	Sum of M34 for each property PP in property list of class PC_i
Growth and maintenance	M41: Number of new classes per month	((Number of elements in PC in the current version of PRO) – (number of elements in PC in the former version of PRO)) / (number of months between publication of the current and former version of PRO)
	M42: number of new classes per top-level class CT_i	M41 for CT_i

Based on our analysis of the product categorization standards work by Hepp et al. (2007), it is evident that there is some commonality between the metrics identified by Hepp et al. (2007) and Burton-Jones et al. (2005). Table 3 summarizes the integrated product ontology metrics The initial set of metrics identified by Burton-Jones et al. (2005) are still applicable for the product domain ontology, however, the pragmatic quality metric is extended with some additional metrics.

Table 3. Product Ontology Metrics [adapted from Burton-Jones et al. (2005)]

Metric	Definition
Overall Quality (Q)	$Q = b_1 \cdot S + b_2 \cdot E + b_3 \cdot P + b_4 \cdot O$ where: $b_1,..,b_4$ = weights
Syntactic Quality (S)	$S = b_{s1} \cdot SL + b_{s2} \cdot SR$ where: SL = Lawfulness SR = Richness
Semantic Quality (E)	$E = b_{e1} \cdot EI + b_{e2} \cdot EC + b_{e3} \cdot EA$ where: EI = Interpretability EC = Consistency EA = Clarity
Pragmatic Quality (P)	$P = b_{p1} \cdot PO + b_{p2} \cdot PU + b_{p3} \cdot PR$ where: PO = Comprehensiveness PU = Accuracy PR = Relevance
Social Quality (S)	$O = b_{o1} \cdot OT + b_{o2} \cdot OH$ where: OT = Authority OH = History

PRELIMINARY VALIDATION

In this section, we provide a preliminary validation of the proposed semiotic metrics suite. We apply the metrics suite to a supply chain scenario and demonstrate its relevance and usefulness. We describe the experimental design, report the results, and discuss the findings.

Validation Scenario

The purpose of the experiment is to demonstrate the feasibility and usefulness of the metrics suite in the specific setting of product ontology. In the context of individualization or mass customization in supply chains, we consider the following 5-tier supply chain model in the IT industry (Figure 3): The ultimate goal of the supply chain is to deliver custom-made desktop computers to end customers. The supply chain consists of a retailer, distributor, OEM, and 1st tier suppliers with S = (A, F) = ({a1, a2, a3, a4, a5, a6, a7}, {f1, f2, f3, f4, f5, f6}).

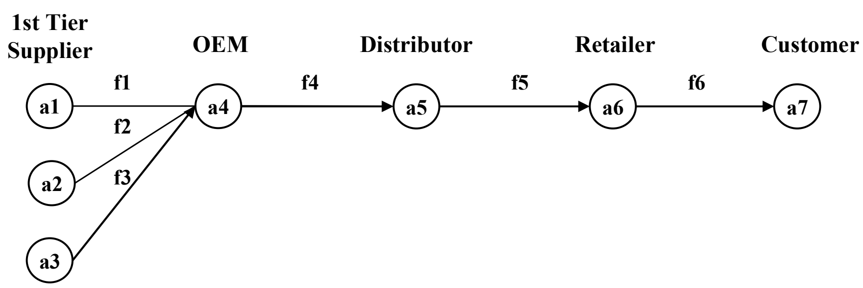

Figure 3. Supply chain model used in validation

Product flows f1, f2, and f3 represent parts, whereas f4 and f5 represent both parts and computers; f6 is computers only. This supply chain model enables different options for individualization. For instance, both the distributor and retailer can create a new bundle of computers and parts and offer it as individualized computers.

In the above supply chain scenario set up for individualization, we study the use of eCl@ss product ontology which is available in both an OWL representation called eClassOWL (Hepp, 2006) and in simple comma-separated value files. The eCl@ss is originally an international classification scheme for goods and services, and thus is of practical importance in many industries and countries (eCl@ss, 2008).

We conduct different levels of validation. First, we pre-check each metric whether it can be applied and what results can be expected. Then, we apply those metrics to the eCl@ss ontology and evaluate how well the product ontology can be used to support individualization in the supply chain scenario. In particular, we focus on the metrics that are the most relevant for our purposes, namely, *Interpretability*, *Consistency*, *Clarity*, and *Relevance*, since they refer to semantics and pragmatics. The results of our analysis are discussed in the next section.

Results
Table 4 summarizes the results of pre-checking of each metric. This assessment of eCl@ss ontology shown in Table 4 does not relate to the supply chain model yet, but provides the general findings.

Table 4. Applicability and results of metrics for eClassOWL

Metric	Applicability and Results
Lawfulness (SL)	eClassOWL is syntactically correct and has been checked formally by respective OWL tools. Its spreadsheet version contains several data errors though; see Hepp et al. (2007).
Richness (SR)	eClassOWL uses rather few OWL language features (missing, e.g., unionOf, intersectionOf, inverseOf).
Interpretability (EI)	The meaning of terms used in eClassOWL can be checked by searching for entries in a reference dictionary
Consistency (EC)	Inconsistencies can occur if is-a relationships are used falsely; here, cross-checks with reference dictionaries could help identify those.
Clarity (EA)	Interpreting terms at the lowest level of the taxonomy should consider the scope or context defined by its super classes.
Comprehen-siveness (PO)	All detailed metrics can be applied; see Hepp et al. (2007) for results.
Accuracy (PU)	Whether a claim made is true or false can only be determined by domain experts. Reasoning cannot help, because there are is-a relationships only.
Relevance (PR)	Whether eClassOWL provides the knowledge required by an application could at least partly be answered by querying the ontology based on explicit requirements.
Authority (OT)	No ontology in the public ontology repositories links to eClassOWL yet.
History (OH)	Accesses to eClassOWL can be recorded by the ontology provider only. eCl@ss is mostly used in non-public applications, thus acceptance can hardly be determined.

Next, we report on applying the individual metrics to the supply chain scenario. Specifically, we highlight our findings with respect to *Interpretability*, *Consistency*, *Clarity,* and *Relevance* metrics using a snapshot of eCl@ss. We choose a sample of 100 class names from the relevant *segment 19* which is entitled 'Information, communication and media technology'.

Lawfulness and Richness In a supply chain context, participants exchange product information. .If the participants use different descriptions, ontologies can be used to resolve differences. To automate this process, ontologies should be machine processable. The eClassOWL representation has been checked by OWL tools and it is syntactically correct, however, the spreadsheet version contains several data errors. With respect to Richness, the eClassOWL uses only a few OWL language features (missing, e.g., unionOf, intersectionOf, inverseOf).

Interpretability of the terms used in eCl@ss is checked by searching for respective entries in WordNet. WordNet returns respective entries for 97 out of 100 terms (which correctly refer to IT). The three missing terms can be explained by typos and not separating words correctly (i.e., 'computersystem', 'mainmemory', and 'harddrive'); these probably could have been retrieved by using a more capable stemmer.

This result shows quite a high interpretability of the terms used in eCl@ss; we did not search for too technical or new terms (such as acronyms of interface standards) because WordNet has not been designed for technical domains. We thereby acknowledge that the results of the interpretability metric also depend on the quality of WordNet, which is used as the reference dictionary. We assume that the share of equal terms is lower in specific product domains (e.g., chemical domain) than in broader ones.

Consistency relates to the correct usage of is-a relationships. We examined the same sample of 100 product classes and made the following observations:

- The general layout of the eCl@ss taxonomy is consistent with separating computer systems from parts. The former is classified into notebook, PDA, server, PC etc.; the latter into graphic card, cooler, drive etc. Thus, the rationale is to arrive at increasingly specialized product classes.
- This rationale is, however, violated by inserting six generic product classes for both computer systems and parts (12 inconsistent classes in total). These classes represent 'other' goods (that do not belong to another class), 'parts' (in the most generic sense), 'accessories', 'assembly', 'maintenance service' and 'repair'.

The usage of such generic product classes cannot only be observed in the sample, but in the entire eCl@ss ontology.

Clarity of the terms used in eCl@ss relates to both product classes and properties. The classes are exclusively part of a four-level taxonomy. Therefore, the context of each term in the sample is already defined by the name of *segment 19*, and on the lower levels by names of classes on levels 2 and 3. Some terms at lower levels make the context explicit by extending the class name. For instance, the class 'Mouse (computer input device)' is a sub class of 'Input device for computer'; such an extension is made only for improving interpretation by human readers and not machines. With regard to product properties, there are some generic properties (e.g., 'manufacturer name', 'supplier product number') which are assigned to every single class, whereas other properties are specific and used in few or even one class only. We can state that clarity is high for all properties in the sample due to how properties are used in eCl@ss in general.

Comprehensiveness and Accuracy metrics are typically computed based on domain expert's feedback. For the subset of eCl@ss ontology under consideration, all detailed metrics can be applied and comprehensiveness and Accuracy rank high. This is demonstrated by Hepp et al. (2007) in their analysis of eCl@ss ontology.

Relevance of the domain knowledge in eCl@ss can be studied by example queries. This metric also ties the ontology to the supply chain model. We define 5 queries each testing whether the eCl@ss provides a respective concept, thus being able to represent the required product. The queries and results are shown in Table 5.

Table 5. Example Queries for Assessing the Relevance of eCl@ss

Query	Assessment of eCl@ss
Q1: Customer a7 searches for a thin client with specific technical features (i.e., dimensions).	'thin client' is linked to the 'personal computer' class by a keyword; requirements can be described using 25 class-specific properties.
Q2: Retailer a6 asks distributors for laptop computers with no software installed in order to create a bundle by himself.	'laptop computer' class allows specifying whether software is included or not by the Boolean property 'software included'.
Q3: Retailer a6 searches for similar offerings from multiple distributors using the product number of the OEM a4.	All product classes possess, besides the supplier product number, manufacturer name and manufacturer product number, thus enabling respective search beyond the previous stage in the supply chain.
Q4: Distributor a5 searches for a replacement of TFT monitor by plasma monitor.	The classes for 'TFT monitor' and 'plasma monitor' share the same super-class 'monitor'. However, the two lists of properties are completely different, which prevents supporting the replacement decision.
Q5: OEM a4 searches for suppliers a1-a3 which can deliver modem cards for all desktop computers.	The product class for PCs contains a suitable property indicating whether a modem card is present or not; however, this property is not linked to the corresponding class 'telecommunication card' because of limitations of the eCl@ss ontology model which defines is-a relationships only.

Authority and History metrics rank very low for the eClassOWL ontology because no ontology in the public ontology repositories links to eClassOWL yet. Similarly, accesses to eClassOWL can be recorded by the ontology provider only. eCl@ss is mostly used in non-public applications, thus acceptance cannot be easily determined.

Implications

The semiotic metrics suite has several implications for users of product ontologies as well as its creators. In particular, applications such as agent-based systems can use the metrics suite to compare and select the most appropriate ontology to be used in a particular task. Ontology evaluation generally ignores specific requirements of a particular organization. Our metrics suite allows the ontology evaluation in the context of a set of products available from an organization in a supply chain. In other words, our framework helps us take into account the context in which the ontology is used in an organization. Since the metrics go beyond pragmatic dimensions such as comprehensiveness, they can help detect weak points in a given product ontology (e.g., lack of interpretability, inconsistencies, and irrelevant concepts).

Existing tools do not adequately support ontology searching and selection. Our proposed metrics suite provides a framework for evaluation of product ontologies and possibly selecting appropriate ontologies for supporting seamless e-commerce activities. The metrics can also be used by product ontology creators such as industry associations or standardization bodies as a means for analyzing the ontology and generating quantitative information about the various quality dimensions, as defined by the metrics suite.

RELATED WORK

The related work can be grouped into two major areas: product ontologies and ontology evaluation. Despite the former's importance, it is rather a specialized field which attracts interest from communities such as knowledge engineering (e.g., Fensel et al., 2001), data management (e.g., Beneventano et al., 2004), e-commerce (e.g., Leukel, 2004), and certainly product data management. While the phrase 'product ontology' is often used to stress the formal specification aspect, other widely used terms, though not equal in meaning, are 'product classification standard' or 'product classification system'. The quality and suitability of such artifacts, however, have rarely been the focus of dedicated research. Many researchers take these ontologies for granted and do not further investigate their structure and content.

To the best of our knowledge, only the work by *Hepp et al.* provides product ontology metrics as well as results of extensive quantitative evaluation (Hepp et al., 2007). While these metrics analyze product ontologies to a great extent and their rationale reflects a lot of domain expertise, they are confined only to the product ontology. Thus, they do not investigate the relationship between the ontology and, for instance, its users, the ontology language, or other ontologies. With regard to the proposed semiotic metrics suite, Hepp et al.'s metrics concern the pragmatic dimension only.

Ontology evaluation in general aims at assessing the relevance of diverse types of ontologies. There is a growing research community which develops methodologies, models, and tools for ontology evaluation, e.g., EON Workshop Series (Garcia-Castro et al., 2007). Studying the ontological quality is made difficult by a number of factors. Contrary to Information Retrieval, for instance, one cannot easily define the metrics 'precision' and 'recall', since these require a clear set of items – here concepts, inter-relations, and properties – being relevant in the respective domain of interest (Brewster et al., 2004). Ontology evaluation can be classified using attributes, for instance as described in (Hartmann, 2004) and (Brank et al., 2005). For the purpose of our work, we focus on one attribute which distinguishes functionality and structure.

The *functionality* of an ontology describes how suitable and appropriate it is for its intended usage in an information system. There are two major approaches. First, one could relate the ontology directly to requirements of the respective task. In this case, such requirements need to be elicited, formalized, and then mapped to elements of the ontology. Second, one could select a particular ontology and compare it to a reference ontology. The shortcoming of both approaches is that both requirements and the reference ontology for the domain of interest can be incomplete, wrong, lacking, and if available, subjective. This is in particular true for broad product ontologies such as UNSPSC[1], eOTD[2] and eCl@ss[3], which all aim at becoming the first reference and global standard. Thus comparing them to another reference ontology is not feasible.

The *structure* of an ontology is formed by its elements and inter-relations. A major stream of research is rooted in the Semantic Web arena and its approaches rely on ontology languages such as OWL and its predecessor DAML+OIL. By systematically checking the actual usage of language features such as classes, properties, axioms, instances etc., one can determine the structural characteristics. For instance, complexity metrics are defined in (Yang et al., 2006); they include number of concepts, relations and paths, and the mean of relations and paths per concept. Similar metrics can be found in (Huang & Diao 2006), which defines metrics for assessing how balanced a taxonomy is. However, both proposals represent only a limited subset of the entire ontology language features and respective structural aspects.

A more elaborate set of metrics describing cohesion can be found in (Yao et al., 2005). Based on graph theory, these metrics determine the degree of relatedness of concepts in an ontology. Though mathematically sound, the results of such metrics cannot easily be interpreted in terms of quality. The same is true for many structural metrics. For instance, whether a big, nested ontology is better than a smaller one depends primarily on the domain of interest. This point of criticism complements the fact that structural metrics in general rely on the expressiveness of ontology languages (thus on what can be described formally). Consequently, these metrics should be regarded as a component of a broader evaluation framework.

The review of related work points out that (1) generic ontology evaluation limits its scope by respective ontology languages and thus cannot fully exploit the quality of domain ontologies, in particular product ontologies which rather rely on size and deepness than on formal complexity, and (2) domain ontology

[1] http://www.unspsc.org
[2] http://www.eccma.org
[3] http://www.eclass-online.com

evaluation requires not only extensive domain expertise, but also an ontological foundation to arrive at both suitable and well-defined metrics. Most current research exploits the pragmatic quality dimension only without taking into account the setting of an ontology. By employing semiotics theory, the proposed metrics suite aims at overcoming these deficits.

CONCLUSIONS

This article proposed a metrics suite for product ontology evaluation based on semiotic theory and demonstrated the feasibility and usefulness of the metrics suite using a supply chain model. The contribution of our research is the comprehensive metrics suite that takes into account the various quality dimensions of product ontology.

Our approach also incorporates the specific setting of product ontologies. This setting is mainly determined by supply chains in which such ontologies are used. Thus, our work has studied supply chains based on a general supply chain model, which allows us to distinguish different roles of actors. In particular, we have addressed the individualization of supply chains which no longer makes it feasible to consider only a small snapshot of a supply chain. For instance, reducing the problem to the relationship between supplier and customer arrives at two-tier supply chain models.

However, finding a suitable product ontology for a particular industry is non-trivial because these ontologies are implemented using a variety of languages, methodologies, and platforms. Assessing their quality and selecting a particular ontology is difficult because of the heterogeneity. We have presented a semiotic set of product ontology metrics that allow for assessing the syntactic, semantic, pragmatic, and social quality, and showed how these metrics could be used for evaluating a real-world product ontology. This metric suite can be used by applications and decision makers to assess the quality of available ontologies in a particular domain. The semiotics based approach sheds light on creating various categories of quality and provides a systematic way to develop quality metrics.

Our proposed product ontology evaluation approach has some limitations. First, it does not provide explicit guidelines for determining the optimal weighting scheme of the various quality dimensions. Second, product ontologies evolve over time and the metric suite does not yet take into account this dynamic nature (Hepp, 2007). Third, it does not include any learning mechanism to update the evaluations based upon feedback from external users. Finally, empirical testing of the approach is needed to validate the relationship between an ontology’s internal attributes reflected in its metrics and its external attributes such as its usefulness for supporting an application.

Our future work involves: (1) testing our metrics suite in more realistic and comprehensive usage scenarios, (2) developing a tool for automated ontology quality assessment, and (3) applying it to ontologies in other domains.

REFERENCES

Beneventano, D., Guerra, F., Magnani, S., & Vincini, M. (2004). A web service based framework for the semantic mapping amongst product classification schemas. *Journal of Electronic Commerce Research, 5*(2), 114-127.

Brank, J., Grobelnik, M., & Mladenic, D. (2005). A survey of ontology evaluation techniques. *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005).*

Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (2004). Data driven ontology evaluation. *Proceedings of the International Conference on Language Resources and Evaluation (LREC-2004).*

Burton-Jones, A., Storey, V.C., Sugumaran, V., & Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering, 55*(1), 84-102.

Coates, J. (1995). Customization promises sharp competitive edge. *Research in Technology Management, 38*(1), 6-7.

Doring, S., Kiebling, W., Preisinger, T., & Fischer, S. (2006). Evaluation and optimization of the catalog search process of e-procurement platforms. *Electronic Commerce Research and Applications, 5*(1), 44-56.

eCl@ss (2008). *International Standard for the Classification and Description of Products and Service*. V8.0, http://www.eclass-online.com.

Fensel, D., McGuinness, D.L., Schulten, E., Ng, W.K., Lim, E.-P., & Yan, G. (2001). Ontologies and electronic commerce. *IEEE Intelligent Systems,16*(1), 8-14.

Garcia-Castro, R., Vrandecic, D., Gomez-Perez, A., Sure, Y., & Huang, Z. (2007). *Evaluation of Ontologies and Ontology-based tools.* http://km.aifb.uni-karlsruhe.de/ws/eon2007.

Hartmann, J. (2005). Methods for ontology evaluation. *Knowledge Web Deliverable D1.2.3*. http://knowledgeweb.semanticweb.org.

Hepp, M. (2007). Possible ontologies: How reality constrains the development of relevant ontologies. *IEEE Internet Computing, 11*(1), 90-96.

Hepp, M. (2006). eClassOWL 5.1. Products and Services Ontology for e-Business. *User's Guide*. http://www.heppnetz.de/eclassowl.

Hepp, M., Leukel, J., & Schmitz, V. (2007). A quantitative analysis of product categorization standards: Content, coverage, and maintenance of eCl@ss, UNSPSC, eOTD, and the RosettaNet Technical Dictionary. *Knowledge and Information Systems, 13*(1), 77-114.

Huang, N., & Diao, S. (2006). Structure-based ontology evaluation. *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE 2006)* (pp. 132-137).

Kirn, S. (2008). *Individualization Engineering.* Cuvillier: Goettingen.

Leukel, J. (2004). Standardization of product ontologies in B2B relationships – On the role of ISO 13584. *Proceedings of the 10th Americas Conference on Information Systems (AMCIS 2004)* (pp. 4084-4091).

Leukel, J., & Sugumaran, V. (2007). Evaluating product ontologies in e-commerce: A semiotic approach. *Proceedings of the 6th Workshop on e-Business (WeB 2007)* (pp. 240-249).

Park, J.M., Nam, J.H., Hu, Q.P., & Suh, H.W. (2008). Product ontology construction from engineering documents. *Proceedings of the International Conference on Smart Manufacturing Application (ICSMA 2008)* (pp. 305-310).

Shim, J., & Shim, S.S.Y. (2006). Ontology-based e-Catalog in e-commerce: Special Section. *Electronic Commerce Research and Applications, 5*(1), 1.

Stamper, R., Liu, K., Hafkamp, M., & Ades, Y. (2000). Understanding the role of signs and norms in organisations – a semiotic approach to information systems design. *Behaviour & Information Technology, 19*(1), 15-27.

Supply-Chain Council (2006). *Supply Chain Operations Reference Model (SCOR®) Version 8.0.* http://www.supply-chain.org.

Yang, Z., Zhang, D., & Ye, C. (2006). Evaluation metrics for ontology complexity and evolution analysis. *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE 2006)* (pp. 162-169).

Yao, H., Orme, A.M., & Etzkorn, L. (2005). Cohesion metrics for ontology design and application. *Journal of Computer Science, 1*(1), 117-113.

Zhdanova, A.V., Krummenacher, R., Henke, J., & Fensel, D. (2005). Community-driven ontology management: DERI case study. *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 73-79).