Assessing the Completeness and Clarity of UML for Conceptual Modeling

Prabodha Tilakaratna and Jayantha Rajapakse

Abstract—Unified Modeling Language (UML) is a popular modeling grammar which is widely being used in the information system modeling process. Although initially UML more focused on system modeling, recent research studies have identified that UML can also be used in conceptual modeling. Therefore, this paper quantitatively assesses the appropriateness of UML for conceptual modeling using two significant factors: (1) to what extent the constructs of UML can represent the real world domain characteristics (completeness) (2) whether UML can represent the real world domain characteristics uniquely using its constructs (clarity). The results indicate that, UML is not capable of achieving a total completeness and clarity. Therefore, the paper recommends that UML needs to be refined in a way to achieve a higher completeness and clarity, to be used for conceptual modeling.

Index Terms—Completeness and clarity, conceptual modeling, ontology, unified modeling language (UML).

I. INTRODUCTION

Accurate development of an information system involves modeling the information system before moving to the implementation [1]. Information system modeling primarily carries out in two phases: (1) modeling the characteristics of a real world domain, which needs to be automated as an information system (conceptual modeling) (2) modeling the characteristics of the information system to be developed (system modeling).

Few decades earlier, in both conceptual and system modeling, the data and process characteristics were captured and modeled separately [2]. Data characteristics were modeled using modeling grammars such as ERM (Entity Relationship Modeling) [3] and process characteristics were modeled using grammars such as DFD (Data Flow Diagrams), workflow specification grammars [4], [5] and business process modeling grammars [6]. Software developers and researchers have identified that the characteristics of real world domains and information system cannot be precisely represented with this distinction. Therefore, they came up with a new information system development concept namely object-oriented concept, which centralized both the data and process characteristics [7]-[8].

With the rapid emergence of object-oriented concept, various grammars evolved for both object-oriented information system modeling and implementation.

Unified Modeling Language (UML) is one such popular modeling grammar, which was developed based on the object-oriented concept [9]. UML is developed in a way where it is intended to be used with all information system development methodologies which supports object-oriented concepts. Also UML can be used with any software development life cycles, and support most of the object-oriented development processes [10]-[11]. UML has thirteen different types of diagrams which have the capability to accurately model almost all the characteristics of an information system [10]. In addition, UML diagrams can easily be transformed into program codes during the information systems implementation [12]. Moreover, although UML is defined as a precise modeling grammar, it is not a barrier to include further improvements in UML concepts [9]. Because of the aforementioned factors, UML is widely being used in the current information system development field.

During the time of the origination of UML, its’ grammatical constructs and rules were defined by assigning the characteristics of information systems [13]. Therefore, UML was more appropriate to be used in system modeling. However, recently some researchers refined UML by assigning the characteristics of the real world domains into UML constructs [13]-[14]. Based on the results obtained, the researchers stated that UML constructs are capable of accurately representing the real world domain characteristics in conceptual models [13]-[16].

However, the characteristics of real world domains can be divided into three main categories: (1) static characteristics (2) state changes of the static characteristics and (3) interactions between the static characteristics. Above research studies only considered the (1) and (2) categories, during the refinement of UML [9]-[10]. Therefore, it cannot be specified that the refined UML includes necessary constructs to represent the real world domain characteristics, which belongs to all the three categories (i.e. completeness of UML).

Moreover, no previous research study carried out to assign the real world domain characteristics of all the three categories, with the UML constructs.

In addition, none of the previous studies assessed whether each real world domain characteristic can be assigned into only a single UML construct, by having a one-to-one assignment between the real world domain characteristics and UML constructs (i.e. clarity of UML). Thus, the aim of this paper is to assess the suitability of UML for conceptual modeling, based on the completeness and clarity. We use the representational theory proposed by Ron Weber [17], and its’ associated concepts of completeness and clarity as the
measurements of this study. From the aforementioned theory and concepts, the following two factors have been identified and used in assessing UML for conceptual modeling:

1) To what extent UML is capable of representing the real world domain characteristics using its grammatical constructs?

2) Whether UML can represent each real world domain characteristic uniquely using its grammatical constructs?

Factor (1) and (2) respectively measure the completeness and clarity of the modeling grammars being used in the conceptual modeling process [17]. Hence, in this study, the appropriateness of UML is being assessed using its completeness and clarity.

The remainder of this paper is structured as follows. Section 2 provides a description about Weber's representational theory [17] including the previous applications of this theory for modeling grammars. Findings and explanations related to the results obtained regarding the appropriateness of UML are given in Section 3. Finally, the paper concludes with a discussion of the overall research implications and future research directions in Section 4.

II. BACKGROUND AND RELATED WORK

Although different object-oriented modeling grammars emerged with the introduction of object-oriented concept, all of them do not possess necessary constructs and rules to represent the real world domain characteristics completely and clearly [18]. Use of inappropriate modeling grammars for conceptual modeling process will create erroneous modeling scripts, which ultimately leads to erroneous information systems [19]. Therefore, researchers have identified the requirement of assessing the appropriateness of modeling grammars before using them for the conceptual modeling process. The investigations performed on this requirement resulted in the introduction of a new theory namely, the ontological expressiveness theory [20].

This theory assesses the strengths and weaknesses of information system development methodologies in expressing the real world domain characteristics seamlessly. While developing the theory, those researchers focused on evaluating the expressiveness of information system development methodologies based on three main factors:

1) The capability to represent the real world domain characteristics.

2) The capability to track the state changes of real world domains.

3) The capability to decompose the conceptual models developed from the real world domains.

Out of the three factors, in order to assess the representational capability (i.e. first factor), Wand and Weber (the two researchers who developed the ontological expressiveness theory) created a model during their study which is known as representational model. This model was used to assess how clearly and completely an information system represents the characteristics of the real world domains, using the underlying modeling grammars.

Weber further analyzed the ontological expressiveness theory and the representational model, and came up with an extension to the ontological expressiveness theory namely, representational theory [17]. The core concept of this theory involves the use of modeling grammars to represent the real world domain characteristics seamlessly in the conceptual modeling process. Representing the real world domain characteristics comprises discussing what exists in the real world domains and how they exist. During the development of the ontological expressiveness theory, Wand and Weber identified what exist in the real world domains, how things behave in the real world domains and how the characteristics of real world domains can comprehensively be observed using philosophical disciplines [21].

Ontology is a subdivision of philosophy, which signifies the abstraction and representation of different domain characteristics. Thus, ontological approach was used in developing both the ontological expressiveness theory and the representational theory. Mario Bunge’s ontology [22], [23] has been chosen for the development of the two theories, because this was the only existing ontology which covered the entire real world by discussing its domain characteristics and the relationships between those characteristics. Besides, the most widely used ontology in the current information systems research field was that of Bunge [15–17], [20]. Therefore, using this ontology, Wand and Weber derived a set of ontological concepts and they argued that this set of concepts has the capability to cover the real world domain characteristics in general.

Hence, these concepts been used in the representational theory to assess the completeness and clarity of modeling grammars in representing the real world domain characteristics. This is known as the representational capability of a modeling grammar [17]. Weber defined that, completeness means the capability of a modeling grammar to represent any given real world domain scenario without losing information. Conversely, clarity means the ability to uniquely represent the real world domain characteristics using the grammatical constructs. Completeness and clarity of modeling grammars were assessed by mapping the grammatical constructs with the previously defined set of ontological concepts [20]. This mapping was performed adhering to the mapping requirements proposed by Wand and Weber in the ontological expressiveness theory [20].

According to these requirements, the mapping can be bi-directional, but it should ensure one-to-one relationships between the two sets being used. However, Weber has figured out that except one-to-one relationships, certain other mapping relationships can also exist between the grammatical constructs and the ontological concepts. Such relationships were identified as incorrect relationships and were named as representational deficiencies. The identified representational deficiencies could be divided into four main categories as construct deficit, redundancy, overload and excess.

Weber has observed that, construct deficit affects and undermines the completeness of modeling grammars [17]. Construct deficit means one-to-zero relationships, where certain ontological concepts do not have matching grammatical constructs. Furthermore, Weber stated that the remaining representational deficiencies (i.e. construct
Construct overload can have more than one matching grammatical constructs. Constructs do not have matching ontological concepts. Zero-to-many relationship, where certain grammatical constructs do not have matching ontological concepts.

Construct redundancy means many-to-one relationship, where one ontological concept has only one matching grammatical construct. Constructs excess means one-to-many relationships, where one ontological concept is represented by more than one grammatical construct.

Redundancy, overload and excess effect on the clarity of the representational theory in the e-Business area [24], [25]. Consequently, representational theory can effectively be used to compare different modeling techniques that were being used in the enterprise system interoperability [24]. Four modeling techniques were used for the comparison and the ontological completeness and clarity of those techniques were assessed using the theory.

Another research study carried out to assess the representational capability of modeling grammars being used in process modeling [6]. Representational theory was used to assess twelve selected process modeling grammars in representing the real world domain characteristics. Based on the analysis they have identified the process modeling grammars which are most suitable to be used in the business process modeling area. Keen and Lakos are two researchers who used representational theory to assess the completeness of six process modeling techniques using some essential features identified by them [26]. Based on the analysis, they concluded that, representational theory can effectively be used in interpreting and comparing the process modeling techniques. Green and Rosemann examined the completeness and clarity of Event-Driven Process Chain (EPC) notation using the representational theory [27].

Some recent research studies focused on utilizing the set of ontological concepts proposed by Wand and Weber [20] to assess the applicability of UML in the conceptual modeling process. Nevertheless, they did not focus on assessing the completeness or clarity of UML in representing the real world domain characteristics by referring to the representational theory [17]. Therefore, this area still remains undiscovered, where further investigations can be carried out.

III. REPRESENTATIONAL CAPABILITY ASSESSMENT

This section exemplifies the investigation with regards to the assessment of the representational capability (completeness and clarity) of UML, which carried out based on the representational theory.

A. Completeness of UML

As specified in the previous section, completeness of UML will be assessed using the construct deficit. In simple terms, if construct deficit exists with a particular modeling grammar, some of the real world domain characteristics would not be able to represent from the modeling script developed using that grammar. To identify whether UML is not capable of representing certain real world domain characteristics, UML constructs have been mapped with the ontological concepts and the results are given in Table I. Each tick appeared in the table indicates that the particular ontological concept can be represented by a construct of UML. Based on the mapping, the total degree of deficit calculated considering all the unmapped ontological concepts relative to the total ontological concepts as follows:

Total degree of deficit = \[ \frac{\text{Number of ontological concepts that do not have matching UML constructs}}{\text{Total number of ontological concepts}} \]

Since the construct deficit illustrates the incompleteness of a modeling grammar, total degree of completeness has been calculated as follows:

Total degree of completeness = \( (1 - \text{Total degree of deficit}) \) %

According to the above equations, the completeness of UML or the ability of UML in representing any real world domain characteristic, has achieved 68.42%. Conversely, from 31.58%, UML do not contain necessary constructs to represent some real world domain characteristics.

B. Clarity of UML

The main requirement of clarity involves, the unique representation of real world domain characteristics using each UML construct. The remaining three representational deficiencies i.e. construct redundancy, overload and excess, will be considered in this sub section for the assessment of the clarity of UML. Firstly, we consider construct redundancy. Redundant constructs mean having more than one grammatical construct to represent the same real world domain characteristic, which may create confusions during the modeling process. The UML constructs have been mapped with the ontological concepts as depicted in Table II, and we have identified the numbers of UML constructs that can be mapped with each ontological concept. The total degree of redundancy was calculated considering all the UML constructs which represented one ontological concept using more than one UML construct as follows.
### Table I: Assessing the Completeness of UML using Constructs Deficit

<table>
<thead>
<tr>
<th>Ontological concept (ontological entity and related concepts)</th>
<th>UML</th>
<th>Ontological concept (ontological state and related concepts)</th>
<th>UML</th>
<th>Ontological concept (ontological event and related concepts)</th>
<th>UML</th>
<th>Ontological concept (ontological system and related concepts)</th>
<th>UML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thing</td>
<td>✓</td>
<td>State</td>
<td>✓</td>
<td>Event</td>
<td>✓</td>
<td>System</td>
<td>✓</td>
</tr>
<tr>
<td>Class</td>
<td></td>
<td>Conceivable state space</td>
<td></td>
<td>Conceivable event space</td>
<td>✓</td>
<td>System composition</td>
<td>✓</td>
</tr>
<tr>
<td>Natural kind</td>
<td>✓</td>
<td>State law</td>
<td>✓</td>
<td>Lawful event space</td>
<td>✓</td>
<td>System environment</td>
<td>✓</td>
</tr>
<tr>
<td>Property</td>
<td>✓</td>
<td>Lawful state space</td>
<td></td>
<td>External event</td>
<td>✓</td>
<td>System structure</td>
<td>✓</td>
</tr>
<tr>
<td>Intrinsic property</td>
<td>✓</td>
<td>Stable state</td>
<td></td>
<td>Internal event</td>
<td>✓</td>
<td>Sub system</td>
<td>✓</td>
</tr>
<tr>
<td>Mutual property</td>
<td>✓</td>
<td>Unstable state</td>
<td></td>
<td>Well defined event</td>
<td></td>
<td>System decomposition</td>
<td>✓</td>
</tr>
<tr>
<td>Emergent property</td>
<td>✓</td>
<td>History</td>
<td></td>
<td>Poorly defined event</td>
<td></td>
<td>Level structure</td>
<td></td>
</tr>
<tr>
<td>State function</td>
<td></td>
<td>Action</td>
<td>✓</td>
<td>Transformation</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional schema</td>
<td>✓</td>
<td></td>
<td></td>
<td>Lawful transformation</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative change</td>
<td>✓</td>
<td></td>
<td></td>
<td>Coupling</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualitative change</td>
<td>✓</td>
<td></td>
<td></td>
<td>Process</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of deficit</td>
<td>25.00%</td>
<td>Degree of deficit</td>
<td>62.50%</td>
<td>Degree of deficit</td>
<td>18.18%</td>
<td>Degree of deficit</td>
<td>28.37%</td>
</tr>
</tbody>
</table>

The degree of deficit: 31.58%
The degree of completeness: 68.42%

### Table II: Assessing the Clarity of UML using Constructs Redundancy

<table>
<thead>
<tr>
<th>Ontological Concept</th>
<th>UML</th>
<th>Ontological Concept</th>
<th>UML</th>
<th>Ontological Concept</th>
<th>UML</th>
<th>Ontological Concept</th>
<th>UML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thing</td>
<td>2</td>
<td>State</td>
<td>4</td>
<td>Event</td>
<td>4</td>
<td>System</td>
<td>2</td>
</tr>
<tr>
<td>Class</td>
<td>0</td>
<td>Conceivable state space</td>
<td>0</td>
<td>Conceivable event space</td>
<td>2</td>
<td>System composition</td>
<td>2</td>
</tr>
<tr>
<td>Natural kind</td>
<td>2</td>
<td>State law</td>
<td>1</td>
<td>Lawful event space</td>
<td>2</td>
<td>System environment</td>
<td>1</td>
</tr>
<tr>
<td>Property</td>
<td>1</td>
<td>Lawful state space</td>
<td>0</td>
<td>External event</td>
<td>1</td>
<td>System structure</td>
<td>1</td>
</tr>
<tr>
<td>Intrinsic property</td>
<td>1</td>
<td>Stable state</td>
<td>0</td>
<td>Internal event</td>
<td>1</td>
<td>Sub system</td>
<td>0</td>
</tr>
<tr>
<td>Mutual property</td>
<td>1</td>
<td>Unstable state</td>
<td>0</td>
<td>Well defined event</td>
<td>0</td>
<td>System decomposition</td>
<td>1</td>
</tr>
<tr>
<td>Emergent property</td>
<td>1</td>
<td>History</td>
<td>0</td>
<td>Poorly defined event</td>
<td>0</td>
<td>Level structure</td>
<td>0</td>
</tr>
<tr>
<td>State function</td>
<td>0</td>
<td>Action</td>
<td>3</td>
<td>Transformation</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional schema</td>
<td>2</td>
<td></td>
<td></td>
<td>Lawful transformation</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative change</td>
<td>1</td>
<td></td>
<td></td>
<td>Coupling</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qualitative change</td>
<td>0</td>
<td></td>
<td></td>
<td>Process</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The degree of redundancy: 43.48%
Total degree of redundancy = \[
\frac{\text{Number of UML constructs that have a mapping to the same ontological concept}}{\text{Total number of UML constructs}}
\]

Next, we have moved to construct excess. Existence of modeling grammar constructs which do not have any meaningful characteristic to be mapped in the intended domain will create construct excess. Excess UML constructs means the constructs of UML which cannot be mapped with any real world domain characteristic. Unavailability of matching real world domain characteristics will make such constructs worthless in the conceptual modeling process.

Therefore, excess UML constructs will not be utilized in creating conceptual models. After mapping the UML constructs with the ontological concepts, the excess UML constructs have been identified which are given in Table III. The total degree of excess was calculated by considering all the excess UML constructs relative to the total UML constructs as follows:

Total degree of excess = \[
\frac{\text{Number of UML constructs that do not have a mapping to any of the ontological concept}}{\text{Total number of UML constructs}}
\]

The final representational deficiency which used to assess the UML clarity is construct overload. This illustrates the situation where one UML construct is used to represent multiple real world domain characteristics. This can create difficulties during the development of conceptual models, because different modelers may use different UML construct to represent the same real world domain characteristic. Therefore, it may be difficult for the stakeholders to grasp the accurate idea given by the conceptual model. From the mapping between the UML constructs and the ontological concepts, the UML constructs have been identified, which can be mapped with more than one ontological concept (Table IV). The total degree of overload has been calculated considering all the UML constructs that have multiple matching real world domain characteristics, relative to the total UML constructs:

Total degree of overload = \[
\frac{\text{Number of UML constructs that can be mapped with more than one ontological concept}}{\text{Total number of UML constructs}}
\]

In order to obtain a single value for the total degree of clarity, the numerical values obtained for the three representational deficiencies need to be merged together. Nevertheless, this merging requires the existence of a common source, thereby relative to the common source, the values of the three representational deficiencies can be added together. The ability of UML to represent any given real world domain characteristic (i.e. completeness) has been considered as the common source, and the values of the three representational deficiencies relative to the total degree of completeness were identified (Table V). Since the three representational deficiencies decrease the clarity, those relative values were used to achieve a single value for the lack-of-clarity of UML. Hence, the final value obtained for the total degree of clarity of UML involves 62.93%.

<table>
<thead>
<tr>
<th>Grammar</th>
<th>Proposed overload grammatical constructs</th>
<th>Degree of overload</th>
</tr>
</thead>
<tbody>
<tr>
<td>UML</td>
<td>Use case relationship, Activity, Activation, Event</td>
<td>8.70%</td>
</tr>
</tbody>
</table>

C. Discussion

Based on the results obtained in the previous two sub-sections, it can be identified that the conceptual models create using UML will not be hundred percent accurate due to its lack of total completeness and clarity. Therefore, refining the UML constructs and rules is required to achieve a total representational capability.

To achieve the remaining 31.58% completeness, new or modified UML constructs need to identify to be mapped with the remaining unmapped real world domain characteristics. Similarly, to remove the lack-of-clarity, UML should be refined in a way to remove one-to-many, many-to-one and zero-to-one relationships that exist between UML constructs and ontological concepts. Nevertheless, refining an existing modeling grammar is a process which needs to be done in a careful manner, since it would be worthless if the refinements make the grammar complex. Also, all the refinements should be done in a way where such modifications can clearly be justified over the existing grammatical constructs and rules.

IV. IMPLICATIONS AND FUTURE WORKS

This research study has significant implications in both the practice and research as follows. When we consider the practical aspect, the research outcomes are significant for all the stakeholders of the conceptual modeling process such as modelers, designers and users. This is for the reason that, in terms of completeness and clarity, modelers and designers can have a clear idea about the extent to which UML is capable of representing the real world domain characteristics. This will help them to decide whether to use this modeling grammar for the modeling process or not. Moreover, although UML will be utilized, the stakeholders would be able to identify that, achieving hundred percent accurate conceptual models is impossible, due to the lack of completeness and clarity of UML in representing the real world domain characteristics. Also, knowing the completeness and clarity of UML will help them in dealing
with the ambiguity situations that occur during the modeling process as well as in viewing the completed conceptual models.

If we consider the results from the research point of view, researchers can compare the representational capability of UML with the other research studies performed on the representational capability of certain modeling grammars [6], [26]. This will help them to have a clear idea about the level of the appropriateness of UML with regards to the other modeling grammars being used in the conceptual modeling process. Also, the same approach which has been used in this paper can be applied to other object-oriented modeling grammars to assess their representational capability. Most significantly, this study can be considered as another research work which proved the validity and the effective use of representational theory in assessing the modeling grammars for conceptual modeling.

This research study will further be expanded with the investigations of refining UML constructs and rules for increasing the completeness and clarity. Modifications required to be carried out for the UML constructs and rules until it reaches a total completeness and clarity. Subsequently, an empirical study will be performed to assess the success and the usability of the refined UML in the practical scenario.

<table>
<thead>
<tr>
<th>TABLE V: ASSESSING THE TOTAL DEGREE OF LACK-OF-CLARITY OF UML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative value of the</td>
</tr>
<tr>
<td>representational deficiency</td>
</tr>
<tr>
<td>Degree of construct redundancy, relative to the degree of completeness</td>
</tr>
<tr>
<td>Degree of construct excess, relative to the degree of completeness</td>
</tr>
<tr>
<td>Degree of construct overload, relative to the degree of completeness</td>
</tr>
<tr>
<td>Total degree of lack-of-clarity</td>
</tr>
</tbody>
</table>

This research study has some limitations as follows. Firstly, the real world domains consist of an infinite number of characteristics and hence it is impossible to map all these real world domain characteristics with UML constructs. Therefore, we utilized the set of ontological concepts proposed by Wand and Weber [20], and as they stated we assume that those concepts can represent the real world domains in general. Secondly, Weber has specified that the selection of the grammatical constructs from a modeling grammar may depend on a person’s perception and the way how each person is looking at that grammar [17]. Also he stated that the mapping between the grammatical constructs and the ontological concepts may differ from person to person. Therefore, even if some other researchers perform the same research study, there can be a possibility that they will get some different values than the ones specified in this paper.

ACKNOWLEDGMENT

We are indebted to Prof. Ron Weber (Faculty of Information Technology - Monash University, Australia) and Mr. Kapila Dias (University of Colombo School of Computing, Sri Lanka) for the helpful comments and reference suggestions given on this paper.

REFERENCES


Prabodha Tilakaratna received the B.Sc. degree in Information and Communication Technology from the University of Colombo School of Computing, Sri Lanka in 2010, BCS (British Computer Society) diploma in 2009. Currently she is a PhD student at the Faculty of Information Technology, Monash University. Before starting the PhD, she has been involved with teaching and research in software engineering at the University of Colombo School of Computing, Sri Lanka.

Jayantha Rajapakse received the B.Sc. degree from the University of Peradeniya, Sri Lanka in 1981, Research Masters degree in Computer Science from the University of Queensland, Australia in 1997, and the Ph.D. degree in Information Systems from the University of Melbourne, Australia in 2006. Currently he is a Senior Lecturer at the Faculty of Information Technology, Monash University. He has been involved with teaching and research in information systems and computer science for over 15 years at the University of Queensland, the University of Melbourne, Queensland University of Technology, Manchester Metropolitan University (Sri Lanka), and Monash University.