Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2002 Proceedings

Americas Conference on Information Systems (AMCIS)

December 2002

ANALYZING UNIFIED MODELING LANGUAGE USING CONCEPT MAPPING

Keng Siau
University of Nebraska-Lincoln

Zixing Shen
University of Nebraska-Lincoln

ANALYZING UNIFIED MODELING LANGUAGE USING CONCEPT MAPPING

Keng Siau and Zixing Shen
University of Nebraska-Lincoln
ksiau@unl.edu

Abstract

The Unified Modeling Language (UML) is a visual modeling language for object-oriented software development. Although the Object Management Group (OMG) adopted UML as its standard modeling language in 1997, the consensus in the academic community and the industry is that further research is needed to evaluate, enhance, extend, and formalize UML. A substantial amount of research has been conducted to this end, but most was based on common sense, informal observation, and intuition. Systematic and empirical studies to analyze, evaluate, and enhance UML have been lacking. This paper attempts to evaluate and study UML from a cognitive perspective. Specifically, the concept mapping approach will be used to investigate the cognitive process involved in using the UML diagrams. The conceptual maps to be developed on UML diagrams will serve as a cognitive basis for UML evaluation, enhancement, and extension.

Keywords: Unified modeling language, concept mapping

Introduction

The Unified Modeling Language (UML), which incorporates various concepts from a large number of different methods, is tailored specifically for object-oriented system development. Because it is a general-purpose modeling language that can be used to visualize, specify, construct, and document the artifacts of a software-intensive system (Booch 1999), UML quickly emerged as the software industry's dominant modeling language. In fact, UML is not only a *de facto* modeling language standard, but also a *de jure* standard. The Object Management Group (OMG) adopted UML as its standard modeling language in late 1997, and is now proposing the UML specification for international standardization.

UML defines nine modeling techniques. They are the Class Diagram, Use-Case Diagram, State-Chart Diagram, Activity Diagram, Sequence Diagram, Collaboration Diagram, Object Diagram, Component Diagram, and Deployment Diagram (Booch *et al.* 1999). The Class Diagram, Object Diagram, Component Diagram, and the Deployment Diagram depict the static aspect of the system. On the other hand, the Sequence Diagram, Collaboration Diagram, State Chart Diagram, and Activity Diagram describe the behavioral (dynamic) aspects of the system. These diagrams are interrelated and used to represent different perspectives of the system.

The objective of this research is to empirically evaluate the effectiveness of UML diagrams. The concept mapping approach developed by Trochim (1989) will be employed for this empirical study.

Literature Review

UML attempts to offer simple, intuitive notations that are understandable to non-programmers (Siau and Cao 2001). UML not only promotes communication and understanding between different project stakeholders (Fowler 2000), but also provides an extensive conceptual base for a broad spectrum of application domains (Selic 1999). For example, Use Case Diagrams, and Activity Diagrams are found to be particularly useful in documenting user requirements (Jackson 1998), Class Diagrams can be used to map the data model (Shah and Slaughter 2000), and Use Case Diagrams reveal the system's scope and purpose (Cockburn

2001). Moreover, applications of UML diagrams are extended to areas such as web applications (Conallen 1999; Hennicker and Koch 2000) and real-time systems (Selic 1999; Herzberg 1999).

Despite its increasing popularity and widespread acceptance, UML has been criticized for its ambiguity and inconsistency (Simon and Graham 1999). Whittle (2000) pointed out that there is no standard formal semantics for any part of UML and, as a result, UML is grossly imprecise. Many UML standard elements (stereotypes, tagged values, and constraints) have sparse semantics and were inconsistently named and organized (Kobryn 1999). The Whole-Part relationship in UML is formalized in a confusing, and thus unsatisfactory way (Barbier *et al.* 2000). In addition, the current version of UML was also found to be inadequate: the core constructs and extension mechanisms defined in UML were not suitable for modeling spatio-temporal data (Price *et al.* 2000), did a poor job in representing configuration patterns (Fontoura and Lucena 2001), and failed to provide support for modeling mobile agents (Klein *et al* 2000).

UML is also found to be very complex. UML consists of various graphical notations, which capture the static system structure, dynamic system structure, system component behavior, and system component interaction. A complexity analysis revealed that each diagram in UML is not distinctively more complex than techniques in other object-oriented methods, but, as a whole, UML is very complex – 2-11 times more complex than other object-oriented methods (Siau and Cao 2001). Previous research also identified deficiencies and flaws in the UML diagramming techniques. For example, the process-centric Use Case Diagrams cannot provide an adequate basis for specifying data-centric Class Diagrams (Halpin 2000). Class Diagrams were found to be not expressive enough to clearly document fundamental aspects in distributed systems development (Lago 2000). Activity Diagrams did not preserve object class boundary (Jackson 1998).

In short, UML needs to be further evaluated, extended, and formalized. Although these criticisms of and comments on UML sound reasonable, they are not supported by empirical evidence and are based mainly on gut feelings and "common sense". Systematic and empirical studies to evaluate UML have been lacking. For example, Dobing and Parsons (2000) stressed that the paucity of empirical research on the effectiveness of various modeling techniques in UML is troubling.

Theoretical Background

Concept mapping, also known as cognitive mapping, is a qualitative methodology that is extensively used for conceptualization purposes. Concept (cognitive) maps are the visual representation of the outcomes of the concept mapping process. They are widely used in multiple disciplines, such as knowledge acquisition (Novak and Gowin 1984), conceptualization of essays and other text (Rico 1983), and strategic management (Eden and Ackermann 1998). This study will utilize the concept mapping approach developed by Trochim and Linton (Trochim and Linton 1986; Trochim 1989) to study and evaluate UML.

According to Trochim and Linton (1986), the process steps (the steps in conducting conceptualization), perspective origins (the persons involved in the process steps), and representation form (the final form the conceptualization is presented in) are the three components of the conceptualization process. The conceptualization process begins with the entity, which refers to "each distinguishable thought or idea expressed verbally as a word, phrase, sentence, or other text unit" (Trochim and Linton 1986, p. 290). Transformation from initial entities to a conceptual representation involves generation of, structuring of, and representation of the conceptual domain. Trochim's approach invites diverse stakeholders to participate in the conceptualization process and generate ideas on a specific topic of interest. Then the interrelationships between these ideas are specified, and multivariate statistical techniques (i.e., cluster analysis and multi-dimensional scaling) are applied to the information, and results are represented in the form of a map for further interpretation and utilization. Because of its role in guiding theory and concept formation, Trochim's approach in concept mapping has been gaining popularity in studies of social services (Galvin 1989), health care (Marquart 1989), education (Caracelli 1989), social technology (Mannes 1989), social science (Linton 1989), and theory development (Valentine 1989).

Research Methodology

Concept mapping consists of six steps. The six steps are:

Step 1: Preparation: Selecting the participants and developing foci are the two tasks in the preparation stage of the concept mapping process. The facilitators must work with the parties involved to decide who will participate. After the participants are selected, the facilitators work with them to develop focus statements for Step 2 and for ratings to guide the conceptualization process.

Step 2: Generation of Statements: Once the participants and focus statements are defined, the actual concept mapping process begins with the generation of a set of statements. Brainstorming can be used to encourage participants to generate as many statements as possible. Facilitators usually record the statements as they are generated and edit them after the brainstorming session to make sure that each is consistent with the focus statements and understandable to each participant in the group. Methods other than brainstorming, such as interviewing and self-reporting, can also be utilized to generate the conceptual domain. For example, the documentary coding method (Wrightson 1976) can be employed to abstract statements or entities from existing text documents such as annual reports, internal organizational memos, and interview records. The outcome of this stage is a comprehensive list of ideas relevant to the topic of interest (in our case, UML).

Step 3: Structuring of Statements: With a set of statements for a given focus, the concept mapping proceeds to configure how the statements are related to each other. Typically, an unstructured card-sorting procedure (Rosenberg and Kim 1975) is used to obtain information about interrelationships. Participants sort the brainstormed statements in any way that make sense to them. After the sorting task is completed, the results are combined across people to get a combined group similarity matrix, which is considered the relational structure of the conceptual domain. The second task in structuring involves rating each statement according to the rating focus statement. A Likert-type response scale is always used to indicate the importance or priority associated with each statement.

Step 4: Representation of Statements: Various maps are produced in this step. First, non-metric multidimensional scaling (Kruskal and Wish 1978) is conducted to draw the point map, which shows the distance between statements. Statements that are closer to each other in the point map are those that have been sorted together more frequently. Second, a hierarchical cluster analysis (Everitt 1980) is conducted to make the cluster map, which represents higher-order conceptual groupings of the original set of statements. Finally, the point-rating map or the cluster-rating map is constructed by adding the averaged rating either by point or by cluster.

Step 5: Interpretation of Maps: The statement list, cluster list, point map, cluster map, point-rating map, and cluster-rating map are presented to participants. The participants are then asked to name the cluster and to see whether the visual structure makes sense. The cluster map and cluster-rating map are refined according to the participants' feedback.

Step 6: Utilization of Maps: Attention is turned back to the original reason for conducting concept mapping at this point in the process. In our case, the concept maps, which include the point map, cluster map, point-rating map, and cluster-rating map, are used to evaluate UML and suggest enhancements and extensions to UML.

Research Procedure

The first two steps of this research are completed (i.e., preparation and statement generation). The focus statement for this study was stated in the form of an instruction to the study participants. Specifically, this instruction was operationalized as: Generate statements (short phrases or sentences) that describe the difficulties in drawing and interpreting UML models.

The subjects for the statement generation were senior and graduate students enrolled in a semester-long UML course at a major Midwest public university. Thirty-six subjects participated in statement generation at the end of the course, by which time they had completed all class assignments, a semester long project, and the final exam. Each subject was asked to generate three statements on difficulties in drawing UML diagrams and three statements on difficulties in interpreting UML diagrams. Around 200 statements concerning difficulties in drawing and interpreting UML diagrams were obtained from the 36 participants. These statements are currently being examined to remove redundancies, awkwardness, or technical jargon.

The subjects who will sort and rate the statements generated will be chosen from graduate students that have completed the UML course. Some of them might have participated in the statement generation, and others might not. A total of 15 students will serve as the subjects in structuring statements. The participants will be given a listing of statements laid out in mailing label format, and be asked to cut the listing into slips with one statement on each slip. Then they will be instructed to group the statement slips into piles "in a way that make sense to you". The only restriction in this sorting task will be that there cannot be: (1) N piles (N= total number of statements), (2) one pile consisting of all the statements (Trochim, 1989).

For the sorting task, the 15 participants will be asked to rate each statement on a 5-point Likert-type response scale in terms of the relative importance of each statement. The operational form for rating is stated below:

Using the following scale, rate each statement on its relative importance in drawing and interpreting UML diagrams.

1	2	3	4	5
relatively less important	somewhat important	moderately important	very important	extremely important

When the sorting and rating data are collected, a statistical analysis will be performed. A multi-dimensional scaling and cluster analysis will be conducted to produce various concept maps for interpretation and utilization.

Conclusion and Expected Contribution

The research will assess empirically the various modeling diagrams in UML. The concept maps to be developed on UML diagrams will provide empirical evidence on the effectiveness of UML diagrams in supporting software development. Also, the concept maps will identify problem areas in UML diagrams that need further modification or enhancement.

References

Barbier, F., Henderson-Sellers, B., Opdahl, A., and Gogolla, M. "The Whole-Part Relationship in the Unified Modeling Language: A New Approach." In K. Siau, and T. Halpin (eds.) Unified Modeling Language: Systems Analysis, Design and Development Issues, (pp.186-210). Hershey, PA :Idea Group Publishing, 2000.
Booch, G. "UML in Action." Communications of the ACM, (42:10), 1999, pp. 26-28.
Booch, Rumbaugh J., and Jacobson, I. The Unified Modeling Language User Guide. Reading, MA: Addison Wesley, 1999.
Caracelli, V. "Structured Conceptualization: A Framework for Interpreting Evaluation Results." Evaluation and Program Planning, (12), 1989, pp. 45-52.
Cockburn, A. Writing Effective Use Cases. NJ: Addison Wesley, 2001.
Conallen, J. "Modeling Web Application Architecture with UML." Communications of the ACM, (42:10), 1999, pp. 63-70.
Dobing, B., and Parson, J. "Understanding the Role of Use Cases in UML: A Review and Research Agenda." Journal of Database Management, (11:4), 2000, pp. 28-36.
Eden, C., and Ackermann, F. Making Strategy: The Journey of Strategic Management. SAGE Publications, 1998.
Everitt, B. Cluster Analysis (2nd Edition). New York, NY: Halsted Press, A Division of John Wiley and Sons, 1980.
Fontoura M., and de Lucena, C. "Extending UML to improve the representation of design patterns." JOOP-Journal Of Object-Oriented Programming, (13:11), 2001, pp. 12-19.
Fowler, M. "Why Use the UML?" Retrieved from World Wide Web: http://www.sdmagazine.com/documents/s=746/sdm0003z/0003z3.htm, 2000.
Galvin, P. "Concept Mapping for Planning and Evaluation of a Big Brother/ Big Sister Program." Evaluation and Program Planning, (12), 1989, pp. 53-57.
Halpin, T. "Supplementing UML with Concepts from ORM." In K. Siau and T. Halpin (eds.) Unified Modeling Language: Systems Analysis, Design and Development Issues, (pp.167-184). Hershey, PA: Idea Group Publishing, 2000.
Hennicker, R. and Koch, N. "Systematic Design of Web Application with UML." In K. Siau and T. Halpin (eds.) Unified Modeling Language: Systems Analysis, Design and Development Issues, (pp.1-20). Hershey, PA: Idea Group Publishing, 2000.
Herzberg, D. "UML-RT as a candidate for modeling embedded real-time systems in the telecommunication domain." UML'99 – The Unified Modeling Language 1999, 1723, 330-338
Jackson, R. B. "Application System Requirements Definition Using the Unified Modeling Language (UML)." In Proceedings of American Conference on Information System, 1998, 689-691.
Klein, C., Rausch, A., Sihling, M., and Wen, Z. "Extension of the Unified Modeling Language for Mobile Agents." In K. Siau. and T. Halpin. (eds.) Unified Modeling Language: Systems Analysis, Design and Development Issues. (pp.118-128). Hershey, PA: Idea Group Publishing, 2000.
Kobryn, C. "UML 2001: A Standardization Process Odyssey." Communications of the ACM, (42:10), 1999, pp. 29-37.
Kruskal, J. B. and Wish, M. Multidimensional Scaling. Beverly Hills, CA: Sage Publications, 1978.

Lago, P. "Rendering Distributed Systems in UML." In K. Siau and T. Halpin (eds.) Unified Modeling Language: Systems Analysis, Design and Development Issues, (pp.129-150). Hershey, PA: Idea Group Publishing, 2000.

Linton, R. "Conceptualizing Feminism: Classifying Social Science Concept." Evaluation and Program Planning, (12), 1989, pp. 25-29.

Mannes, M. "Using Concept Mapping for Planning the Implementation of a Social Technology." Evaluation and Program Planning, (12), 1989, pp. 67-74.

Marquart, J. "A Pattern Matching Approach to Asses the Construct Validity of an Evaluation Instrument." Evaluation and Program Planning, (12), 1989, pp. 37-43.

Novak, J. D. and Gowin, D. B. Learning How to Learn. Cambridge, England: Cambridge University Press, 1984.

Price, R., Tryfona, N., and Jensen, C. "Extend Spatio-temporal UML: Motivations, Requirements, and Constructs." Journal of Database Management, (11:4), 2000, pp. 14-27.

Rico, G. L. Writing the Natural Way: Using Right-Brain Techniques to Release your Expressive Powers. Los Angeles, CA: J.P. Tarcher, 1983.

Rosenberg, S. and Kim. M. P. "The Method of Sorting as a Data-Gathering Procedure in Multivariate Research." Multivariate Behavioral Research, (10), 1975, pp. 489-502.

Selic, B. "Turning Clockwise: Using UML in the Real-Time Domain." Communications of the ACM, (42:10), 1999, pp. 46-54.

Siau, K. "Information Modeling and Method Engineering: A Psychological Perspective." Journal of Database Management, (10:4), 1999, pp. 44-50.

Siau, K. and Cao, Q. "Unified Modeling Language (UML)—a Complexity Analysis." Journal of Database Management, (12:1), 2001, pp. 26-34.

Simons A. J. H. and Graham, I. "30 Things That Go Wrong In Object Modelling With UML 1.3." In H. Kilov, B. Rumpe, and I. Simmonds (eds.) Precise Behavioral Specification of Businesses and Systems. Kluwer Academic Publishers. 1999.

Shah, D. and Slaughter, S. "Data Modeling and UML". In K. Siau and T. Halpin (eds.) Unified Modeling Language: System Analysis, Design and Development Issues, (pp.43-60). Hershey, PA: Idea Group Publishing, 2000.

Trochim, W. and Linton, R. "Conceptualization for Evaluation and Planning." Evaluation and Program Planning, (9), 1986, pp. 289-308.

Trochim, W. "An Introduction to Concept Mapping for Planning and Evaluation." Evaluation and Program Planning, (12:1), 1989, pp. 1-16.

Valentine, K. "Contribution to the Theory of Care." Evaluation and Program Planning, (12), 1989, pp. 17-23.

Whittle, J. "Formal Approaches to System Analysis Using UML: An Overview." Journal of Database Management, (11:4), 2000, pp. 4-13.

Wrightson, M. "The Documentary Coding Method." In R. Axelrod (ed.) The Structure of Decision: The Cognitive Maps of Political Elites. Princeton, NJ: The Princeton University Press, 1976.