On the UML use in the Brazilian industry: A state of the practice survey

Kleinner Farias, Lucian Gonçales, Vinicius Bischoff
Graduate Program on Applied Computing (PPGCA)
Univ. of Vale do Rio dos Sinos, São Leopoldo, Brazil
kleinnerfarias@unisinos.br,{lucianj, viniciusbischof}
@edu.unisinos.br

Bruno da Silva[1], Everton Guimarães[2], Jacob Nogle[1]
[1]Dept. of Computer Science & Software Engineering
California Polytechnic State University,
San Luis Obispo, USA
{bcdasilv, jnogle}@calpoly.edu
[2]Department of Computer Science, Drexel University,
Philadelphia, USA
etg38@drexel.edu

Abstract— Context: **The Unified Modeling Language (UML) has become the standard for modeling software. Several surveys on the UML usage have been proposed in recent years. However, none of them explores the UML use in specific regional scope, and thus little is known about the practices and perceptions of UML use from the perspective of practitioners in the Brazilian industry.** *Objective:* **This paper reports on a survey focused on identifying the state-of-the-practice of the Brazilian industry for what concerns the UML usage in real-world settings.** *Method:* **In total, 222 practitioners from 140 different Information Technology companies have answered an on-line (or printed) questionnaire concerning their UML use experiences, the difficulty in adopting UML and what should be done to increase the UML adoption in practice.** *Result:* **The results show that: (1) 60 participants (28.2%) have used UML in their daily work, while 73.2% have not; (2) 55.41% of the surveyed participants did not disagree with the statement that UML is the "lingua franca" in software modeling; (3) 61.26% reported to find that the automatic creation of UML diagrams to represent a big picture of the system under development would be useful to boost UML use.** *Conclusion:* **The UML is not often used in the work life of participants. In addition, no relationship was identified between the use of UML and the participant company being a software factory.**

Unified Modeling Language; UML; Practice; Industry

I. INTRODUCTION

The Unified Model Language (UML) [1] is a unification generated from the main methods of software modeling and provides a common notation design covering software analysis to software deployment. Everything in software can be detailed when providing and maintaining software for costumers, resulting in a vast set of diagrams. In the context of the development process, UML assumes that its adoption implies a series of benefits, such as providing a common understanding between team members, comprehension of development details, and increased efficacy in software development. Some studies argue that these benefits are consequences of a full and formal application, where the UML must be applied during the whole software project, and the practitioners have a firm grasp on the usage of this language. As this reality is uncertain, several surveys on the UML usage have been proposed in recent years with the purpose to investigate how the UML is used in practice.

However, there is still a lack of understanding regarding practices and perceptions of UML usage from the perspective of practitioners. In fact, the state-of-the-art UML use in industry diverged on how software industry applies it as well as how practitioners use UML in practice. In addition, the current literature about the UML use in industry was not yet capable of concluding if UML is the de facto standard of modeling languages. Specifically, existing surveys have focused on collecting opinions from participants from different parts of the world. This turns a problem because it assumes that perceptions and fragments from participants opinions spread worldwide are valid on a local and regional scope.

To account for this, this paper focuses on identifying the state-of-the-practice of the Brazilian industry for what concerns the UML usage in world-wide companies. Specifically, this work seeks to investigate how UML is being used in practice, in relation to the relevance of the UML in software projects, i.e. whether the use of UML in software projects and applications is frequent, finding a convergence among developers on the status of UML as a lingua franca, and suggestions for possible improvements in the UML. Exploring these issues is important because we have learned from empirical studies (e.g., [14]) that UML usage has a significantly positive impact on the functional correctness of realized changes in source code in the context of maintenance and evolution tasks.

To achieve these objectives, this study reports on a survey with practitioners of the Brazilian industry. This survey consists of three research questions, and was developed following well-defined guidelines and previous studies, such as [3][4][5][6]. Specifically, a search was conducted on the literature to pinpoint gaps, and to grasp how other studies designed their questionnaires and collected their data. Moreover, we carefully selected a representative set of participants. Thus, all practitioners work or have already acted professionally in several companies in the last years. In total, 222 practitioners from 140 Information Technology companies have answered an on-line (or printed) questionnaire concerning their UML usage experiences, the difficulty in adopting UML, and what should be done to increase the UML usage. In particular, we collected the opinions from practitioners about their practices and perception regarding UML.

II.   RELATED WORK

The literature developed deep discussions on reasons for using UML or not [2][8][10]. However, they did not focus on a particular geographic delimitation. Petre [2] reported an interview-based qualitative investigation involving 50 software engineers in 50 companies over a period of 2 years. The participants were primarily in North America and Europe, but some were from Brazil, India, and Japan. Based on a survey collected from these geographical regions the author found that among the 50 interviewed engineers, 35 reported that they do not use UML. This group reported several reasons for this, including corporate-wide decisions, notation-related issues, and high cost of keeping models synchronized and consistent. This work had an extended version published in [8] which confirmed the previous results of [2]. In a similar manner, the results presented in the study of Gorschek, Temepro, and Angelis [10] pointed out that design models are not used very extensively in industry, and when they are used, the use is informal, with minimal or no tool support, and the notation is not necessarily UML. In addition, they observed that the use of models decreased with an increase in experience and increased with higher level of qualification.

Some studies are optimistic in relation to potential benefits that UML provides during the collaboration and communications between team members [11][12][16]. In [16], the authors collected evidences that UML benefits the collaboration and communication on organizations globally distributed. However, Störrle [12] pointed in his study that there is a possible association between cultural differences and modeling usage, which was considered worth exploring in the future. Finally, the survey conducted by Ho-Quang et al. [11] not target UML practice on industrial closed source projects, but the only the open sources ones.

III.   METHODOLOGY

A. Goal and Research Questions

The Goals (G) of this study are to **(G1)** understand the diffusion and relevance of UML use in the Brazilian companies; and **(G2)** identify improvement points to increase the UML adoption in real-world projects. Based on these goals, the survey aims at addressing three Research Questions (RQ). The two first research questions (RQ 1 and RQ2) address the first goal (G1), while the third question addresses the second goal (G2). The research questions are formulated as follows:

- **RQ1: What is the frequency of UML use in practice?**

- **RQ2: Is UML the "lingua franca" for software modeling?**

- **RQ3: What might improvement points increase the UML use?**

B. Target Population and Data collection

Population selection. According to Kitchenham, the sampling frame of a target population is represented as the finite set of all its members [4]. Thus, a framing population consists of Brazilian software professionals—including developers, software architects, and project managers. These participants represent ones who are in a position to answer the questions and to whom the results of the survey apply [4]. The participants were selected based on two key criteria: the level of theoretical knowledge and practical experience related to software modeling and programming in mainstream projects.

Data collection. In order to enable the data collection, a process of three steps was followed. First, we designed a survey followed well-established guidelines such as [3][4], thereby reducing threats to internal validity. For this, the questionnaire questions concerned on focusing on scrutinizing the research gaps of previous studies; grasp the structures of previously developed questionnaire. In addition, the questionnaire design was based on some of the findings contained on Petre research [2]. In the second step, we concerned in inviting the practitioners. Some practitioners from our industrial contact networks were invited to participate of our research. These participants were from four cities in the southern region of Brazil, including Porto Alegre, São Leopoldo, Canoas, and Novo Hamburgo. Also, working students within six graduate courses at the University of Vale do Rio dos Sinos were invited to respond the questionnaire. Finally, in the last step, data was collected through an on-line questionnaire created by means of Google Docs. We chose this strategy because the questionnaire could be administered quickly, and could also easily collect data from a large number of subjects in geographically diverse locations.

IV.   STUDY RESULTS

A. RQ1: The frequency of UML use

Figure 1 presents the collected data. Among the 213 participants that answered this question in our study, 60 (28.2%) have used UML in their work day, while 73.2% have not. This result reinforces the finding reported by Petre [2], in which the author reports that 35 out of 50 subjects in her study do not use UML in practice. In order to advance knowledge about the use of UML, we investigated whether the use of UML could be influenced by the type of company. For this, we classify the companies of 113 participants as software factory and not software factory. Table 6 summarizes categorical data about the UML use and company type to create a contingency table. For this, we have applied a series of statistics for checking if there is a relationship between these variables, including Chi-Square and Cramer's V. We highlight that a chi-square test assumes that observations are independent of one another and that each observation can be assigned to one and only one category.
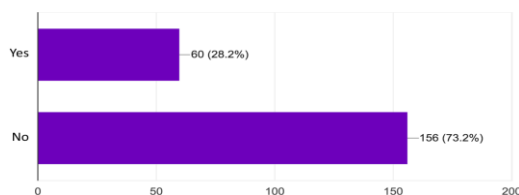

Figure 1. The level of practical experience of participants.

The chi-squared distribution with 1 degree of freedom was 0.115 with p-value = 0.734. Considering the Fisher's exact test, the p-value produced was 0.445. These small p-values (<0.05) indicates that there is not an association between the UML use and company type. Moreover, the Cramr's V was also calculated to eliminate any threat related to statistical conclusion validity.

This test is a measure of association between two nominal variables and varies from 0, indicating no association between the variables, to 1, representing a complete association. The value 1 means that the two variables are equal to each other. The low value of the Cramr's V collected (0.032) also confirmed the previous conclusion.

TABLE I.   2-WAY CONTINGENCY TABLE

	Factory	No Factory	Total
UML use	19	17	36
No UML use	38	39	77
Total	55	56	113
	50.4	49.6	100

Conclusion of RQ1: The UML is not often used in the work life of participants. In addition, no relationship was identified between the use of UML and the participant company being a software factory.

B. RQ2:UML as "lingua franca"

Among the 222 participants, we obtained 217 (about 97.75%) complete questionnaires considering the UML as "lingua franca" in software modeling and model-driven development. Figure 3 presents the obtained data regarding the RQ1. The data indicate that 55.41% of the surveyed did not disagree with the statement that UML is "the lingua franca" in software modeling. On the other hand, 42.34% agreed that UML is not the "de facto standard" in software modeling, indicating the use of other forms for representing design models, such as drawing or sketching freehand shapes in whiteboard. More specifically, we have recorded that 10% strongly agree, 32% agree, 14% are neutral, 23% disagree, and 19% strongly disagree.

Furthermore, when asked whether the UML would be the "lingua franca" in model-driven development projects, the number of participants that did not disagree with this statement increased from 55.41% to 68.02%—a growth of 22.76%. In addition, following an inverse trend, the amount of surveyed that disagreed also decreased from 42.34% to 29.73%—a drop of 29.79%. In particular, the collected questionnaires indicated 11.71% strongly agree with this statement, 36.94% agree, 19.37% are neutral, 21.17% disagree, and 8.56% strongly disagree.
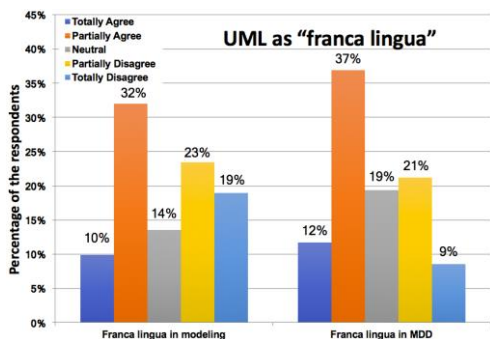

Figure 2. The improvements suggested by participants.

We have learned from previous experimental studies, such as [2][11][13] that there are a set of issues that challenge the effectiveness of UML as a lingua franca. Nevertheless, practitioners have elaborated ad hoc practices that employ UML effectively in reasoning about and communicating about design, both individually and in collaborative dialogues. For this reason, we believe that such practices have favored the tailored use UML use in companies, justifying the high number of participants who agreed that UML is widely used in real-world settings.

Conclusion of RQ2: Although there is no unanimity regarding the adoption of UML as standard modeling language in companies and in MDD projects, the participants see the UML use as broad in companies.

C. RQ3: Critical Improvements for UML tools

Figure 4 presents the answers from 217 participants about four suggested improvement points for UML to promote the UML use. These results do not only reveal some existing gaps that prevent the wider use of UML within project teams in real-world settings, but also can be seen as drivers to make the UML a richer and production-ready modeling language. The results show that the UML use could increase if the creation of diagrams representing global aspects of the system could be done automatically, if there were tools that would automatically create function-oriented diagrams rather than generic diagrams, if a round-trip engineering mechanism were a reality, if automatic update of UML diagrams and source code in response to automatically detected inconsistencies between them were allowed, and if there were tools that could provide effective support for collaborative modeling among distributed teams, allowing developers to be aware of changes that developers are making at runtime, something like Google Docs.

In particular, among 222 participants that suggested improvements to UML use, 136 people (61.26%) reported to find that the automatic creation of UML diagrams to represent a big picture of the system under development would be useful to boost UML use. The number of participants who did not disagree is even greater. In total, 210 (94.59%) do not disagree with the use of a Big Picture view as a mechanism to favor the use of the UML. This result is also found considering the other improvement points, i.e., feature-oriented diagrams, round-trip engineering, and collaborative modeling tools.
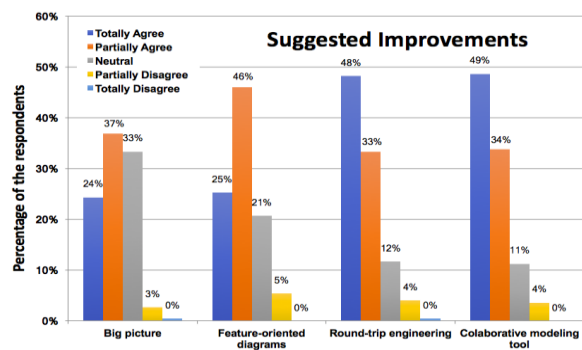

Figure 4. The improvements suggested by participants.

The participants reported the lack of modelling tools, which could support collaboration between teams. For example, this would allow developers to be aware of changes that developers are making at runtime, something that is already supported by

Google Docs. In addition, upcoming modeling tools need to support round-trip engineering for synchronizing related UML diagrams and source code. Given that modeling all structural and behavioral aspects of a software system within a single model is not a trivial task, the UML has proposed a set of diagrams to support a multi-view modeling approach. Thus, different aspects of a system under development are represented by several models, the views. The key challenge identified by the participants is to maintain such different views and their source code consistent and synchronized. According to Chaudron, software development teams use UML during communication and planning of joint implementation efforts [11]. The need for round-trip engineering emerges when project teams need to modify source code and UML diagrams have to be updated to reflect such changes (or vice versa). If UML diagrams are not properly updated, some critical inconsistency may occur.

> ***Conclusion of RQ3:*** *The collected data suggest that creating a "Big Picture" view of the system under development automatically, using functional-oriented diagrams, supporting round-trip engineering, and having more effective collaboration resource would be important improvements to be incorporated into upcoming modeling tools.*

D. Discussion

1) UML usage and Company Issues. Although the study participants believe, for the most part, that the UML is a "franca lingua" in companies, and that they have theoretical knowledge about UML modeling, they do not use UML frequently. This suggests that there may be some culture issue in the companies, or even in relation to the type of development process adopted, that does not favor the wider use of UML.

2) Improviment points for Modeling Tools: We have learned from previous studies, such as [2][8][10][13][15], that UML has been used within teams in the industry for communicating and coordinating their work. Despite this, the results of previous studies, such as [2][8], confirmed in this study, demonstrate that UML has not been widely adopted. The insights from this paper indicate that if the suggested improvement points are implemented by the next modeling tools, UML usage could undergo a positive reversal.

V. CONCLUSIONS AND FUTURE WORKS

This paper reported on a survey aimed at identifying the state-of-the-practice of the Brazilian industry for what concerns the UML usage in world-wide companies. The main results show that: (1) 60 (28.2%) have used UML in their work day, while 73.2% have not; (2) 55.41% of the surveyed did not disagree with the statement that UML is the "lingua franca" in software modeling; (3) 61.26% reported to find that the automatic creation of UML diagrams to represent a big picture of the system under development would be useful to boost UML use. Moreover, we might also point out some important improvements to be incorporated into upcoming modeling tools, including the creation of a "Big Picture" view of the system under development automatically, the support to functional-oriented diagrams and round-trip engineering, and more effective collaboration. The results of this research reinforced some evidences already found on state-of-the-art literature about UML in practice, specifically concerning the UML use, which is

barely applied on software projects. In overall, great part of participants know about UML, but they had not used UML in their projects. This work is an initial effort of our research agenda to explore and analyze the UML adoption in industry. Future works will concentrate effort to investigate more aspects regarding the practice of UML in industry such as the benefits and issues that hinder the UML use in practice.

REFERENCES

[1] OMG, "UML: Infrastructure specification," version 2.4. https://www.omg.org/spec/UML/2.4.1/Superstructure/PDF, acessed 13 March 2018.

[2] M. Petre, "UML in practice," International Conference on Software Engineering (ICSE' 2013), 2013, pp. 722–731

[3] C. Wohlin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell, A. Wesslen, "Experimentation in Software Engineering," Springer, 2012.

[4] B. Kitchenham, S. Pfleeger, "Personal opinion surveys, Guide to Advanced Empirical Software Engineering," Springer London, 2008, pp. 63–92.

[5] M. Torchiano, F. Tomassetti, F. Ricca, A. Tiso, G. Reggio, "Relevance, benefits, and problems of software modelling and model driven techniques—a survey in the italian industry," Journal of Systems and Software, vol. 86, num. 8, 2013, pp. 2110–2126.

[6] D. Budgen, A. J. Burn, O. P. Brereton, B. A. Kitchenham, R. Pretorius, "Empirical evidence about the UML: a Systematic Literature Review," Software: Practice and Experience, vol. 41, num. 4, 2011, pp. 363–392.

[7] M. R. Chaudron, W. Heijstek, A. Nugroho, "How e ective is UML modeling?," Software & Systems Modeling, vol. 11, num 4, 2012, pp. 571–580.

[8] M. Petre, " 'no shit' or 'oh, shit!': responses to observations on the use of uml in professional practice," Software & Systems Modeling, vol. 13, num. 4, 2014, pp. 1225–1235.

[9] J. Singer, S. E. Sim, T. C. Lethbridge, "Software engineering data collection for field studies," Guide to Advanced Empirical Software Engineering, Springer, 2008, pp. 9–24.

[10] T. Gorschek, E. Tempero, and L. Angelis, "On the use of software design models in software development practice: An empirical investigation," Journal of Systems and Software, vol. 95, 2014, pp. 176-193.

[11] T. Ho-Quang, R. Hebig, G. Robles, M. R. Chaudron, and M. A. Fernandez, "Practices and perceptions of UML use in open source projects," In International Conference on Software Engineering: Software Engineering in Practice Track (ICSE 2017), 2017, pp. 203-212.

[12] H. Störrle, "How are Conceptual Models used in Industrial Software Development?: A Descriptive Survey," In Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE 2017), 2017, pp. 160-169.

[13] G. Scanniello, C. Gravino, M. Genero, J. A. Cruz-Lemus, and G. Tortora, "On the impact of UML analysis models on source-code comprehensibility and modifiability," ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 23, num. 2, 2014, pp. 1-13.

[14] W. Dzidek, E. Arisholm, L. C. Briand, "A realistic empirical evaluation of the costs and benefits of UML in software maintenance," IEEE Transactions on Software Engineering, *34*(3), pp.407-432.

[15] C. F. J. Lange, M. R. V. Chaudron and J. Muskens. "In practice: UML software architecture and design description," IEEE Software, vol. 23, no. 2, pp. 40-46, March-April 2006.

[16] A.M. Fernández-Sáez, M.R. V. Chaudron and M. Genero. "An industrial case study on the use of UML in software maintenance and its perceived benefits and hurdles". *Empirical Software Engineering*, pp. 1-65, 2018.