

Variability Resolution and Product Configuration with SMarty: An Experimental Study on UML Class Diagrams

¹Anderson S. Marcolino, ¹Edson Oliveira Jr, ¹Itana M.S. Gimenes and ²Ellen F. Barbosa

¹Informatics Department, State University of Maringá, Avenida Colombo, 5790, Maringá-PR, Brazil

²Maths Science Institute and Computing, University of São Paulo, Avenida do Trabalhador São-Carlense, 400, São Carlos-SP, Brazil

Article history

Received: 04-05-2017

Revised: 30-06-2017

Accepted: 07-07-2017

Corresponding Author:
Edson Oliveira Jr
Informatics Department, State
University of Maringá,
Avenida Colombo, 5790,
Maringá-PR, Brazil
Email: edson@din.uem.br

Abstract: Variability management is one of the most important activities during software product line development and evolution. Current literature presents several approaches for variability management, especially based on UML, such as, PLUS and SMarty. A systematic process with guidelines support SMarty. Existing literature for these kind of approaches provides slight experimental evidence of their effectiveness at product configuration. Thus, this is considered fundamental for transferring technology to the industry. This paper provides experimental evidence on the product configuration capability of SMarty by comparing it to PLUS, one of the most cited product-line method in literature. The experimental study provides incipient evidence that SMarty is more effective for resolving variabilities and configuring consistent products at UML class level. Thus, overall obtained results indicated the capability of SMarty at configuring specific products.

Keywords: Class Diagrams, Experimental Study, Software Product Line, UML, Variability Management

Introduction

The Software Product Line (SPL) approach is aimed at establishing common features in a products family. SPL development and evolution encompasses essential activities. The management of variabilities is one of the most important activities for the success of SPLs.

Even though feature interaction and feature analysis have been the traditional approaches to SPL, there are important approaches for variability management (Capilla *et al.*, 2013; Chen *et al.*, 2009; Galster *et al.*, 2013; Thurimella and Bruegge, 2012). Many of them use other methods, besides the feature interaction and analysis, such as the UML diagrams. The support received by the use of UML models may facilitate the understanding of the development team, managers and stakeholders and receive the supports of UML tools and other UML mechanisms, such as Object Constraint Language (OCL), eXtensible Markup Language (XML) export and import files and other facilities. Chen and Babar (2011) present a systematic literature review on the evaluation of approaches for variability management based on 97 reviewed primary studies, 25 (25.7%) are UML-based proposals. In their survey on

variability modeling in industrial practice, Berger *et al.* (2013) point out that 28.6% of the interviewed companies use UML as a modeling notation for representing variability, reported as the second most frequently adopted notation, the first being feature modeling (77.1%). In addition, according to Chen and Babar (2011), UML is a widely adopted standard notation empowered by its profiling mechanism, which allows the specification of SPL artifacts.

In this study, two UML-based variability management approaches were considered: The Stereotype-based Management of Variability (SMarty) approach (Marcolino *et al.*, 2013; 2014a; 2014b; Oliveira Jr *et al.*, 2010a) and the Product Line UML-based Software Engineering (PLUS) method (Gomaa, 2004). SMarty has been proposed and empirically evaluated. It defines the SMartyProfile, which is an UML profile and SMartyProcess, a systematic process. SMarty also provide guidelines to apply its profile to class, use case, sequence and component diagrams. PLUS is a highly cited method, similar to the SMarty approach due to the use of UML models to allow modeling common and variable SPL elements in use cases and classes by means of stereotypes.

The current scenario of SPL indicates that despite the existence of several variability management approaches in the literature, the majority of this kind of approach has not yet been evaluated using rigorous scientific methods (Chen and Babar, 2011). A systematic literature review performed by Ahnassay *et al.* (2014) shows that a substantial amount of SPL empirical evaluations had not been adequately designed or reported. The need for improvement in quality research design and quality reporting, especially for variability management approaches, is only reinforced further with the findings of such a review. Bosch *et al.* (2015) claim that the heterogeneity of notations and tools shows that industry has not yet solved the variability management problem and continues to experiment with solutions and approaches. This makes it difficult to rely on the available evidence.

Motivated by this scenario, two previous studies were conducted to evaluate the SMarty effectiveness. SMarty was evaluated with PLUS, with relation to their effectiveness for representing and identifying variabilities in use case diagrams (Marcolino *et al.*, 2013). Then, SMarty was investigated with the approach by Ziadi *et al.* (2003) with relation to their effectiveness for representing identifying variabilities in sequence diagrams (Marcolino *et al.*, 2014a).

Recently, SMarty was experimentally compared to PLUS on the effectiveness taking into consideration variability identification and representation in class diagrams (Marcolino *et al.*, 2014b). Therefore, in this study, we focused on experimentally compare the effectiveness of SMarty and PLUS with respect to the capability of resolving variabilities and deriving consistent product configurations from class diagrams.

Effectiveness has been adopted in several researches (Abdelnabi *et al.*, 2004; Basili and Selby, 1987; Coteli, 2013; Martinez-Ruiz *et al.*, 2011) as a way of measuring whether a certain task achieved its goals with completeness and/or accuracy for a given domain, taking into account the number of errors made by a user (ISO/IEC, 2004). The work of Basili and Selby (1987) uses effectiveness for detecting faults at comparing testing strategies; participants are asked to detect faults by applying a certain strategy. Reinhartz-Berger and Sturm (2014) count the number of correct and incorrect answers from variability modeling activities using UML. Abdelnabi *et al.* (2004) use effectiveness to identify the number of real defects found by applying a code reading technique to object-oriented frameworks.

In a previous study (Marcolino *et al.*, 2014b), SMarty was empirically compared to PLUS taking into account the capability of identifying and representing variability in UML class diagrams. Although SMarty had slight better results, we decided to carry out a new experiment aiming at evaluating the effectiveness of resolving

variabilities and deriving product configurations on both SMarty and PLUS.

In this study, effectiveness is a way of measuring the number of correctly and incorrectly variability resolution and, consequently, product configuration for a given variability management technique (SMarty or PLUS). We have already successfully applied the effectiveness measure in Geraldi and Oliveira Jr (2017; Giron *et al.*, 2017; Marcolino *et al.*, 2013; 2014a; 2014b).

This paper is structured according to several sections. The second section presents the main concepts of variability management, the PLUS method and the SMarty approach. The third section presents the controlled experiment to evaluate SMarty and PLUS effectiveness at resolving variability and deriving product configurations. The fourth section presents related work. The fifth section presents our conclusions and discusses future work.

Background

This section discusses essential concepts of SPL and variability management, which are essential to understanding the PLUS method and the SMarty approach.

Variability Management

Variability management is a crucial activity in the SPL lifecycle Bragança and Machado (2006; Chen *et al.*, 2009; Galster *et al.*, 2013; Gomaa, 2004; Korherr and List, 2007; Thurimella and Bruegge, 2012; Ziadi *et al.*, 2003), especially for products configuration and derivation. Such activity provides significant benefits, such as: higher reusability of core assets; decreasing of time to market; and Return On Investment (ROI) (Nikollik, 2012; Pohl *et al.*, 2005; Thurimella and Bruegge, 2012).

Variability management generally relies on four concepts (Table 1): Variability, variation point, variant and variant constraints.

The majority of the variability management approaches in the literature do not make it explicit the effectiveness of deriving product configurations, in different artifacts (Chen *et al.*, 2009), in particular UML-based ones. These approaches are based on stereotypes to allow the representation of variabilities. However, the application of such stereotypes are not systematically guided. In addition, industry needs evidence on their effectiveness to have more confidence on their adoption (Catal, 2009).

The SMarty approach (Fiori *et al.*, 2012; Oliveira Jr *et al.*, 2010a) was developed to enhance the process of variability identification and representation. However, it is still lacking providing evidence on its effectiveness. We chose Gomaa's PLUS method (Gomaa, 2004) to compare with SMarty as it is a well-known and cited method in the literature. Although both SMarty and PLUS support use case and class diagrams, in this study we are focused only in class diagrams.

Table 1. Variability management: Main concepts

Concept	Description
Variability (Bosch, 2004)	It is the capability of an artifact to be customized for a given domain.
Variation Point (Pohl <i>et al.</i> , 2005)	It represents a location where variation takes place.
Variant (Pohl <i>et al.</i> , 2005)	Possible element to resolve a variation point.
Variant constraints (Bosch, 2004)	Relationships among variants for resolving a variation point.

Table 2. Stereotypes of the PLUS method (Gomaa, 2004)

Stereotype	Description
<< kernel >>	Used to represent mandatory elements for Use Case and Class diagrams.
<< optional >>	Used to represent optional elements for Use Case diagrams. It represents an element that can be selected or not in a specific product. For Class diagrams is used semantically to indicate mutually exclusive or inclusive elements.
<< alternative >>	Indicates mutually exclusive elements.

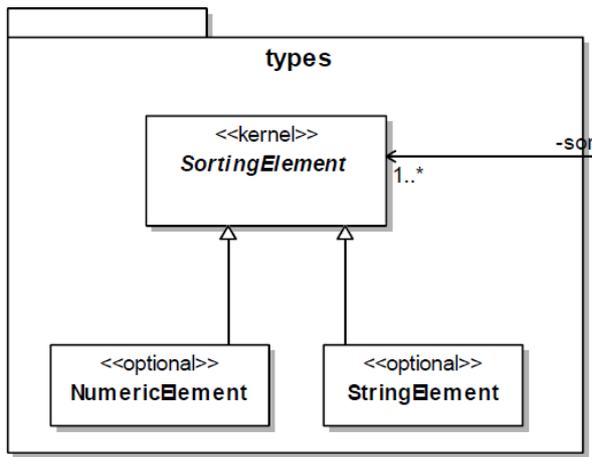


Fig. 1. An SPL feature with PLUS stereotypes

The PLUS Method

PLUS encompasses different activities for SPL development: Requirements, analysis and design. The analysis activity is concerned with static modeling as feature/class dependency modeling.

The PLUS activity of class modeling aims at explicitly modeling commonalities and variabilities. PLUS does not have an UML profile. Although PLUS has no explicit meta attributes and meta classes for modeling variability, it allow the application of stereotypes to tag variation points and variants.

Table 2 presents the stereotypes of PLUS for use case and class diagrams.

Figure 1 shows an example of the application of the PLUS stereotypes and its possible limitation. It exemplifies a sorting algorithm SPL feature. The abstract class *SortingElement* is mandatory tagged with <<kernel>>.

As a rule for variability constraint, one variant specified as optional, may be or not selected in the configuration (Bosch, 2004). However, in the specification of PLUS, the variants, even tagged as

optional for class diagrams, represent an alternative, that is, at least one needs to be selected, as shown in the example of Fig. 1.

As the presented issue is hypothetically defined, experimental evidence is necessary. Furthermore, the consequent problem in understanding the models and checking complementary models may influence the time for generating software configurations and their quality, even considering its probably easier application, due to the reduced number of stereotypes. Two experimental studies were conducted to test these hypotheses: one to evaluate PLUS effectiveness in the process of modeling the variabilities in UML class diagrams, with regard to SMarty; and another to identify the issue on understanding and on generating concise product configurations. Both experiments are described next.

The SMarty Approach

SMarty has an UML profile (SMartyProfile) and a systematic process (SMartyProcess). The profile has several stereotypes to represent variability (Table 3).

SMartyProfile is based on the inter-relationship of the main concepts of SPL with respect to the variability management. These concepts are applied in the elements of interest of the UML metamodel, as observed in Fig. 3. Based on the relationship among the variability management concepts and the UML diagrams, Fig. 3 presents the UML SMartyProfile 5.1.

Figure 2 presents an example on the SMarty stereotypes. It represents a sorting algorithm feature, the same from Fig. 1. The abstract class *SortingElement* is compulsory and represents a variation point.

Besides the motivation to verify the limitations pointed out for PLUS method, the evaluation allows verifying the potential of the profile and process from SMarty approach, as well as its evolution. The next sections present two experimental studies performed to collect evidences in the two perspectives of effectiveness for SMarty in comparison to the PLUS method.

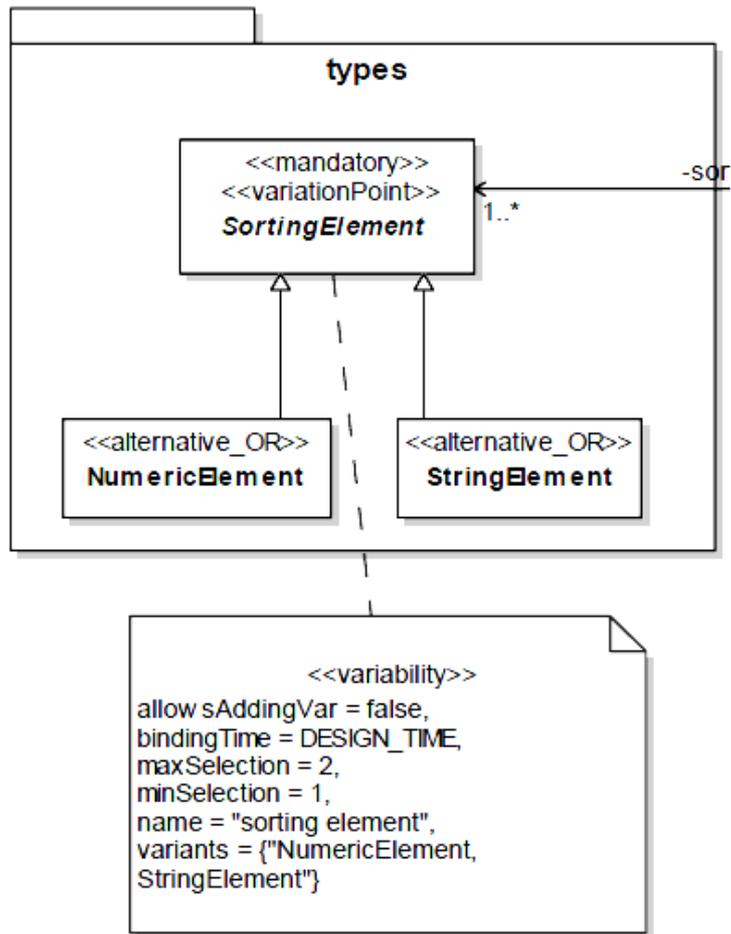


Fig. 2. Example of SPL feature modeled with Smarty

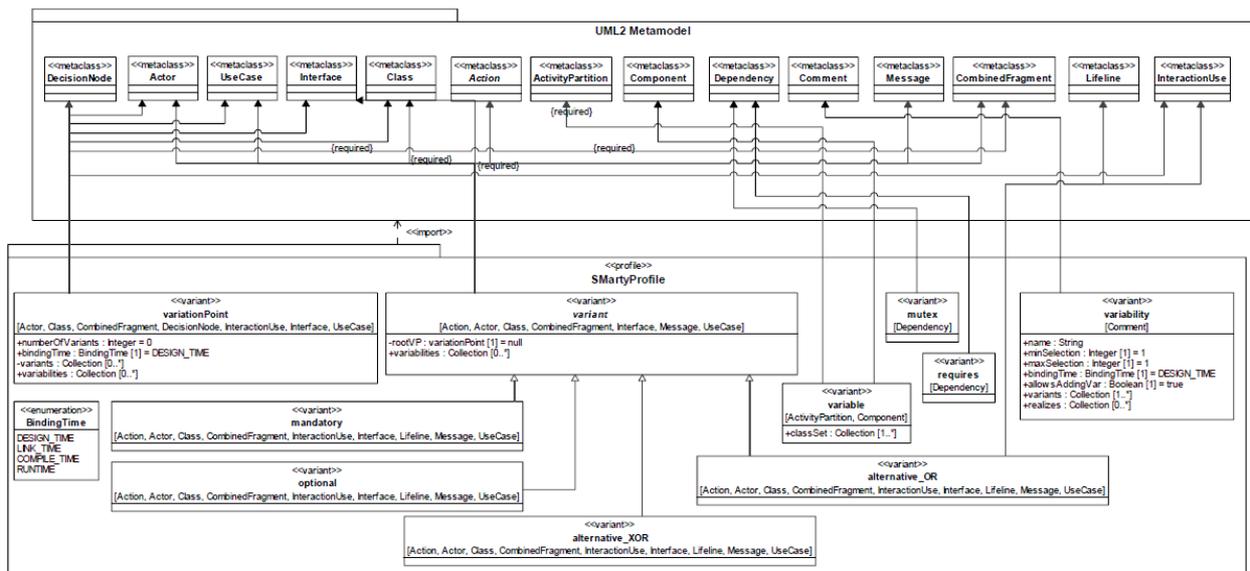


Fig. 3. The SmartyProfile 5.1 OliveiraJr *et al.* (2010a)

Table 3. Overview of the SMartyProfile stereotypes (OliveiraJr *et al.*, 2010a)

Stereotype	Description
<< variability >>	Extends the metaclass Comment.
<< mandatory >>	A compulsory variant.
<< optional >>	A variant that should be selected or not.
<< alternative_OR >>	A variant part of a group of inclusive variants.
<< alternative_XOR >>	A variant part of a group of exclusive variants.
<< mutex >>	A mutually exclusive relationship between two variants.
<< requires >>	A require relationship between two variants.
<< variable >>	Extends the metaclass Component.

Effectiveness to Resolving Variabilities and Configuring Products

This section presents an experimental study with regard to the effectiveness of SMarty and PLUS with relation to product configuration activities from UML classes with modeled variability.

This study is a quasi-experiment because we could not randomize the participants selection due to a small population. Next, we describe this study phases.

Definition

The goal of this experiment was to compare PLUS and SMarty, for the purpose of identifying which is more effective, with respect to resolving variabilities and deriving consistent product class diagram configurations, from the point of view of product line architects, in the context of master and Ph.D. students from the State University of Maringá and the University of São Paulo (USP) involved in software engineering researches.

The following research questions were defined for this study:

- R.Q.S.1: Which methodology is more effective at resolving and deriving consistent product configurations from class diagrams: SMarty or PLUS?
- R.Q.S.2: Which approach/method requires the largest number of consultations to SPL descriptions to provide an accurate understanding of variabilities to deriving a specific product configuration: SMarty or PLUS?
- R.Q.S.3: What is the influence of the prior participant SPL and variability knowledge in the application of the method or approach for deriving consistent product configurations from UML class diagrams?

Planning

The local context took into consideration the E-Commerce and AGM SPLs aiming at the resolution and configuration of specific products from UML class diagrams. For the training session, specific exercises were applied to simulate the study

execution, making the participants familiar with the activities they should perform.

The relevance of the study instrumentation led us to evaluate it previously in a pilot project. During the pilot project, we realized that each participant should handle only one approach due to the time spent if we considered both approaches. In addition, for the selection of participants, the same premise was taken into consideration.

In this study, instrumentation was composed of: The free consent term to the experimental study; a questionnaire for characterization of participants; the training material; SPLs descriptions; the UML class diagrams from each SPL; and an experimental form, in which the participant signed the configurations that they preferred and indicated if it was verified the SPLs descriptions to provide additional understanding of the classes. Participants were split in two groups, according to the characterization questionnaire. Fatigue was the main reason for such a splitting, as realized during the pilot project.

Hypotheses Formulation: we formulated and tested the following hypotheses:

- Null Hypothesis (H_0): Both X and Y approaches are equally effective in terms of resolving and generating specific product configurations from UML class diagrams:

$$H_0 : \mu(\text{effectiveness}(X)) = \mu(\text{effectiveness}(Y))$$

- Alternative Hypothesis (H_1): X approach is less effective than Y approach:

$$H_1 : \mu(\text{effectiveness}(X)) < \mu(\text{effectiveness}(Y))$$

- Alternative Hypothesis (H_2): X approach is more effective than Y approach:

$$H_2 : \mu(\text{effectiveness}(X)) > \mu(\text{effectiveness}(Y))$$

Effectiveness Marcolino *et al.* (2014b) is the dependent variable taking into consideration the studies

from Martinez-Ruiz *et al.* (2011; Coteli, 2013; Basili and Selby, 1987). We calculates effectiveness as follows:

$$effectiveness(z) = nVarC - nVarI$$

Where:

- z = A given variability management approach
- $nVarC$ = The number of correct identified and modeled variability elements (variation point or variant) according to the z approach
- $nVarI$ = The number of incorrect identified and modeled variability elements (variation point or variant) according to the z approach. Both false positive and false negative incorrect variabilities are considered to compose this number

Independent variables are: The variability management approach - a factor with two treatments X (PLUS) and Y (SMarty); and the SPL - a factor with two treatments E-Commerce and AGM.

Figure 4 illustrates dependent and independent variables (factors).

We could not randomize the selection of participants since the population of volunteers was quite restricted. Thus, random capacity was done at the assignment of the approach (X or Y) to each participant.

Balancing was done as tasks of the experiment were assigned in equal number to participants.

With regard to review mechanism, we collected evidence to test the hypotheses with: descriptive statistics, normality tests (both Shapiro-Wilk and Mann-Whitney-Wilcoxon test) and Spearman for correlation between the participant’s knowledge in SPL/variability and the effectiveness of each treatment (approach).

Execution

Twenty-four software engineering participants were involved in our study: 12 master students and 12 Ph.D. candidates.

Amongst instrumentation handled by participants, the main object to collect data was the experimental form, in which the configurations of the products was selected based on UML class diagrams of each SPL. For each configuration, one selection box was assigned with “yes”, if the configuration should be included, or “no”, if it should not be marked.

As a “configuration” we considered each class annotated with PLUS (X) or Smarty (Y) in the respective SPL class diagram (E-Commerce or AGM), intercalated in an equal number.

An additional selection box was included for the indication of a possible consultation in the SPL description/specification, when a class model was not enough to allow a participant to resolving a variability, thus selecting a configuration. This answer was taken in consideration to R.Q.S.3.

The participation procedures were as follows: A participant resolves and derives two specific products, one based on the E-Commerce SPL and other based on the AGM SPL; and both products are represented in UML classes diagrams modeled with X or Y approaches, randomly distributed to participants in an equal number.

Analysis and Interpretation

Collected results from deriving products are summarized in Table 4, which shows the calculated effectiveness, the knowledge level and the number of consultations to SPL specification documents. Table 5 summarizes this experiment data analysis.

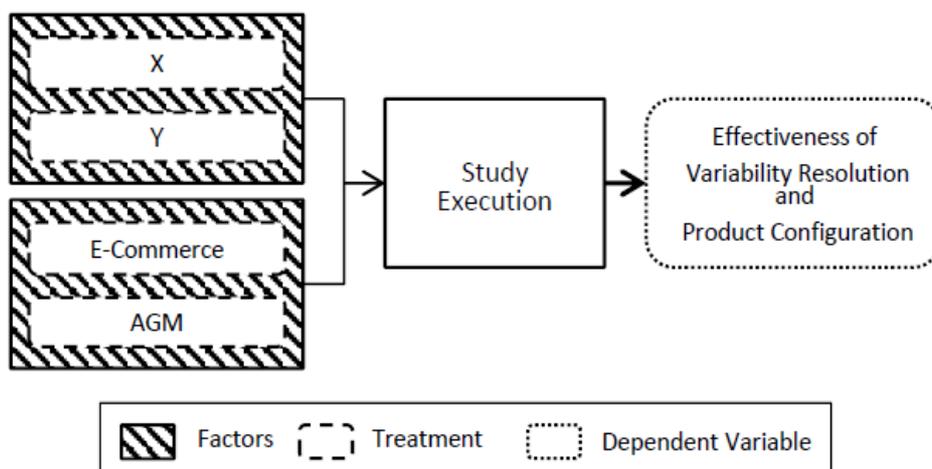


Fig. 4. Dependent and independent variables of the experiment

Table 4. Collected data and descriptive statistics

X Approach (PLUS Method)						Y Approach (SMarty 5.1)					
Participant #	Correct Identified Variability Elements	Incorrect Identified Variability Elements	Eff.	Knowl. Level	Number of Checks in SPL Spec.	Participant #	Correct Identified Variability Elements	Incorrect Identified Variability Elements	Eff.	Knowl. Level	Number of Checks in SPL Spec.
1	14	0	14	5	0	1	10	4	6	5	0
2	12	2	10	4	14	2	14	0	14	4	6
3	13	1	12	3	0	3	10	2	8	3	2
4	14	0	14	3	14	4	14	0	14	3	0
5	11	3	8	3	0	5	14	0	14	3	0
6	10	4	6	2	6	6	14	0	14	2	0
7	14	0	14	2	1	7	9	5	4	2	7
8	14	0	14	2	0	8	10	4	6	2	7
9	10	4	6	2	10	9	14	0	14	2	4
10	13	1	12	2	0	10	13	1	12	2	2
11	9	5	4	2	0	11	7	7	0	2	0
12	11	3	8	2	0	12	14	0	14	2	1
13	14	0	14	2	0	13	11	3	8	2	0
14	12	2	10	2	7	14	14	0	14	2	12
15	13	1	12	2	2	15	14	0	14	2	0
16	14	0	14	2	0	16	10	4	6	2	0
17	13	1	12	2	14	17	13	1	12	2	0
18	12	2	10	2	9	18	14	0	14	2	0
19	14	0	14	2	10	19	10	4	6	2	0
20	14	0	14	2	3	20	12	2	10	2	0
21	9	5	4	2	8	21	14	0	14	2	0
22	13	1	12	1	2	22	14	0	14	1	0
23	13	1	12	1	0	23	14	0	14	1	0
24	9	5	4	1	2	24	13	1	12	1	0
Mean	12.29	1.71	10.58	2.21	4.25	Mean	12.33	1.58	10.75	2.21	1.71
Std. Dev.	1.74	1.57	3.13	0.98	5.71	Std. Dev.	2.07	2.02	4.07	0.87	3.12
Median	13.00	1.00	12.00	2.50	0.50	Median	13.50	0.50	13.00	2.00	0.00

Approaches Effectiveness (R.Q.S.1)

Normality Tests

We applied Shapiro-Wilk to test normality of the e-Commerce and AGM SPLs samples. As the results for the normality test indicated for both samples a non-normal result, X approach (N = 24) and Y approach (N = 24), a nonparametric test was conducted for the sample to identify which approaches have a significant effectiveness in the interpretations leading to a correct instantiation of products.

Mann-Whitney-Wilcoxon Test

The values identified in the test demonstrate that does not exist a significant difference between the samples. The *p* was calculated to be compared with the level of significance of 95% ($\alpha = 0.05$), to confirm the result. The calculated value for *p* was 0.533 and in comparison $p = 0.533 > \alpha = 0.05$ confirm that the null hypothesis (H_0) must be accepted. Therefore, there are no statistical difference between the median of the effectiveness in relation to the capacity of interpretation and generation of correct product configurations based in SPLs designed in UML class diagrams with the PLUS method or with the SMarty 5.1 approach.

To evaluate the results of the calculated effectiveness, the number of checks in the SPLs specification documents for the creation of the products were considered for analysis. These checks consequently influence the effectiveness for the approaches. Besides this, the Spearman correlation test was calculated to analyze the possible influence of the participants' previous knowledge, which may also influence the effectiveness.

Comparison of the Quantity of Checks in the SPL Description to Support the Understanding of Classes (R.Q.S.2)

The descriptive statistic of the quantity of classes in which additional information was checked by the participants and induced to a correct selection, increasing the calculated effectiveness, is present in Table 4.

These values correspond to each class assigned as "Yes" and led the increasing of the effectiveness. They were considered for analysis because, with consults to the description of the product to verify their meaning which, according to the interpretation of the participant it was not clear in the class model, it was possible a correct selection of them. Thus, the consult resulted in a greater effectiveness for the approach, which classes were consulted.

Table 5. Summary of the effectiveness of X and Y approaches and the spearman's correlation

Element	Class	
	PLUS (X)	SMarty 5.0 (Y)
Selection of participants	20 master and Ph.D. students. N(X) = 10	N(Y) = 10
Effectiveness (mean)	23	13.6
Shapiro-Wilk normality test	p = 0.0062 (0.0062 < 0.05)	p = 0.5286 (0.5286 > 0.05)
Statistical test	Mann-Whitney-Wilcoxon p = 0.020 (0.020 < 0.05)	
Result	PLUS is more effective.	
Correlation	p = 0.39 Weak positive correlation for E-Commerce. p = 0.44 Weak positive correlation for AGM.	p = 0.23 Weak positive correlation for E-Commerce. p = 0.37 Weak positive correlation for AGM.

Normality Test

We also applied the Shapiro-Wilk normality test for the sample of the number of checks in the SPL description which led to a hit and, therefore, the increasing of the effectiveness for their respective approach (Table 4). The results are presented below as follows:

Total of Checks for Approach X (N = 24)

The normality test indicated, for a mean (N) of size 24 with 95% of significance level ($\alpha = 0.05$), $p = 0.0001$ ($0.0001 < 0.05$) and a calculated value of $W = 0.7888 < W = 0.9160$, i.e., the sample was non-normal.

Total of Checks for Approach Y (N = 24)

The normality test indicated, for a mean (N) of size 24 with 95% of significance level ($\alpha = 0.05$), $p = 0.000001$ ($0.000001 < 0.05$) and a calculated value of $W = 0.6187 < W = 0.9160$, i.e., the sample was non-normal.

Test Mann-Whitney-Wilcoxon for the Sample of Checks in the Descriptions of the SPLs for X and Y Approaches

The values analyzed through the test presented a statistical difference. The value of p , to be compared with the significance level of 95% ($\alpha = 0.05$), was 0.048, that is $p = 0.048 < \alpha = 0.05$. Therefore, the null hypothesis (H_0) was reject, proving that the mean of checks for each approach influence the number of hits and, consequently, the calculated value for effectiveness.

The lower the number of checks is, the greater is the support of the approach considered. Approach X presents a mean of 4.25 (total of 102) and approach Y presents a mean of 1.71 (total of 41). Thus, the number of checks indicates that, by the interpretation of the SPLs class diagrams, there is support enough to guarantee a better understanding of the variabilities represented by SMarty. This result comprises with the issue presented in the second section.

Despite the reduced number of stereotypes to be applied, it is necessary additional checks to generate software products configurations with PLUS. On the other hand, the set of guidelines gives an additional

support to interpret all the variabilities and elements graphically represented on the SPLs class diagrams, reducing the need of additional documents.

Besides the results obtained, the Spearman's correlation was calculated to verify if there is any influence in the effectiveness values with regard to the participant's previous knowledge that, summed up with the number of checks, may have influence in the effectiveness final value.

Correlation among the Effectiveness of the Approaches and the Participants Variability Characterization (R.Q.S.3)

Spearman's Correlation

The following values were calculated and applied in the Spearman's scale (Fig. 5):

- X Approach: $\rho = 0.4956$ - weak positive correlation
- Y Approach: $\rho = -0.1015$ - weak negative correlation

Analyzing results obtained for the Spearman's correlation, we observed that X approach had a weak positive correlation. In other words, it was more influenced by the knowledge level of the participants in comparison with the results of Y approach, which presented a weak negative correlation.

An issue that may attenuated the influence of the knowledge of SMarty is the existence of SMartyProfile. It supports both the approach application as the interpretation of the different elements of the UML diagrams that it encompasses. This result is in agreement with a key approach for aspects such as time learning and applying new techniques.

The need of training and the cost (in terms of time) for the adoption of new approaches in industry is an issue that influences the adoption of new technologies. Thus, SMarty, with its guidelines, allows an easier adoption due to the set of elements it supports. By means of its UML profile, which facilitates its use in UML modeling tools, or by its set of guidelines, which facilitates the implementation and understanding of the elements represented, SMarty is able to generating products that are more concise.

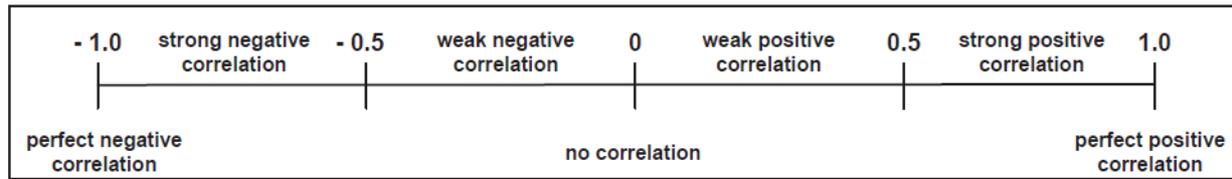


Fig. 5. Spearman correlation scale (Spearman, 1987)

Table 6. Summary of results

Effectiveness (R.Q.S.1)		Checks in SPL's Descriptions (R.Q.S.2)		Knowledge Level and Effectiveness Correlation (R.Q.S.3)	
PLUS	SMarty 5.1	PLUS	SMarty 5.1	PLUS	SMarty 5.1
Shapiro-Wilk Sample non-normal	Normality Test Sample non-normal	Shapiro-Wilk Sample non-normal	Normality Test Sample non-normal	No normality test because the knowledge level is represented by an ordinal scale.	
Mean 10.58	10.75	Mean 4.25	1.71	Not applied.	
Mann-Whitney-Wilcoxon Test p = 0.533 >= 0.05		Mann-Whitney-Wilcoxon Test p = 0.048 <= 0.05		Spearman's Correlation p(Corr.1) = 0.4956 p(Corr.2) = -0.1015	
Accept the Null Hypothesis (H ₀)		Reject the Null Hypothesis (H ₀)		Weak positive correlation.	Weak negative correlation.

The general results to answer the research questions of Study#2, as well as the statistical tests applied are summarized in Table 6.

Validity Evaluation

The validity evaluation for this experimental study is discussed next.

Threats to Conclusion Validity

We dealt with a major concern with regard to the sample size (N = 24). We are aware of this and we will mitigate such threat by trying to increasing sample size in prospective studies. As we could not apply the random capacity to participants selection, we could not infer the generalization of the results. We will treat this limitation in prospective study replications.

Threats to Construct Validity

Effectiveness is calculated based on the ability of the participants in modeling variability by considering the X and Y approaches and the hits and errors in the selection of the configurations, for this study. We guaranteed the independent variable variability modeling approach by running a pilot project before the experiment execution.

Threats to Internal Validity

We addressed the following issues:

- Differences among participants: As we took into consideration a medium sample of participants (N = 24), we tried to reduce variations in their skills by realizing a training session with tasks in the same

order. Participants had nearly same level of experience in variability concepts and UML modeling. We verified this with the application of questions with no tasks

- Fatigue effects: Our study lasts for 80 min on average, thus we understand that fatigue could not affect the results. We took this in consideration from the pilot project results and other previous similar experiments carried out by our group (Marcolino *et al.*, 2013; 2014b), which evidenced a non-fatigue effect for the participants
- Influence among participants: Although we could not really control this factor, a human observer supervised the experiment tasks. Thus, we believe that this issue did not affect the internal validity of this study
- Other effects: We implemented different training sessions in similar way, thus, trying to minimize any biases

Threats to External Validity

We detected two main threats:

- Instrumentation: We did not use real class diagrams, as the E-Commerce and the AGM are not commercial SPLs. More experimental studies must be conducted adopting real SPLs, mainly developed by the industry
- Participants: Software Engineering master students and Ph.D. candidates were selected. Although most students are not practitioners, we can benefit from

students to perform experiments as largely and well discussed by Carver *et al.* (2003; Falessi *et al.*, 2017; Höst *et al.*, 2000; Salman *et al.*, 2015)

Related Work

A systematic literature review was carried out based on search engines, such as IEEE, ACM and others (Fig. 6), in order to investigate the existing experimental studies, which provide evidence on the effectiveness of variability management approaches.

The systematic review found 178 studies, from which 28 were duplicated, remaining 150. From these studies, only two were considered relevant and selected for fully reading: (i) “Empirical Validation of Complexity and Extensibility Metrics for Software Product Line Architectures” (Oliveira Jr *et al.*, 2010b); and (ii) “Assessing the Influence of Stereotypes on the Comprehension of UML Sequence Diagrams - A Family of Experiments” (Genero *et al.*, 2008).

Besides the results of the systematic review two studies realized experimental evaluations in evaluation of architecture of software product line Gonzalez-Huerta *et al.* (2015) and for a UML-based approach to modeling variabilities, such as SMarty and PLUS (Reinhartz-Berger and Sturm, 2014).

Gonzalez-Huerta *et al.* (2015), after an analysis of literature identified a low number of empirical validations of evaluation methods for software architectures. Such a low number of empirical studies, quantitative and qualitative comparisons with existing methods has been neglected and the low number of studies replications. With these facts, they propose the evaluation study for their model-driven approach Quality

Driven Architecture Derivation and Improvement (QuADAI). QuADAI approach applies architectural transformations to a product line architecture derived from the architecture of an SPL to ensure desired quality attributes for a product.

The objective of the experiment was to compare the effectiveness, efficiency, perceived ease of use, perceived usefulness and intention to use with regard to participants using the product evaluation and transformation activities of QuADAI as opposed to the Architecture Tradeoff Analysis Method (ATAM). Participants measure an architecture to check the fulfillment of the non-functional requirements with the two methods. Such as our study, which is part of a family of experimental studies, Gonzalez-Huerta *et al.* (2015) presented a new study and their family of experimental studies consolidate in a meta-analysis. The experimental analysis indicated that participants produced their best results when applying QuADAI and the meta-analysis aggregate the results obtained in the individual experiments.

Reinhartz-Berger and Sturm (2014), such as Gonzales and in our study, identify a lack in experimental studies to verify the comprehensibility of UML profiles for performing application engineering tasks. They refer the comprehension of tasks that assess the participants ability to use the knowledge represented in the schema, where participants are requested to determine whether and how certain information is available from the schema. Results were evidenced in different UML diagrams, i.e., use case, class and sequence diagrams and for reuse-related aspects.

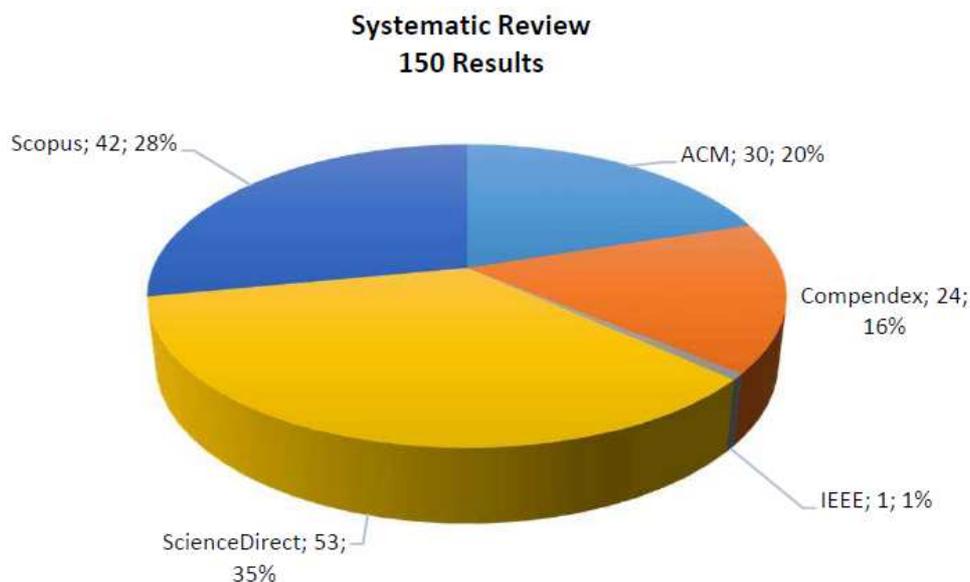


Fig. 6. Systematic review results

The studies retrieved from the systematic review were useful to provide some indications to the experimental evaluation described herein. However, we highlight that none of them is directly related to variability management approaches. In recent study (Reinhartz-Berger and Sturm, 2014), it was identified the evaluation of comprehensibility of UML-based variability representations, however the study only evaluates the approach based in answers of a questionnaire. In the present study, it was proposed the configuration of products from the UML-based SPL diagrams, comparing the effectiveness of two approaches: SMarty and PLUS. Finally, such as identified in the related studies, there is a need of experimental evaluations in the area, which also motivated our work.

Conclusion and Future Work

We provided experimental evaluation of the SMarty effectiveness as compared to the PLUS method for UML class diagrams at product configuration. Results show that PLUS has more effectiveness than SMarty for UML class diagrams. This might be related to the simplicity of PLUS as it has only a couple of stereotypes. SMarty has a larger set of stereotypes, thus, requires more discipline; however, it gives more support for the identification and representation of variabilities (Marcolino *et al.*, 2014a), which leads to a non-ambiguous specification. This is important to generate valid SPL products, evidenced with this experimental study. In addition, SMarty is supported by a UML profile fully compliant with most UML tools.

This study provided data about the SMarty capability to be understood and generating software product configurations, in a concise way, as compared to the PLUS method. The latter is easier to apply but generates ambiguous interpretation of its elements. If the configurations are inconsistent, the derivation of specific products will present lower quality of the products and consequently will decrease the benefits in adoption of SPLs methodology. In this study, there is no statistical significance of the values of effectiveness. However, in relation to the number of additional consults to SPL descriptions, it was identified statistically that PLUS needs more consults than SMarty, influencing the calculated effectiveness.

The experiment results cannot be generalized due to its already discussed limitations. Thus, new experiments must be carried out using real SPLs.

Variability management is essential for the success in the adoption of SPL and, even the results cannot be generalized, their initial evidence could be used to conduct new experiments (by the use of laboratory packages), consolidate new approaches and provide initial evidence of the used approaches to conduct new and more deep investigations in academic or industrial set.

This paper contributes as it provides initial evidence data for the effectiveness of the SMarty and makes it explicit how to plan and conduct experimental evaluations with regard to the effectiveness of variability management approaches.

Acknowledgement

The authors thank CAPES a Brazilian Funding Agency for providing Anderson S. Marcolino a two-year Master's degree scholarship and Master's students and Ph.D. candidates from the State University of Maringá and the University of São Paulo for attending and providing us feedback in our study.

Author's Contributions

Anderson S. Marcolino: Coordinated the experiment at DIN/UEM and ICMC/USP. Performed design, execution, analysis and interpretation and sharing of the experiment. Wrote the majority of the manuscript.

Edson Oliveira Jr: Coordinated the experiment at DIN/UEM. Contributed to the experiment planning, design and execution, as well as data analysis and dissemination. Reviewed manuscript writing and proofreading.

Ellen F. Barbosa: Coordinated the experiment at ICMC/USP and contributed with data analysis. Contributed to technically review the paper.

Itana M.S. Gimenes: Coordinated the research project (PROCAD) at UEM to which the experiment was carried out. Contributed to technically review the paper.

Ethics

Participation during pilot and experiment trials are voluntary and participants are made known that their feedbacks will be contributing to a non-profit research project.

References

- Abdelnabi, Z., G. Cantone, M. Ciolkowski and H.D. Rombach, 2004. Comparing code reading techniques applied to object-oriented software frameworks with regard to effectiveness and defect detection rate. Proceedings of the International Symposium on Empirical Software Engineering, Aug. 20-20, IEEE Xplore Press, Redondo Beach, CA, USA, pp: 239-248.
DOI: 10.1109/ISESE.2004.1334911
- Ahnassay, A., E. Bagheri and D. Gasevic, 2014. Empirical evaluation in software product line engineering. Tech. Rep. TR-LS3-130084R4T, Ryerson University, Laboratory for Systems, Software and Semantics.

- Basili, V. and R. Selby, 1987. Comparing the effectiveness of software testing strategies. *IEEE Trans. Software Eng.*, SE-13: 1278-1296. DOI: 10.1109/TSE.1987.232881
- Berger, T., R. Rublack, D. Nair, J.M. Atlee and M. Becker *et al.*, 2013. A survey of variability modeling in industrial practice. *Proceedings of the 7th International Workshop on Variability Modelling of Software-Intensive Systems*, Jan. 23-25, ACM, New York, pp: 1-8. DOI: 10.1145/2430502.2430513
- Bosch, J., 2004. Preface. *Proceedings of the Groningen Workshop on Software Variability Management: Software Product Families and Populations*, (PFP' 04), pp: 1-2.
- Bosch, J., R. Capilla and R. Hilliard, 2015. Trends in systems and software variability [Guest editors' introduction]. *IEEE Software*, 32: 44-51. DOI: 10.1109/MS.2015.74
- Bragança, A. and R.J. Machado, 2006. Extending UML 2.0 metamodel for complementary usages of the/spl Lt/extend/spl Gt/relationship within use case variability specification. *Proceedings of the 10th International Software Product Line Conference*, Aug. 21-24, IEEE Xplore Press, Baltimore, MD, USA, pp: 123-130. DOI: 10.1109/SPLINE.2006.1691584
- Capilla, R., J. Bosch and K. Kang, 2013. *Systems and Software Variability Management: Concepts, Tools and Experiences*. 1st Edn., Springer, New York, ISBN-10: 3642365833, pp: 317.
- Carver, J., L. Jaccheri, S. Morasca and F. Shull, 2003. Issues in using students in empirical studies in software engineering education. *Proceedings of the 9th International Symposium on Software Metrics*, Sept. 5-5, IEEE Xplore Press, Sydney, NSW, Australia, pp: 239-249. DOI: 10.1109/METRIC.2003.1232471
- Catal, C., 2009. Barriers to the adoption of software product line engineering. *ACM SIGSOFT Software Eng. Notes*, 34: 1-4. DOI: 10.1145/1640162.1640164
- Coteli, M.B., 2013. Testing effectiveness and effort in software product lines a thesis. MSc Thesis, Middle East Technical University.
- Chen, L. and M.A. Babar, 2011. A systematic review of evaluation of variability management approaches in software product lines. *Inform. Software Technol.*, 53: 344-362. DOI: 10.1016/j.infsof.2010.12.006
- Chen, L., M. Ali Babar and N. Ali, 2009. Variability management in software product lines: A systematic review. *Proceedings of the 13th International Software Product Line Conference*, Aug. 24-28, Carnegie Mellon University Pittsburgh, USA, pp: 81-90.
- Falessi, D., N. Juristo, C. Wohlin, B. Turhan and J. Münch *et al.*, 2017. Empirical software engineering experts on the use of students and professionals in experiments. *Empirical Software Eng.* DOI: 10.1007/s10664-017-9523-3
- Fiori, D.R., I.M.S. Gimenes, J.C. Maldonado and E. Oliveira Jr, 2012. Variability management in software product line activity diagrams. *Proceedings of the International Conference on Distributed Multimedia Systems*, (DMS' 12), Knowledge Systems Graduate School, Skokie, IL, pp: 89-94.
- Galster, M., D. Weyns, D. Tofan, B. Michalik and P. Avgeriou, 2013. Variability in software systems-a systematic literature review. *IEEE Trans. Software Eng.*, 99: 282-306. DOI: 10.1109/TSE.2013.56
- Genero, M., J.A. Cruz-Lemus, D. Caivano, S. Abrahao and E. Insfran *et al.*, 2008. Assessing the influence of stereotypes on the comprehension of UML sequence diagrams: A controlled experiment. *Proceedings of the 11th International Conference on Model Driven Engineering Languages and Systems*, Sept. 28-Oct. 3, Springer, Toulouse, France, pp: 280-294. DOI 10.1007/978-3-540-87875-9_20
- Geraldi, R.T. and E. Oliveira Jr, 2017. Towards initial evidence of SMartyCheck for defect detection on product-line use case and class diagrams. *J. Software*, 12: 379-392. DOI: 10.17706/jsw.12.5.379-392
- Giron, A.A., I.M.S. Gimenes and E. Oliveira Jr, 2017. Case study of test case generation based on metamodel for model transformations. *J. Software*, 12: 364-378. DOI: 10.17706/jsw.12.5.364-378
- Gomaa, H., 2004. *Designing Software Product Lines with UML: From Use Cases to Pattern-Based Software Architectures*. 1st Edn., Addison-Wesley Professional, ISBN-10: 0201775956, pp: 736.
- Gonzalez-Huerta, J., E. Insfran and S. Abrahão and G. Scanniello, 2015. Validating a model-driven software architecture evaluation and improvement method: A family of experiments. *Inform. Software Technol.*, 57: 405-429. DOI: 10.1016/j.infsof.2014.05.018
- Höst, M., B. Regnell and C. Wohlin, 2000. Using students as subjects-a comparative study of students and professionals in lead-time impact assessment. *Empirical Software Eng.*, 5: 201-214. DOI: 10.1023/A:1026586415054
- ISO/IEC, 2004. *Software engineering-product quality-Part 4: Quality in use metrics*. ISO/IEC.
- Korherr, B. and B. List, 2007. A UML 2 profile for variability models and their dependency to business processes. *Proceedings of the 18th International Workshop on Database and Expert Systems Applications*, Sept. 3-7, IEEE Xplore Press, Regensburg, Germany, pp: 829-834. DOI 10.1109/DEXA.2007.96

- Marcolino, A., E. Oliveira Jr, I.M.S. Gimenes and J.C. Maldonado, 2013. Towards the effectiveness of a variability management approach at use case level. *Proceeding of the International Conferences on Software Engineering and Knowledge Engineering, (SEKE' 13)*, pp: 214-219.
- Marcolino, A., E. Oliveira Jr and I.M. Gimenes, 2014a. Variability identification and representation in software product line UML sequence diagrams: Proposal and empirical study. *Proceedings of the Brazilian Symposium on Software Engineering, Sept. 28-Oct. 3, IEEE Xplore Press, Maceió, Brazil*, pp: 141-150. DOI: 10.1109/SBES.2014.11
- Marcolino, A.S., E. Oliveira Jr, I.M.S. Gimenes and E.F. Barbosa, 2014b. Empirically based evolution of a variability management approach at UML class level. *Proceedings of the 38th Annual International Computers, Software and Applications Conference, Jul. 21-25, IEEE Xplore Press, Vasteras, Sweden*, pp: 354-363. DOI: 10.1109/COMPSAC.2014.58
- Martinez-Ruiz, T., F. Garcia, M. Piattini and J. Munch, 2011. Modelling software process variability: An empirical study. *IET Software*, 5: 172-187. DOI: 10.1049/iet-sen.2010.0020
- Nikolik, B., 2012. Software quality assurance economics. *Inform. Software Technol.*, 54: 1229-1238. DOI: 10.1016/j.infsof.2012.06.003
- Oliveira Jr, E., I.M.S. Gimenes and J.C. Maldonado, 2010a. Systematic management of variability in UML-based software product lines. *J. Univ. Comput. Sci.*, 16: 2374-2393. DOI: 10.3217/jucs-016-17-2374
- Oliveira Jr, E., J. Maldonado and I. Gimenes, 2010b. Empirical validation of complexity and extensibility metrics for software product line architectures. *Proceedings of the 4th Brazilian Symposium on Software Components, Architectures and Reuse, Sept. 27-29, IEEE Xplore Press, Brazil*, pp: 31-40. DOI: 10.1109/SBCARS.2010.13
- Pohl, K., G. Bockle and F. Linden, 2005. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer-Verlag, Berlin, ISBN-10: 3540243720, pp: 467.
- Reinhartz-Berger, I. and A. Sturm, 2014. Comprehensibility of UML-based software product line specifications. *Empirical Software Eng.*, 19: 678-713. DOI: 10.1007/s10664-012-9234-8
- Salman, I., A.T. Misirli and N. Juristo, 2015. Are students representatives of professionals in software engineering experiments? *Proceedings of the 37th IEEE International Conference on Software Engineering, May 16-24, IEEE Xplore Press, Florence, Italy*, pp: 666-676. DOI: 10.1109/ICSE.2015.82
- Spearman, C., 1987. The proof and measurement of association between two things. *Am. J. Psychol.*, 100: 441-471. DOI: 10.2307/1422689
- Thurimella, A.K. and B. Bruegge, 2012. Issue-based variability management. *Inform. Software Technol.*, 54: 933-950. DOI: 10.1016/j.infsof.2012.02.005
- Ziadi, T., L. Helouet and J.M. Jezequel, 2003. Towards a UML profile for software product lines. *Proceedings of the 5th International Workshop on Software Product-Family Engineering, Nov. 4-6, Springer, Siena, Italy*, pp: 129-139. DOI: 10.1007/978-3-540-24667-1_10