Resumo:

Software Product Lines (SPL) are a well known solution to create reusable software products. Usually, a set of product variants are analyzed and commonalities and variabilities are extracted to create the product line. However, this process is still not well defined. Some authors argue that a set of guidelines is needed to formalize this process. The main objective of this work is to propose a process that assembles a feature retrieval process for SPL re-engineering. The process main goal is to assemble different feature retrieval processes for each scenario, aiming for a high level of reuse. The main goal of the assembled process is to detect and extract features from a set of product variants, covering the first two phases of the SPL re-engineering process. This process intends to provide enough flexibility regarding artifacts, strategies and techniques used for feature retrieval. This flexibility is given by allowing the choice and combination of different techniques for feature retrieval based on information gathered during the process' execution. This information includes: team experiences and skills, domain engineering artifacts, requirements artifacts, product artifact types and extensions, and technologies used to develop the products. A set of guidelines was created to help whoever is performing this process to choose the techniques that better fit their situation. To create this work, a set of 70 studies was analyzed focusing on two topics of interest: combination of strategies and techniques for feature extraction, and adaptability/flexibility/customization of the proposal to attend different scenarios. After the end of this analysis, a process for feature retrieval was created. To obtain an early feedback about the process a survey was applied in specialists of the SPL and re-engineering fields. The survey contained questions about the contribution, relevance, originality and reliability of the proposed process. The survey applied in specialist of the area resulted in a positive feedback as a great majority of the participants stated that this proposal may be a relevant contribution to the field. The participants, however, also pointed that some improvements should be made regarding the flexibility/adaptability of the process. After applying some changes to improve the aspects of flexibility and reliability of the proposal, a quasi-experiment was performed. The subjects were eight students of software engineering with some development experience. These subjects were divided in four groups according to their knowledge and skills. Two teams would apply this process while the other two would apply a different approach. The measures of interest here
were the effort (time) to perform the processes and the precision (similarity) of the retrieved features. The quasi-experiment results showed that the teams that applied this process got a reduced effort and a higher precision than the teams that performed a different approach. For future work, this process will be extended to cover another phases of the SPL re-engineering process. Also, a case study in a real development environment will be planned and executed to give this proposal more reliability.

**Palavras-chave:** Software Product Line; Software Re-engineering; Software Process

**Modalidade de Participação:** Iniciação Científica
PAEPSPL: A FEATURE RETRIEVAL RE-ENGINEERING PROCESS

1. INTRODUCTION

As it happened in the automobile industry, the introduction of product lines in software production caused big changes to the software industry. The benefits and motivation for implementing a Software Product Lines (SPL) are many: reduced maintenance effort, reduced complexity, increased reusability and better cost estimative (POHL et al., 2005). While many systems are now being developed proactively as a software product line, there are still those that emerge when a company has one or more software products and wants to create a software product line out of them.

One way to create a SPL is the software product line re-engineering process. The goal of this process is to take a number of product variants and transform them into a SPL with the use of techniques, methods and tools. This process can be divided in three main phases: detection, analysis and transformation (ASSUNÇÃO et al., 2017). During the detection phase, the variability points and commonalities of the products are identified and extracted through the use of feature retrieval techniques. The second phase, analysis, is where the discovered features are organized as a feature model. The last phase, transformation, is when artifacts linked with these features are managed and modified in order to create the SPL.

However, this process is still not well defined. Otsuka et al. (2011) and Ziadi et al. (2012) argue that a set of guidelines is needed to formalize this process. Other authors, such as those of Martinez et al. (2015) and Stoermer; O'brien (2001), have pointed that guidelines may lead to an automated support for this process.

The main objective of this work is to propose a process that assembles a feature retrieval process for SPL re-engineering. The process main goal is to assemble different feature retrieval processes for each scenario, aiming for a high level of reuse. The main goal of the assembled process is to detect and extract features from a set of product variants, covering the first two phases of the SPL re-engineering process.

This process intends to provide enough flexibility regarding artifacts, strategies and techniques used for feature retrieval. This flexibility is given by allowing the choice and combination of different techniques for feature retrieval based on information gathered during the process' execution. This information includes: team experiences and skills, domain engineering artifacts, requirements artifacts, product artifact types and extensions, and technologies used to develop the products. A set of guidelines was created to help whoever is performing this process to choose the techniques that better fit their situation.

2. METHODOLOGY

The first step to conduct this work was to search in digital libraries for another software re-engineering process. A Systematic Mapping Study (SMS) on SPL re-engineering process (ASSUNÇÃO et al., 2017) was found, and it was used as a guideline to construct this work. In the SMS, Assunção et al. (2017) mapped 119 studies, these studies were selected as the universe of studies for this work. However, in order to reduce this large set of studies into a more precise one, the following inclusion criteria (IC) were applied. IC01) The study must address the
detection or analysis phase of the re-engineering process; IC02) The study must present the use of two or more techniques for feature retrieval; IC03) The study must present some kind of customization/adaptability in their proposal. To be included in this work the studies should answer “yes” to all the ICs. After finishing this process, 70 studies were selected to be analyzed.

The studies analysis was done focusing on two topics of interest: combination of strategies and techniques for feature extraction, and adaptability/flexibility/customization of the proposal to attend different scenarios. After the end of this analysis, a process for feature retrieval was created.

To obtain an early feedback about the process a survey was applied in specialists of the SPL and re-engineering fields. The survey contained questions about the contribution, relevance, originality and reliability of the proposed process.

After applying some changes to improve the aspects of flexibility and reliability of the proposal, a quasi-experiment was performed. The subjects were eight students of software engineering with some development experience. These subjects were divided in four groups according to their knowledge and skills. Two teams would apply this process while the other two would apply a different approach. The measures of interest here were the effort (time) to perform the processes and the precision (similarity) of the retrieved features.

3. RESULTS and DISCUSSION

The Prepare, Assemble and Execute (PAEPSPL), illustrated in Figure 1, is a process that assembles and executes a feature retrieval process for SPL re-engineering. The process was design to allow adaptation according to different scenarios and re-applicability. This may reduce the cost and effort of re-applying this process in different or similar scenarios. During Prepare, data such as team experience, skills, and products documentation are analyzed. This information is used to choose most recommended retrieval techniques for that scenario. During the second phase, Assemble, these techniques are assembled in different activities of a generic process, see Figure 2, generating an assembled process. For instance, Formal Concept Analysis (FCA) and Latent Semantic Indexing (LSI) can be assembled as extraction techniques to retrieve the features and create the Feature Model, as showed in Figure 3. And finally, during Execute the assembled process is executed to retrieve features of the product variants.
To help the selection of retrieval techniques, a guidelines section was created in the process documentation. These guidelines described the techniques mapped from the studies analyzed, give examples of use, list input and outputs artifacts used for each technique, point to related techniques and tools, and most important, describes recommended situations, as Table 1 illustrates using the example of clustering technique.

Table 1: Guideline for the clustering technique

<table>
<thead>
<tr>
<th>Technique</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>Group elements based on their dependences</td>
</tr>
<tr>
<td>Inputs</td>
<td>Source Code</td>
</tr>
<tr>
<td>Outputs</td>
<td>Feature tree; Feature clusters;</td>
</tr>
<tr>
<td>Tools</td>
<td>Cluster 3.0; C Clustering Library</td>
</tr>
<tr>
<td>Related Techniques</td>
<td>Formal Concept Analysis</td>
</tr>
<tr>
<td>Recommended Situations</td>
<td>Products with high level of dependencies between feature implementations.</td>
</tr>
</tbody>
</table>

The techniques were divided in three groups: i) **Static analysis** - Dependency Analysis and its variations, Rule-Based Techniques, Data-Flow analysis and its variations, and Clustering; ii) **Information retrieval** - LSI, Vector Space Model (VSM) and FCA; iii) **Support strategies** - Expert Driven Extraction and Heuristics.

Regarding the Heuristics techniques, a set of heuristics proposed by different works in the SPL re-engineering area were mapped:

**H1. Filtering**: Filter elements that do not contribute directly to the feature retrieved.
H2. **Score Modification**: Rank elements based on their lexical and syntactical properties, eliminating not relevant elements.

H3. **Compare**: Calculates the similarity degree for each pair of input model elements.

H4. **Match**: Use empirical similarity thresholds and analyze a pair of model elements, returning pairs that are considered similar in a match result.

These heuristics were included in the process documentation because they can be combined with any other feature retrieval technique, giving the process more flexibility.

The survey applied in specialist of the area resulted in a positive feedback as a great majority of the participants stated that this proposal may be a relevant contribution to the field. The participants, however, also pointed that some improvements should be made regarding the flexibility/adaptability of the process.

After the process received this improvements the quasi-experiment was performed. The quasi-experiment results showed that the teams that applied this process got a reduced effort and a higher precision than the teams that performed a different approach.

Both empirical evaluations demonstrated that this process is a contribution that reduces the effort of applying the SPL re-engineering process by providing guidance to assemble a feature retrieval process according to a specific scenario. This guidance reduces the effort and the process obtained a higher level of precision of retrieved features.

### 4. FINAL CONSIDERATIONS

This work presented a process that assembles a feature retrieval process for SPL re-engineering. This kind of process is needed because its adaptability and reusability may reduce the effort of applying and re-applying the feature retrieval when constructing a SPL. The proposal was created with the information gathered when analyzing a set of proposals of the re-engineering field mapped by Assunção et al. (2017). The result was a generic process that collects information about the domain, the team that will perform the re-engineering, and product artifacts. This information is used to choose the most recommended techniques for that scenario and assembles them into a feature retrieval process. This feature retrieval process is executed and the features are retrieved from the products.

To evaluate this proposal a survey was applied in specialist of the area. The survey responses indicated that this proposal may be a relevant contribution to the area. The survey also pointed some needed improvements regarding the flexibility of the process. After these improvements were made, a quasi-experiment was performed. The quasi-experiment results showed that this process reduces the effort and increases the precision of feature retrieval in comparison to another approach.

For future work, this process will be extended to cover another phases of the SPL re-engineering process. Also, a case study in a real development environment will be planned and executed to give this proposal more reliability.

### 5. REFERENCES

MARTINEZ, J., ZIADI, T., BISSYANDÉ, T. F., KLEIN, J., & LE TRAON, Y. Bottom-up adoption of software product lines: a generic and extensible approach. In:


