Abstract. Within the realm of software product line engineering, feature modeling is one of the widely used techniques for modeling commonality as well as variability. Feature models incorporate the entire domain application configuration space, and are therefore developed collectively by teams of domain experts. In large scale industrial domains, feature models become too complex both in terms of maintenance and configuration. In order to make the maintenance and configuration of feature models feasible, we propose to modularize feature models based on the well established Distributed Description Logics formalism. Modular feature models provide for an enhanced collaborative/distributed feature model design, more efficient feature model evolution and better reusability of feature model structure. We also develop methods for the configuration and configuration verification of a modular feature model based on standard inference mechanisms. We describe and evaluate our proposed approach through a case study on an online electronic store application domain.

Keywords: Feature models, Variability Modeling, Software Product Lines, Modular Software Design

1. Introduction

A key aspect of software product line engineering is capturing the common characteristics of a set of software-intensive applications in a specific problem domain [18]. Product line engineering allows for the rapid development of variants of a domain specific application by using a common set of reusable assets often known as core assets. Such an approach supports the management of commonality as well as variability in the software development process [10, 11]. Feature modeling is an important conceptual tool that offers modeling support for software product lines. It provides for addressing commonality and variability both formally and graphically, describing interdependencies of the product family attributes (features) and expressing the permissible variants and configurations of the product family.

Feature modeling has received great attention both in industry and academia since its inception in the early 90’s by the Software Engineering Institute. Its concepts were developed within the Feature Oriented Domain Analysis (FODA) methodology [12]. In FODA, the distinguishing factors of the product family members are abstracted away until no difference between the members of the given problem domain exists. Other methods have been since developed on this basis such as Organization Domain Modeling (ODM) [14], FeatuRSEB [15], Feature-Oriented Reuse Method (FORM) [13] and cardinality-based feature modeling [10].

Different success stories of the use of product families and related techniques have been reported in the literature encompassing fields other than Computer Science such as the auto-manufacturing and cellular phone design industries. As an example, Clements and Northrop [1] reported that Nokia was able to increase its production capacity for new cellular phone models from 5-10 to around 30 per year. The main challenge that Nokia faced was that as the market demand and customer taste changed rapidly, different products with varying specifications and user interfaces needed to be developed to keep up with the requests. The use of product family engineering made it possible to quickly and efficiently create the required software and hardware variations for the different products and use variability to customize the newly configured cellular phones. More recently, Microsoft has also become interested in software product families. It has incorporated the feature modeling process at the front end of its Software Factory life cycle to provide support for software product family engineering, which has been applied to its Health Level 7 (HL7) [2] software factory initiative. This initiative employs feature modeling to make the rapid design and development of HL7 conformant applications easier and more accessible to health informatics practitioners.

In complex and large domains, due to the complexity of the problems, feature models are often collaboratively developed by different teams. For instance, one of the scenarios for collaboration between the teams is that each team focuses on one set of conceptually-related features (e.g., one team focuses on developing all features related to the payment aspects of an online store, while another team focuses on developing the features related to shipping the sold items, etc.), then all the features developed by the different teams are merged together forming a comprehensive feature model. The other approach is that each team focuses on one specific set of goals of the product family and develops one feature model by identifying all of its required features regardless of their conceptual relevance (e.g., one team focuses on developing a feature model for easily deployable and cheap to install online stores, another team on a feature model for an online store that ensures highly secure and safe transactions, etc.). Once these feature models are developed, they can be put together to form one representative unique feature model. Therefore as it can be seen in these two scenarios, the required collaborative process for feature model development makes the maintenance of these models difficult to perform, which requires suitable representation mechanisms to store and integrate multiple feature models.

Furthermore, in light of the large-scale industrial applications of product families, the desire to create an overarching feature model for a product family entails the development of very large feature models that need to be customized before they can be used for a specific application in a given domain. In order to develop an instance of a product family from the relevant feature model, its most desirable features need to be selected from the feasible configuration space of the product family, whose entirety may not be of interest for a specific application at hand. The selection of the best set of features for a product would be based on the strategic goals, requirements and limitations of the stakeholders, as well as the integrity constraints of the feature model. Once the desired features are specified, the feature model can be customized such that it includes the wanted features and excludes the non-relevant ones. A final fully-specific feature model with no points for further customization is called a configuration of the feature model based on the selected features. In many cases, a configuration is gradually developed in several stages.
In each stage, a subset of the preferred features are selected and finalized and the unnecessary features are discarded yielding a new feature model whose set of possible configurations are a subset of the original feature model. This feature model is referred to as a specialization of the former feature model and the staged refinement process constitutes staged configuration [3].

The adopted representation mechanism that enhances feature model maintenance should also provide means for efficient specialization, configuration and validation of feature models.

1.1. Problem Description

As mentioned earlier, there are several challenges towards the maintenance and specialization of large-scale feature models. In this paper, we are specifically looking at addressing the following problem statements:

1. Given a complex target domain with a large feature space that demands the involvement of multiple experts, what would be the most appropriate feature model representation mechanism such that expert collaboration, and feature model maintenance is facilitated. For example, in the context of the two mentioned feature model development collaboration scenarios, what would be an acceptable approach for the representation of feature models in such a way that experts can efficiently perform their tasks individually and also share feature model information collectively;

2. The specialization of feature models is performed based on the requirements of the stakeholders and the target application.

   In many cases, the stakeholders only specify a limited number of required features and expect a complete configuration of the feature model accordingly. Since the configuration needs to be developed based on feature interactions, feature model integrity constraints and the requested stakeholder requirements, what would be an appropriate method in the context of the developed representation discussed above for finding a suitable feature model configuration that satisfies the stakeholders.

The first issue is related to the maintenance of large-scale feature models, while the latter is concerned with their manipulation and specialization.

1.2. Approach Overview

In order to tackle to the two mentioned challenges, we take a holistic approach where both problems are viewed in a unified framework. Fundamentally, we adopt Description Logics [4] as the underlying formalism for representing feature models, and strategize in this context such that appropriate techniques are developed for our purpose. Due to their advantages discussed below, modularity and concept covering lay at the heart of our framework:

1. Modularity is an engineering technique that builds larger systems by combining smaller components. A modular design allows for the rise of important design attributes such as the increase in extendability, reusability, parallel and collaborative design progression and others. Such an approach in feature model development extends our ability for decomposing the problem domain into smaller subdomains that can be analyzed and modeled simultaneously, while also providing for the unification of these subdomains later. To support for collaborative feature model design through modularity, we customize the Distributed Description Logic framework to represent modular feature models;

2. Concept covering is a computational minimization process that evaluates whether a certain conceptual structure wraps around another, or how extended the structure should be to be able to do that. We provide a correspondence between the concept covering problem in Description Logics [5] and feature model configuration and show how an acceptable configuration can be developed based on the feature model integrity constraints and the stakeholder requirements. We further demonstrate how concept covering can be beneficial in pinpointing to the stakeholder requests that are hampering the development of a suitable feature model configuration, when one cannot be developed.

In summary, we are interested in looking at how large and complex feature models can be maintained through an appropriate representation mechanism, and furthermore how these large feature models can be configured. For the first issue, we explore the possibility of feature model modularization in Description Logics, in the context of which we will benefit from existing concept covering methods to create appropriate configurations for large-scale feature models.

The remainder of this paper is organized as follows: Section 2 provides an overview of the main concepts of feature modeling along with an introduction to the sample online electronic store feature model. The representation of feature modeling notations in Description Logics is discussed in Section 3. The modular representation of feature models in Distributed Description Logics is shown in Section 4, followed by the formalization of the feature model configuration process through concept covering. In Section 6, two case studies are presented. Section 7 provides some discussions. In Section 8 elaboration on the related work is provided. Finally, the paper is concluded in Section 9.

2. Feature Modeling

Features are important distinguishing aspects, qualities, or characteristics of a family of systems [6]. They are widely used for depicting the shared structure and behavior of a set of similar systems. To form a product family, all the various features of a set of similar/related systems are composed into a feature model. A feature model is a means for representing the possible
configuration space of all the products of a system product family in terms of its features. Feature models can be represented both formally and graphically; however, the graphical notation depicted through a tree structure is more favored due to its visual appeal and easier understanding. More specifically, graphical feature models are in the form of a tree-like structure whose root node represents a domain concept, e.g., a domain application, and the other nodes and leafs illustrate the features. In this context, a feature is a concept property related to a functional or non-functional requirement, e.g., domain application task, modeled in a way to capture commonalities or possibly differentiate among product family variants.

2.1. Overview of Feature Models

In a feature model, features are hierarchically organized and can be typically classified as:

- **Mandatory**, the feature must be included in the description of its parent feature;
- **Optional**, the feature may or may not be included in its parent description given the situation;
- **Alternative feature group**, one and only one of the features from the feature group can be included in the parent description;
- **Or feature group**, one or more features from the feature group can be included in the description of the parent feature.

Figure 1 depicts the graphical notation of the feature relationships. The tree structure of feature models falls short at fully representing the complete set of mutual interdependencies of features; therefore, additional constraints are often added to feature models and are referred to as **Integrity Constraints (IC)**. The two most widely used integrity constraints are:

- **Includes**, the presence of a given feature (set of features) requires the **existence** of another feature (set of features);
- **Excludes**, the presence of a given feature (set of features) requires the **elimination** of another feature (set of features).

In addition to this information, several feature modeling notations such as the cardinality-based feature models [7] provide means for expressing the plausible cardinality of the features.

2.2. The Online Electronic Store Feature Model

The online store is a commonly used example in the feature modeling literature, which is employed to represent the behavior of feature modeling frameworks [8]. Figure 2 depicts a small feature model designed to depict some of the aspects of such an online store. The sample electronic store consists of three mandatory features, namely: Payment, Tax, and Password, and one optional feature: Shipping. Furthermore, the Shipping feature has been represented by an or-group of sub-features, while an alternative-group of sub-features constitutes the Tax feature. It is easy to see that with this design, various configurations for an electronic store can be developed. For instance, one store may decide to use airmail to ship its goods, while another can only provide pickup service to its customers.

Clearly, not all possible configurations of the features of e-store produce valid electronic store programs. For instance, an automatic fraud detection feature would not be so helpful while the payments are made in the form of cash. Therefore, not all possible configurations of the feature model are valid and some restrictions on the acceptable feature configurations exists. Such restrictions are expressed in the form of feature model integrity constraints. Several examples of such integrity constraints for the online electronic store feature model are: 1) cash payment includes pickup shipping; 2) credit card payment includes automatic fraud detection; 3) never changing the password includes the inclusion of special characters in the password; and 4) manual fraud detection excludes credit card payment; 5) e-delivery includes credit card payment; 6) credit card payment includes special character passwords; 7) pickup excludes universal tax rate.

Now based on the structure of the feature model and its integrity constraints, valid feature model configurations can be developed and validated.
3. Feature Model Representation in Description Logics

Our proposed formal framework borrows from Semantic Web languages and particularly Description Logics [4]. Description Logics (DLs) are a family of logic formalisms for knowledge representation, whose basic syntax elements are concept names, role names, and individuals. Intuitively, concepts stand for sets of objects in the domain, and roles link objects in different concepts. Individuals are used for special named elements belonging to concepts. Formally, concepts are interpreted as subsets of a domain of interpretation \( \Delta \), and roles as binary relations, which are subsets of \( \Delta \times \Delta \).

DL formulas (concepts) are given semantics by defining the interpretation function over each construct. For example, given two generic concepts such as \( C \) and \( D \), their conjunction \( C \cap D \) is interpreted as set intersection: \( (C \cap D)_I = C_I \cap D_I \), and also the other boolean connectives are given the usual set-theoretic interpretation, whenever present. The interpretation of constructs involving quantification on roles needs to make domain elements explicit. Concepts can be used in inclusion assertions \( C \sqsubseteq D \), and definitions \( C \equiv D \), which pose restrictions on possible interpretations according to the knowledge elicited for a given domain. The semantics of inclusions and definitions is based on set containment: an interpretation \( I \) satisfies an inclusion \( C \sqsubseteq D \) if \( C_I \subseteq D_I \), and satisfies a definition \( C \equiv D \) when \( C_I = D_I \). A DL theory, also known as TBox or ontology, is a set of inclusion assertions and definitions. A model of a TBox \( T \) is an interpretation satisfying all inclusions and definitions of \( T \). It is possible to introduce a variety of constructors in DL languages, making them more expressive. However, it is well known that this may lead to an explosion in computational complexity of inference mechanisms. Hence, a trade-off is necessary. In this paper, we refer to \( \text{ALCH} \) (the logic \( \text{AL} \) plus complex concept negation and role hierarchy).

Given the syntax and semantics of \( \text{ALCH} \), we are able to model the structure and integrity constraints of feature models in the \( \text{ALCH} \) Description Logic, hence it is expressive enough for our purpose. In the following similar to [8], we show how the structural knowledge, the integrity constraints and the stakeholder requests can be represented in the selected DL language.

3.1. Structural Knowledge

The structural knowledge base, denoted \( SKB \), consists of the description of the hierarchical feature relationships of a feature model. The fundamental feature hierarchy introduced in Section 2 is expressed in this knowledge base. As base constructs, we define \( \text{Feature} \) as the parent of all available features, and also let \( \text{hasfeature} \) be the core role that relates parent features with their child features (It is important to note that in the following formalization \( \text{Feature} \) and \( \text{hasFeature} \) are actual constructs in the formalism metamodel whereas \( \text{Child} \), \( \text{Child}i \) and \( \text{Parent} \) are placeholders for feature names that would be replaced at...
MODULAR FEATURE MODELS: REPRESENTATION AND CONFIGURATION

the metamodel instantiation phase. This can be seen in actual examples later).

\[ \text{Feature} \sqsubseteq \top \]

\[ \text{hasFeature} \sqsubseteq \text{ObjectProperty} \]

**Mandatory** a mandatory feature can be viewed as a prerequisite for the existence of its parent feature;

\[ \text{Child} \sqsubseteq \text{Feature} \]

\[ \text{Parent} \sqsubseteq \text{Feature} \]

\[ \text{Parent} \sqsubseteq \exists \text{hasFeature.Child} \]

In case multiple features are mandatory, they can be formulated by

\[ \text{Child}_i \sqsubseteq \text{Feature}, \text{ for } 1 \leq i \leq n \]

\[ \text{Parent} \sqsubseteq \exists \text{hasFeature.Child}_i \]

which means that all of the child features should be present in the configuration of the parent feature.

**Optional** optional features are free to appear or not in the formation of the parent feature; hence, both states of being false or true are acceptable and therefore, does not require their direct connection with their parent features;

\[ \text{Child} \sqsubseteq \text{Feature} \]

however, it should be noted that due to the nature of optional features and the fact that they may not be present in the configuration of their parent feature, they often appear in conjunction with other sibling features.

**Or feature group** an or feature group allows the feature parent to become configured with the appearance of at least one of its child features:

\[ \text{Child}_i \sqsubseteq \text{Feature}, \text{ for } 1 \leq i \leq n \]

\[ \text{Parent} \sqsubseteq \exists \text{hasFeature.Child}_i \]

\[ \text{Parent} \sqsubseteq (\exists \text{hasFeature.Child}_1) \sqcup \ldots \sqcup (\exists \text{hasFeature.Child}_n) \]

This formulation shows that the elements of an alternative feature group cannot co-exist in the configuration of their parent feature.

More complex clauses can be formulated based on these primitive relationships. As an example, the shipping feature shown in Figure 2 can be represented using an or feature group:

\[ \text{Shipping} \sqsubseteq \text{Feature}, \text{Pickup} \sqsubseteq \text{Feature}, \text{E} - \text{delivery} \sqsubseteq \text{Feature} \]

\[ \text{Airmail} \sqsubseteq \text{Feature}, \text{RegularMail} \sqsubseteq \text{Feature} \]

\[ \text{Shipping} \sqsubseteq (\exists \text{hasFeature.Pickup}) \sqcup \ldots \sqcup (\exists \text{hasFeature.RegularMail}) \]

As another example, the fraud detection feature can be modeled as:

\[ \text{FraudDetection} \sqsubseteq \text{Feature}, \text{Manual} \sqsubseteq \text{Feature}, \text{Automatic} \sqsubseteq \text{Feature}, \text{Manual} \sqsubseteq \neg \text{Automatic} \]

\[ \text{FraudDetection} \sqsubseteq (\exists \text{hasFeature.Manual}) \sqcup (\exists \text{hasFeature.Automatic}) \]

Finally, for the feature model root in addition to its feature relationships, we define the root as:

\[ \text{root} \sqsubseteq \exists \text{hasFeature.Feature} \]

For instance, the E-Store feature model root is represented as:

\[ \text{E} - \text{Store} \sqsubseteq \exists \text{hasFeature.Feature} \]

\[ \text{E} - \text{Store} \sqsubseteq \exists \text{hasFeature.Tax} \sqcap \exists \text{hasFeature.Password} \sqcap \exists \text{hasFeature.Payment} \]

The information represented in this form are referred to as structural knowledge of the feature model and will be stored and referred to as \( \text{SKB} \).
3.2. Integrity Constraints

Feature interdependencies that cannot be captured in structured hierarchical format are represented within the Integrity Constraints base (IC).

Includes the presence of a given feature could entail the automatic inclusion (the need for presence) of one or more other features, which is represented as:

\[ \exists \text{hasFeature.BaseFeature} \sqsubseteq \exists \text{hasFeature.IncludedFeature} \]

For instance, cash payment includes pickup shipping can be represented as:

\[ \exists \text{hasFeature.Cash} \sqsubseteq \exists \text{hasFeature.Pickup} \]

Excludes the inclusion of one feature may result in the exclusion of one or more features, which is shown by:

\[ \exists \text{hasFeature.BaseFeature} \sqsubseteq \neg (\exists \text{hasFeature.ExcludedFeature}) \]

For instance, manual fraud detection excludes credit card payment would be depicted as:

\[ \exists \text{hasFeature.Manual} \sqsubseteq \neg (\exists \text{hasFeature.CreditCard}) \]

All integrity constraints will be stored and represented through the integrity constraint base (IC).

3.3. Stakeholder Requests

Stakeholders and product developers can specify a set of basic features that they would like to see in the final product. For instance, in the e-store feature model configuration process, they may require the inclusion or exclusion of certain features. Such requirements and requests are referred to as the stakeholders’ hard constraints. The satisfaction of hard constraints is either feasible or not, which makes the configuration process based on hard constraints a crisp one. Hard constraint can be represented by complex concepts in Description Logic format. For instance, a request by the stakeholders to have both airmail delivery and credit card payment features available can be represented by a concept description defined as: \( \text{Airmail} \sqcap \text{CreditCard} \). We represent the stakeholders’ hard constraints through \( SR_h \).

Summing up with regards to the knowledge domain modeling we have the following formalization:

- \( SKB \): The feature model structural knowledge, i.e. the definition of features and the feature model hierarchical tree structure, is represented using DL concepts and role relations modeled as described in Section 3.1.
- \( IC \): Integrity constraints, i.e. the includes and excludes relations between features, are modeled using subsumption assertions as described in Section 3.2.
- \( SR_h \): Stakeholder’s hard constraints, i.e. features that are requested by the stakeholders to be present in final configuration, are represented using complex concept descriptions explained above.

4. Modularity for Feature Models

In this section, we first show how the collaboration scenarios for developing feature models can produce multiple feature models and hence require a modular design. Based on this, the DDL framework is customized to represent modular feature models.

4.1. Collaborative Feature Model Design

As was discussed earlier, large domains require the collaboration of several teams of modeling and domain experts to fully cover the wide range of available features. The collaborative process structure can be defined based on how the feature model is to be developed. In the following we introduce two sample collaboration scenarios:

1. It is possible that in a complex domain, some of the required applications have already been analyzed and legacy software applications exist. For this reason, feature models can be more easily developed for the segments of the target domain which already possess a running application or a designed conceptual model. In case several applications of that product family exist, teams of experts can be assembled to focus on each one of these existing applications in order to develop the corresponding feature models for these applications. Later, it is possible to integrate each of the developed individual feature models by unifying their common features and incorporating their differences into one representative feature model. Figure 3 shows this case where multiple existing fragmented feature models for the domain of e-store have been developed by different teams of experts. As it can be seen in this figure, each individual feature model can encompass all of the aspects of the domain; however, it may not be complete in all of the aspects.
2. The second possible form of collaboration is shown in Figure 4. In this approach, the target domain is divided into non-overlapping subdomains, based on which different teams of experts are assembled such that each of the teams focuses on one specific subdomain of interest. The outcome of the work of each team would be a complete subdomain analysis and its corresponding feature model design. As shown in Figure 4, the individual feature models do not have any overlap with each other. The integration of these feature models will entail a complete feature representation for the domain.

In both collaboration scenarios, multiple individual feature models are developed by the individual teams. So the main challenge is to interrelate the individual feature models such that they form a meaningful and comprehensive feature model. In other words, each of the individual feature models can be viewed as an independent module of a larger feature model; therefore, the large feature model constitutes several individual feature models and their interrelationships. This approach allows for a modular approach towards feature model design.

4.2. Modular Feature Models

Given the set of individual feature models developed individually by different teams, it is important to investigate how these feature models, referred to as feature model modules, are interrelated and integrated in terms of their integrity constraints and feature relationships such that they form one representative whole. We base our formalization on the notions in DDL. In the following, we would like to show how individual modules can be integrated such that they form a large modular feature model.

**Definition 1** Given a set \( \mathcal{F} = \{ \mathcal{F}_i \}_{i \in I} \) of feature model modules represented in \( \mathcal{ALCH} \) Description Logic, we denote the TBox of \( \mathcal{F}_i \) as \( \mathcal{T}_i \). Each TBox contains all the required structural knowledge as well as the integrity constraints of \( \mathcal{F}_i \), i.e., \( \mathcal{T}_i = \text{SKB}_i \sqcup \mathcal{IC}_i \). In this context, the set of all feature model modules can be represented as a corresponding family of consistent TBoxes \( \mathcal{T} = \{ \mathcal{T}_i \}_{i \in I} \).

In order to unambiguously refer to the features in different feature modules, they are prefixed with the index of the feature module, e.g., \( i : \text{Pickup} \sqsubseteq \text{Feature} \) shows that the pickup feature has been defined in feature module \( i \) (i.e., \( i \) denotes the index of the feature model module from which the feature, Pickup, has been derived).

Furthermore, to interrelate the individual feature modules, **bridge rules** in DDL are employed, and are referred to as **module bridges**.

![Figure 3. Feature model development from the integration of available fragmented feature models.](image)
Definition 2  Given a family of feature model modules and their corresponding TBoxes $T = \{T_i\}_{i \in I}$, a module bridge from module $i$ to $j$ is a feature model integrity constraint of one of the following forms:

\[
i : X \sqsubseteq \rightarrow j : Y, \text{ an into-module bridge;}
\]

\[
i : X \sqsupseteq \rightarrow j : Y, \text{ an onto-module bridge.}
\]

where $X$ and $Y$ are concepts or roles from $T_i$ and $T_j$, respectively. In case, one specific feature is represented differently in two feature modules, an equivalence-module bridge can be derived as a pair of into- and onto-module bridges, denoted $i : X \sqsubseteq \rightarrow j : Y$ to show that the features refer to the same concept.

Example 1  Let us consider the e-store feature model represented through four feature model modules in Figure 4. Now, recalling that features in different modules can be interrelated through integrity constraints, their relationships can be represented as module bridges. The following are the possible examples of module bridges:

\[
1 : \exists\text{hasFeature.Cash} \sqsubseteq \rightarrow 3 : \exists\text{hasFeature.Pickup}
\]

\[
3 : \exists\text{hasFeature.E - delivery} \sqsubseteq \rightarrow 1 : \exists\text{hasFeature.CreditCard}
\]

\[
1 : \exists\text{hasFeature.CreditCard} \sqsubseteq \rightarrow 4 : \exists\text{hasFeature.SpecialCharPassword}
\]

\[
2 : \neg(\exists\text{hasFeature.UniversalTaxRate}) \sqsupseteq \rightarrow 3 : \exists\text{hasFeature.Pickup}
\]

Module bridges allow for the integration of different feature modules through the application of integrity constraints as onto- or into-module bridges. Simply stated, module bridges can be viewed as integrity constraints between the features of two different feature modules. It is also possible to relate duplicate features that are present in multiple feature modules through equivalence-module bridges.

Given the notion of module bridges, it is possible to define modular feature models as a collection of feature modules and their interrelating module bridges. To this end, we define a feature model sentinel as a feature module that constitutes the core feature modeling concepts and feature model configurations of a modular feature model and unifies all of the individual feature modules into a whole.
Definition 3 A feature model sentinel is a feature module $S$, which is connected pair-wise to all feature modules, and consists at least of the following feature description concepts and roles:

- $\text{GenericFeature} \sqsubseteq \top$
- $\text{hasGenericFeature} \sqsubseteq \text{ObjectProperty}$
- $\text{GenericProductLine} \sqsubseteq \exists \text{hasGenericFeature}.\text{GenericFeature}$
- $\text{OurProductLine} \sqsubseteq \text{GenericProductLine}$

A feature model sentinel serves as the backbone of a modular feature model by containing the basic required axioms for defining a feature model. Feature definitions in the individual feature modules are extensions of these base concepts included in the sentinel. Later, we will show that stakeholder requests, $\text{SR}_h$, and any developed feature model configuration will be stored in the feature model sentinel. Interestingly, the implication of a sentinel is that it makes the individual feature modules independent of the final modular feature model that they are going to appear in. For instance, given a set of five feature modules, one can create a sentinel to form a modular feature model from three of them to address some problem domain of interest, while another person can create a different sentinel that brings all five of these modules into play to create a completely different modular feature model and addresses a different problem. In essence, sentinel is for modular feature models what glue code is for component-based software development.

Generally speaking, a feature model sentinel is the connecting module that shapes a modular feature model defined as follows:

Definition 4 A modular feature model is a triple $(F, B, S)$ where $F = \{F_i\}_{i \in I}$ represents the individual feature modules, $B = \{B_{ij}\}_{i \neq j \in I}$ denotes the set of all available module bridges, and $S$ is the feature model sentinel.

In order to characterize the topology of a modular feature model, a corresponding module bridge graph can be developed.

Definition 5 Given a modular feature model $(F, B, S)$, a module bridge graph is a directed graph with a set of nodes corresponding to the collection of $F = \{F_i\}_{i \in I}$ and an arc from node $i$ to node $j$ when the set of module bridges $B_{ij} \in B$.

Although $B_{ii}$ is not permitted in a modular feature model, we allow it to be present in the module bridge graph. The module bridge graph can be extended such that its arcs represent the number of module bridges between the connected nodes yielding a weighted module bridge graph. Module bridge graphs are appropriate underlying data structures for defining and analyzing structural quality metrics for feature models, such as cohesion, coupling, reusability, and others. Further analysis of module bridge graphs are outside the scope of this paper and will be discussed in separate work. Figure 5 shows the module bridge graph and its corresponding adjacency matrix for the modular feature model depicted in Figure 4.

Example 2 Let us consider the feature modules depicted in Figure 4. A modular feature model developed from these four individual feature modules consists of a feature model sentinel $S$, four feature modules $F_i$, $i = 1..4$, and several feature bridges including those described in Example 1.

As it can be seen in Figure 6, the sentinel sits at the center of the modular feature model and provides the base required constructs. The sentinel is connected to all feature modules through several module bridges, denoted as the set $B_S$. Each of the modules between the sentinel and the other feature modules ($B_{S,i}$) consists of the following axioms (not shown in the figure, in other words the arrow labeled $B_S$ will be carrying the following axioms):
Figure 6. A modular feature model with four feature modules and a sentinel.

\[
i : \text{Feature} \rightarrow S : \text{GenericFeature}
\]

\[
i : \text{hasFeature} \rightarrow S : \text{hasGenericFeature}
\]

Also the sentinel is given access to the top feature(s) of each module through a module bridge \((B_{Si})\) in order for us to be able to build the overall structure of the modular feature model in the sentinel. For instance, \(B_{S1}\) consists of \((S : \text{Payment} \equiv 1 : \text{Payment})\), which shows that the payment feature being referred to in the sentinel is equivalent to that defined in feature module 1. With the presence of these module bridges, a complete feature model can be put together in the sentinel for e-store as follows:

\[
E - \text{store} \equiv \exists \text{hasGenericFeature}.\text{Payment} \sqcap \exists \text{hasGenericFeature}.\text{Tax} \sqcap \exists \text{hasGenericFeature}.\text{Password}
\]

\[
\text{Shipping} \equiv \text{GenericFeature}
\]

Note that the features such as Payment, Tax, Password, and Shipping have already been defined in their respective feature modules and have been imported to the sentinel through the module bridges. So, the sentinel and the module bridges form a complete integration of individual feature modules and create a structured modular feature model for the e-store domain.

It is important to mention that the design teams responsible for each of the individual feature modules need to come together to designate the required module bridges in order to have a nice integration of the feature modules. The development and representation of feature models in a modular manner provides many benefits and advantages, such as:

- **Distributed/collaborative development**: Given that the modular formalism for feature models allows different aspects of a feature model to be encapsulated into several different feature modules, it is possible for the design team to divide the domain of discourse into separate subdomains. Once the subdomains are identified, different teams can work on each of
these subdomains as a result of which several feature modules are developed, each of which constitutes the features of one of the subdomains. The various design teams can then collaboratively work together to integrate these feature modules into a comprehensive modular feature model. The other possibility that is offered is that since the teams are working on their own subdomain and feature module, they can work independently of each other in a distributed setting:

- **Feature model maintenance and evolution:** One of the important issues that designers face is the need for continuous update and refinement of their models referred to as model maintenance and evolution. Given the fact that within the modular feature model formalism, feature modules focus on the features of a specific subdomain of discourse and hence are independent of each other from the perspective of the subdomain that they cover, designers can update or maintain each feature module independently of the other modules. Having said that, this does not mean that changes within a feature module do not affect the modular feature models that contain this feature module, but rather it means that changes in one feature module will be contained in that module and will not propagate to other modules and hence it enhances the process of evolution and maintenance;

- **Module reusability:** Generally, product line engineering is one of the important approaches for re-use base engineering. Modular feature models further enhance reusability by providing feature modules. As mentioned earlier, feature modules capture the features of a specific subdomain of discourse and can be used in the creation of any larger modular feature model that covers that specific subdomain. For this reason, each feature module does not necessarily need to be used only in one modular feature model and can be reused in different contexts for other purposes as well. Therefore, any one feature module can be reused in the construction of many other modular feature models. Besides reusability, it can indirectly increase productivity, promote knowledge transfer between projects, and reduce the likelihood of repeating design mistakes.

In the following, we investigate how a given configuration of a modular feature model can be validated to see whether they conform to its structural knowledge and integrity constraints.

### 4.3. Configuration Verification

A configuration of a feature model is the result of the selection of a certain number of feature model elements and the removal of the others. A feature model configuration corresponds with the requirements of a specific domain application. Designers often configure a feature model such that the maximum number of stakeholder requests and requirements for the given application are satisfied; however, in large feature models, due to the structural and integrity constraints imposed on the available features, not all configurations are valid. With the increase in the number of selected features, keeping track of the valid configurations becomes harder to be performed manually. In this paper, since the representation of modular feature models is performed in Description Logics, it is possible to use standard Description Logic inference mechanisms to perform the configuration validation process. Any Description Logic-based system provides at least two basic standard reasoning mechanisms:

**Concept satisfiability** \( T \models D \nsubseteq \bot \). Given a DL TBox \( T \) and a concept \( D \), does there exist at least one model of \( T \) assigning non-empty extension to \( D \)?

**Subsumption** \( T \models D \subseteq C \). Given a DL TBox \( T \) and two concepts \( C \) and \( D \), is \( D \) more general than \( C \) in any model of \( T \)?

So, given a feature model \( F \) expressed in terms of Description Logic TBox \( T \), which conveys a software product line description, such as that of the electronic online store, \( PL = SKB \sqcup IC \), and a feature model configuration \( C \), using the two standard reasoning mechanisms it is possible to see whether 1) the feature model configuration fully conforms with the feature model description using the entailment of concept subsumption; 2) they are at least compatible with each other but the configuration may contain features that are not included in the feature model through the satisfiability of concept conjunction; 3) the feature model configuration is not at all compatible with the feature model by the entailment of concept disjointness. Hence, it is quite straightforward to verify whether a feature model configuration developed manually is a valid configuration or not. However, since concept satisfiability and subsumption perform on a monolithic TBox, we need to temporarily convert our modular feature model representations into a monolithic representation. Distributed Description Logics provides the means to do so through the bridge operator.

**Definition 6** Given a set of module bridges \( B_{ij} \) from feature module \( F_i \) to \( F_j \), an operator \( B_{ij}(\cdot) \), taking as input features in \( F_i \) and producing a set of corresponding features for them in \( F_j \), is called the bridge operator.

Details of the bridge operator can be found in [9]. For our purpose, we need to perform the bridge operator on and from all available feature modules onto the feature model sentinel. In this way, the sentinel will temporarily contain the representative monolithic form of the modular feature model, over which the feature model configuration verification process can be easily performed.

It is important to mention that from a high level designer perspective the verification process is performed on the modular form of the feature models, and the designers are oblivious to the temporary transformation of feature modules into a temporary monolithic sentinel. This is also true for the algorithms proposed in the following sections, where the designers interacting with the feature model will only see a modular feature model representation, which preserves all of its discussed benefits.

In the following section, we will benefit from and customize a method originally proposed by Ragone et al. for distributed Web service composition [10]. In their method, concept abduction is used to address the concept covering problem in Description Logic, which is employed for creating an automated Web service composition process. As we will see, their approach
provides inexact composition solutions when an exact solution cannot be found. We benefit from the similarity of the Web service composition problem and modular feature model configuration, in which both problems are trying to create a customization for a set of distributed lowly coupled entities based on some given criteria, to create a configuration process for modular feature models.

5. Modular Feature Model Configuration

It is common practice that feature model configurations are derived based on stakeholders’ requests, i.e., the stakeholders specify the set of important features that are required to be present in the final product. Based on these stakeholder requests, the designers would then try to customize the feature model such that the maximum number of these requests are satisfied. However due to feature interdependencies, it is possible that not all of the requests are satisfiable. Furthermore, the complexity of large industrial-scale feature models makes it hard for the designers to manually find an acceptable configuration for the feature model based on the requests; therefore, the development of automated feature model configuration techniques that would facilitate this process is desirable. For instance, it would be beneficial to have an automated configuration tool that would take as input the stakeholder requests, the structural knowledge and the integrity constraints of the feature model and provide as output a suggested configuration.

In order to develop such an automatic configuration algorithm for modular feature models\(^1\), we base our work on the notion of concept covering as presented in [10]. It is important to note that existing feature model configuration algorithms all only work with monolithic feature models, whereas the following proposed algorithm is designed to work with modular feature models. The following sections introduce how concept covering in Description Logics can be employed to perform feature model configurations in modular feature models.

5.1. The Concept Covering Problem

In the context of the modular feature model configuration verification process, it was shown that concept subsumption can be used to evaluate whether a given feature model configuration, denoted \(C\), is consistent with the structural knowledge of a feature model, denoted \(PL\). In fact if the relation \(T \models C \subseteq PL\) holds, the configuration is consistent within the framework of the feature model; however, if \(T \models PL \cap C \not\subseteq \bot\) holds, since satisfiability and subsumption only return a boolean truthfulness value, it is not possible to find an explanation for why the entailment of concept disjointness holds. The Concept Abduction Problem (CAP) was introduced and defined in [5], to provide an explanation hypothesis when subsumption does not hold.

\[\text{Definition 7} \quad \text{Let } S, D \text{ be two concepts and } T \text{ a set of axioms in a Description Logic } \mathcal{L}, \text{ where both } S \text{ and } D \text{ are satisfiable in } T. \text{ A CAP, denoted as } \langle \mathcal{L}, S, D, T \rangle, \text{ is finding a concept } H \text{ such that } T \not\models S \cap H \subseteq \bot, \text{ and } T \models S \cap H \subseteq D.\]

Given a CAP, if \(H\) is a conjunction of concepts and no sub-conjunction of concepts in \(H\) is a solution to the CAP, then \(H\) is an irreducible solution. In [5], the minimality criteria for \(H\) and a polynomial algorithm, to find solutions which are irreducible, for \(ALCN\ DL\), have been proposed. More recent work has extended support for CAP in \(SH\ DL\) [11], which is an extension of \(ALCH\ DL\) that we use in this paper \((ALCH \subset SH)\) and can hence be applied to our specific problem.

As described in [12], the solution to a CAP can be read as why does not \(S \not\subseteq D\) hold?, what is a possible explanation with respect to \(T\)? In other words \(H\) represents what is expressed, explicitly or implicitly (that is, entailed because of the TBox), in \(D\) and is not present in \(S\), or also which part of \(D\) is not covered by \(S\)?

So, how can a solution to CAP be useful for modular feature model configuration? To see the relationship, let us suppose that stakeholder requests are represented as \(SR_h\), and some feature model configuration \(C\) has been provided by the designers to satisfy the stakeholders’ requests. It is intuitive that \(H\), in a CAP solution, is an explanation hypothesis for the missing part in the available feature model configuration \(C\) needed to completely satisfy a stakeholders’ requests in \(SR_h\). A not-full match with the stakeholders’ request is then due to \(H\). This is an observation that can also be expressed by defining \(H\) as: what is not covered by \(C\) with respect to \(SR_h\).

Based on this last remark, we use solutions to sets of CAP to create a correspondence between modular feature model configuration and concept covering and solve concept covering problems as it is shown in the following. In particular, we show that we are able, through concept abduction, to formulate the definition of concept covering, and we introduce a greedy algorithm\(^2\) to compute concept covering, or in other words feature model configurations based on stakeholders’ requests, in terms of concept abduction [13, 14]. Such an algorithm provides as output the obtained feature model configuration and, when such a complete configuration does not exist, a logical explanation on what remains missing in the provided feature model configuration with regards to the stakeholders’ requests.

Here, we use the following definition for the concept covering problem:

\[\text{Definition 8} \quad \text{Let } D \text{ be a concept, } \mathcal{R} = \{S_1, S_2, ..., S_k\} \text{ be a set of concepts, and } T \text{ be a set of axioms, all in some DL } \mathcal{L}, \text{ e.g., } \mathcal{L}=ALCH, \text{ where } D \text{ and } S_1, ..., S_k \text{ are satisfiable in } T. \text{ Let also } \prec_T \text{ be an order relation over the concepts in } \mathcal{L} \text{ taking into account the TBox } T. \text{ The Concept Covering Problem } (CCoP) \text{ denoted as } V=\langle \mathcal{L}, \mathcal{R}, D, T \rangle \text{ is finding a pair } \langle \mathcal{R}_c, H \rangle \text{ such that}
\]

1. \(\mathcal{R}_c \subseteq \mathcal{R}\), and \(T \not\models \cap_{S_i \in \mathcal{R}_c} S_i \equiv \bot;\)
2. \(H\) is a solution to \(\langle \mathcal{L}, S, D, T \rangle\), and \(H \prec_T D\).\]
We call \( \langle R_c, H \rangle \) a solution for \( V \), and say that \( R_c \) covers \( D \). Finally, we denote SOLCCoP\((V)\) the set of all solutions to a CCoP \( V \).

In the above definition, the elements for the solution \( \langle R_c, H \rangle \) of a CCoP represent respectively:

- \( R_c \): Which concepts in \( R \) represent the cover for \( D \) w.r.t. \( T \).
- \( H \): What is still in \( D \) and is not covered by concepts in \( R_c \).

Intuitively, \( R_c \) is a set of concepts that completely or partially cover \( D \) w.r.t. \( T \), while the abduced concept \( H \) represents what is still in \( D \) and is not covered by \( R_c \).

Di Cugno et al. [15] argue that concept covering problem is similar, but has remarkable differences when compared to classical set covering. CCoP is not trivially a different formulation of a classical minimal set covering (SC) problem, as an exact solution to a CCoP may not exist. Furthermore in SC elements are not related with each other, while in CCoP elements are related with each other through axioms within the TBox, and while the aim of an SC is to minimize the cardinality of \( R_c \), the aim of a CCoP is to maximize the covering, hence minimizing \( H \).

Several solutions can exist for a single CCoP, depending on the strategy that is adopted for choosing concepts present in \( R_c \). This fact gives way to the definition of best cover and exact cover.

**Definition 9** ([16]) A best cover for a CCoP \( V \), w.r.t. an order \( \prec_T \) on the adopted DL, is a solution \( \langle R_c, H_b \rangle \) for \( V \) such that there is no other solution \( \langle R_c', H' \rangle \) for \( V \) with \( H' \prec_T H_b \).

There is no solution \( \langle R_c', H' \rangle \) for \( V \) such that \( H' \), the remaining part of \( D \) still to be covered, is ‘smaller’ than \( H_b \) with respect to an order \( \prec_T \). Observe that, since the solution for a concept abduction is not unique in general, there could be two solutions \( (R_c, H_1), (R_c, H_2) \) such that the first is a best cover, while the second one is not. However, when an exact cover exists, it is independent of any order \( \prec_T \) on the adopted DL. However, for our purpose the SIM-DL algorithm presented in [17] can be used to compute an order \( \prec_T \) for Description Logic ALC\(N/R\). SIM-DL can be used for our purpose since we use A\(\mathcal{L}\)C\(\mathcal{H}\) which is a subset of A\(\mathcal{L}\)C\(\mathcal{N}/R\).

**Definition 10** ([16]) An exact cover for a CCoP \( V \) is a solution \( \langle R_c, H_e \rangle \) for \( V \) such that \( T \models H_e \equiv T \).

### 5.2. CCoP for Modular Feature Model Configuration

Now, it can be seen that a solution to a concept covering problem corresponds with a configuration for a given modular feature model and a set of stakeholders’ requests. Here, the aim is to find a cover for the stakeholders’ requests in the context of the modular feature model. An exact cover is equivalent to a feature model configuration that satisfies all of the stakeholders’ requests. However, in cases where such a configuration cannot be found, the best possible configuration is strived for, which will satisfy the maximum number of stakeholder requests. Therefore, it is possible to define the feature model configuration process as a specialization of the concept covering problem:

**Definition 11** (Extends Definition 8) Let \( SR_h \) be the stakeholders’ requests, \( F = SKB \cup IC \) be a given modular feature model defined in terms of its structural knowledge and integrity constraints, and \( T \) be a set of axioms used to define \( F \), all in a DL \( L \). Let also \( \prec_T \) be an order relation over \( L \) taking into account the TBox \( T \) of \( F \). The modular feature model configuration problem is defined as CCoP \( V \equiv \langle L, F, SR_h, T \rangle \). All possible feature model configurations for modular feature model \( F \) based on \( SR_h \) are attained by SOLCCoP\((V)\).

Following from this correspondence, a solution \( \langle C, H \rangle \) for \( V \equiv \langle L, F, SR_h, T \rangle \) represents a feature model configuration such that:

- \( C \) shows the features of the modular feature model in \( F \) that form the structure of the produced configuration;
- \( H \) consists of all the features that are still remaining in the set of stakeholders’ requests \( SR_h \) that have not been addressed by the configuration provided in \( C \).

So informally stated, a feature model configuration that would satisfy all of the requests of the stakeholders would be a pair \( \langle C, H \rangle \), \( H \equiv T \). This means that all members of \( SR_h \) are present in \( C \). However, if such a configuration does not exist, a configuration \( \langle C, H \rangle \), where for any other possible derivable configurations such as \( \langle C', H' \rangle \), we have \( H \prec_T H' \), is the best possible achievable configuration, which leaves the least number of unsatisfied stakeholder requests. It is desirable that an exact configuration is sought first, and in case of it being unattainable, a (near) best configuration be developed.

It is well-known that even the basic set covering problem is NP-hard. For concept covering to be of any practical use in a feature model configuration process, it has to be reasonably fast. With this issue in mind, we have extended the tractable polynomial greedy concept covering algorithm proposed for Web Service composition in [10] making it suitable for feature model configuration, building on and extending a classical greedy set covering one [18]. The algorithm takes as input the structural knowledge and integrity constraints of a modular feature model along with a set of specific stakeholder requests, and develops a corresponding configuration.
Algorithm \texttt{BuildConfiguration}(\mathcal{F}, \mathcal{SR}_h, \mathcal{T})

\textbf{input}  concepts \(\mathcal{SR}_h, \mathcal{S}_i \in \mathcal{F}\) and \(\mathcal{F} = \bigcup_j \mathcal{B}(j) \cup \mathcal{S}\), where \(\mathcal{SR}_h\) and \(\mathcal{S}_i\) are satisfiable in \(\mathcal{T}\) and \(\mathcal{S}\) is the feature model sentinel

\textbf{output} \(\langle C, H \rangle\)

\begin{algorithmic}
\begin{itemize}
\item[] \textit{/* \([\bigtriangleup]\): Add all of the implications of the stakeholder requests to \(\mathcal{SR}_h\) */}
\item[] \textbf{for each} \((f \subseteq \exists\text{hasFeature.} \hat{f}) \in \mathcal{F}\)
\item[] \hspace{1em} \textbf{if} \(f \in \mathcal{SR}_h\) \textbf{then}
\item[] \hspace{2em} \(\mathcal{SR}_h = \mathcal{SR}_h \cap \hat{f}\);
\item[] \hspace{1em} \textbf{end if}
\item[] \textbf{end for each}
\item[] \(C = \emptyset\);
\item[] \(\mathcal{SR}_{hu} = \mathcal{SR}_h\);
\item[] \(H_{\text{min}} = \mathcal{SR}_h\);
\item[] \textbf{do}
\item[] \hspace{1em} \(S_{\text{MAX}} = \top\);
\item[] \hspace{1em} \textit{/* \([\spadesuit]\): Perform a greedy search among \(\mathcal{S}_i \in \mathcal{F}\) */}
\item[] \hspace{2em} \textbf{for each} \(\mathcal{S}_i \in \mathcal{F}\)
\item[] \hspace{3em} \textbf{if} \((C \cup \{\mathcal{S}_i\})\) is a cover for \(\mathcal{SR}_{hu}\) and \((\mathcal{T} \models C \cup \{\mathcal{S}_i\} \neq \bot)\) then
\item[] \hspace{4em} \(H = \text{solveCAP}(\langle L, \mathcal{S}_i, \mathcal{SR}_{hu}, \mathcal{T} \rangle); \text{ //solveCAP solves a concept abduction problem.}\\
\item[] \hspace{4em} \textit{/* \([\heartsuit]\): Choose \(\mathcal{S}_i\) based on an order \(\prec_T\) */}
\item[] \hspace{5em} \textbf{if} \(H \prec_T H_{\text{min}}\) \textbf{then}
\item[] \hspace{6em} \(S_{\text{MAX}} = \mathcal{S}_i\);
\item[] \hspace{6em} \(H_{\text{min}} = H;\)
\item[] \hspace{5em} \textbf{end if}
\item[] \hspace{2em} \textbf{end if}
\item[] \hspace{1em} \textbf{end for each}
\item[] \hspace{1em} \textit{/* \([\clubsuit]\): If a new \(\mathcal{S}_i\) is found then add \(\mathcal{S}_i\) to \(C\) and remove it from \(\mathcal{F}\) */}
\item[] \hspace{2em} \textbf{if} \(\mathcal{T} \models S_{\text{MAX}} \neq \top\) \textbf{then}
\item[] \hspace{3em} \(\mathcal{F} = \mathcal{F}\{\mathcal{S}_i\}\);
\item[] \hspace{3em} \(C = C \cup \{\mathcal{S}_i\}\);
\item[] \hspace{3em} \(\mathcal{SR}_{hu} = H_{\text{min}};\)
\item[] \hspace{3em} \textit{/* \([\infty]\): Enforce modular feature model integrity constraints */}
\item[] \hspace{4em} \textbf{for each} \((\exists\text{hasFeature.} f \subseteq \exists\text{hasFeature.} \hat{f}) \in \mathcal{F}\)
\item[] \hspace{5em} \textbf{if} \(f \in C\) \textbf{then}
\item[] \hspace{6em} \(\mathcal{F} = \mathcal{F}\{\hat{f}\};\)
\item[] \hspace{6em} \(C = C \cup \{\hat{f}\};\)
\item[] \hspace{5em} \textbf{end if}
\item[] \hspace{4em} \textbf{end for each}
\item[] \hspace{4em} \textbf{for each} \((\exists\text{hasFeature.} f \subseteq \neg\exists\text{hasFeature.} \hat{f}) \in \mathcal{F}\)
\item[] \hspace{5em} \textbf{if} \(f \in C\) \textbf{then}
\item[] \hspace{6em} \(\mathcal{F} = \mathcal{F}\{\hat{f}\};\)
\item[] \hspace{5em} \textbf{end if}
\item[] \hspace{4em} \textbf{end for each}
\item[] \hspace{1em} \textbf{end if}
\item[] \hspace{1em} \textit{/* \([\heartsuit]\): Continue searching until no \(\mathcal{S}_i\) is found */}
\item[] \hspace{2em} \textbf{while} \(S_{\text{MAX}} \neq \top\),
\item[] \hspace{3em} \textbf{return} \(\langle C, \mathcal{SR}_{hu}\rangle;\)
\item[] \textbf{end do}
\item[] \textbf{end algorithm}
\end{itemize}
\end{algorithmic}

This algorithm attempts to create a configuration that covers \(\mathcal{SR}_h\) as much as possible, using the concepts \(\mathcal{S}_i \in \mathcal{F}\). In the first step, stakeholders’ requests are expanded such that they include all of their required child features \([\bigtriangleup]\). Then, \texttt{BuildConfiguration} adopts a greedy approach towards the selection of the most appropriate features to be included in the developed configuration \([\spadesuit]\), i.e., it solves the concept abduction problem for all the available features in \(\mathcal{F}\), and selects the feature that provides the closest match for \(\mathcal{SR}_h\) \([\heartsuit]\). Once such a feature is identified, it is added to the configuration and removed from the list of stakeholder constraints \([\clubsuit]\). Furthermore, the addition of this feature to the configuration requires the enforcement of the integrity constraints, i.e., the removal of the features in an excludes constraint and addition of features in an includes relation with the recently selected feature \([\infty]\). This process is repeated until no more appropriate features can be added to the configuration \([\heartsuit]\). Finally, \texttt{BuildConfiguration} returns the developed configuration \(C\). If the configuration does not cover all of the stakeholder requests, i.e., it is not an exact cover, the explanation for why an exact cover was not achieved is presented in \(\mathcal{SR}_{hu}\).
6. Case Studies

In this section, we will provide two case studies. The first one builds on the electronic store feature model that we have used throughout the paper to explain step-by-step how a modular feature model can be configured based on our proposed procedure. In this case study, we assume that the monolithic feature model shown in Figure 2 has been modularized and represented as depicted in Figures 4 and 6. In the second case study, we will investigate and compare the performance of the configuration process in terms of the time taken to create a configuration on both the monolithic and modular feature models.

6.1. Case Study 1

Let suppose that in the context of the modular feature model discussed in Example 2, the stakeholders have requested a simple configuration of the e-store where the required features are credit card payment, pickup item delivery and a never changing password option. The required features represent the base/minimum set of features that are mandatory (highly desirable) to be present in the final configuration. The stakeholders’ request can be represented as:

$$SR_h \equiv e \setminus \text{store} \setminus \text{CreditCard} \setminus \text{Pickup} \setminus \text{Never}$$

In order to form a configuration of the modular electronic store feature model, we need to employ the BuildConfiguration algorithm. The algorithm evaluates the stakeholders’ requests in several rounds and creates a desirable configuration given the requests. So given $SR_h$, we show in the following just the first round of the BuildConfiguration algorithm is performed. The subsequent rounds of the algorithm are performed similarly. The following steps are marked with the corresponding symbol in the algorithm:

(△) To formulate the configuration, BuildConfiguration first extends to $SR_h$ with all of the child features of the requested features, which were originally in $SR_h$. As it can be seen, it is only e-store that has child features in this example, so the expansion of e-store in $SR_h$ entails:

$$SR_h \equiv e \setminus \text{store} \setminus \text{CreditCard} \setminus \text{Pickup} \setminus \text{Never} \setminus \text{Payment} \setminus \text{Tax} \setminus \text{Password}$$

Now, the new features in the expanded $SR_h$ need to be expanded; hence, Password and Tax should be unfolded. This process is performed until no further expansion is possible.

(♣) The second step consists of a greedy search between the modular feature model features to find the one that provides the most covering over $SR_h$.

(◊) Here, the algorithm chooses a feature that provides the most covering for $SR_h$ from the features available in the modular feature model. This is calculated based on the ordering created by $\prec_T$. It is easy to see that since the selection of the CreditCard feature provides a more covering specialization of the feature model for the given stakeholder requests in $SR_h$, it will be selected.

(♠) The feature chosen by the greedy feature selection process, in this case CreditCard, can be added to the configuration that is being gradually developed. Therefore, $C = \{\text{CreditCard}\}$. Also this feature is removed from the list of available features so that it is not processed in the future rounds of the greedy algorithm.

(∞) Once the selected feature is added to the configuration, it must be ensured that its related integrity constraints are also enforced. For this sake, since the credit card feature has two integrity constraints that require the inclusion of the AutomaticFraudDetection and SpecialCharacterPasswords features (see Section 2.1); therefore, they will also be added to the configuration. In this step all of the integrity constraints of the features in C are enforced, so now because the AutomaticFraudDetection and SpecialCharacterPasswords features are in C their integrity constraints will also be enforced if any.

(◇) Finally, the algorithm moves onto the next suitable feature (e.g., Pickup) to form the remaining segments of the configuration.

This process is repeated until an acceptable configuration is developed. As it can be seen, if the features initially requested by the stakeholders are consistent, i.e., both them and the features enforced by the integrity constraints can appear simultaneously in a configuration, the BuildConfiguration algorithm is able to create a sound and complete configuration based on the requests (see Section 7 for further details). However, if the requests are not consistent then the algorithm will try its best to satisfy the maximum number of the requests in the developed configuration according to its greedy nature.

6.2. Case Study 2

Let us now analyze the execution time performance of the modular feature model configuration process in contrast with monolithic feature models.

In the second case study, we have studied the impact of modularization on the time required for the configuration (specialization) of a given feature model based on some set of stakeholder requests. To analyze the execution time of the configuration...
process, we report the results of four scenarios. Each of the scenarios are performed over large-scale synthesized feature models created by the SPLOT Feature Model Generator [19], which is being recently used in the research community for evaluating feature model configuration methods and the execution time evaluations are based on the framework provided by [20]. The main reason why we chose synthetic feature models for our work is that many of the real world feature models that are designed within the industry are proprietary and are not shared with the academic community. The only shared feature models that are available through the SPLOT repository are rather small feature models that are not really suitable for showing the performance of the proposed modular feature model configuration algorithm. Hence, we chose the SPLOT Feature Model Generator to create large feature models based on the characteristics of real world feature models that would allow us to analyze the performance of our proposals.

The scenarios for our four experiments that we have performed are described as follows:

Scenario 1 –

**Purpose:** The goal of the first scenario was to analyze if modularization enhances the execution time of the configuration process. In other words, would the time needed for configuring a modular feature model be less than that needed for configuring the same feature model but with a monolithic design?

**Procedure:** For this scenario, several feature models with exactly 2000 features were generated using the SPLOT feature model generator tool and were each partitioned into different number of modules. Modularization was performed such that features with more integrity constraints were kept together as much as possible. The number of requests from the stakeholders was set to 20. The generated feature models were configured based on the algorithm proposed in this paper and the average configuration running time was calculated.

**Observation:** Figure 7.a shows that the configuration process for a feature model with only one module (monolithic feature model) takes much longer in terms of running time compared to the other cases (i.e., modular feature models with 2000 features and varying number of modules with more than one module).

**Interpretation:** This observation shows that given the same number of features in a feature model, modularization can enhance execution time of configuration. It should also be noted that although modularization has reduced the execution time, the execution time reduction for modular feature models with more than 6 modules is not significant, e.g. compare the modular feature model with 6 and 9 modules in Figure 7.a. This can be due to the fact that in this scenario we have kept the number of features at a constant of 2000 and therefore, increasing the number of modules without an increase in the number of features would not enhance the execution time any further. So given this observation, we believe that when the execution time for configuration of the modular feature model remains relatively the same, as is the case for modular feature models with 6, 9, 12, and 15 modules, the decision on how many modules is required is dependant on the semantics of the feature model. The feature model designers would be the best for deciding on the best number of modules for the modular feature model.

Scenario 2 –

**Purpose:** In the second scenario, the goal was to analyze whether an increase in the number of features in a given feature model effects the configuration execution time or not. In other words, does the number of features in a modular feature model impact the configuration time?

**Procedure:** For this purpose, different feature models were generated using the SPLOT feature model generator, each of which consisted of different number of features. Each of these feature models were partitioned into numerous modules. Similar to the first scenario, the number of stakeholder requests was set to 20. These feature models were then configured and the average running time was recorded.

**Observation:** Figure 7.b depicts the execution time of the configuration algorithm on these different feature models. It can be seen that as the number of features increases the configuration execution time also increases.
**Interpretation:** The important observation from 7.b is that although the configuration execution time rises with the increase in the number of features, but at the same time a comparison between the modular feature models with the same number of features shows that those modular feature models that have a higher number of modules have a smaller configuration execution time.

**Scenario 3**

**Purpose:** The intention of the third scenario is to investigate whether the increase in the number of stakeholder requests has any substantial effects on the configuration execution time of similar feature models or not. In other words, does the increase in the number of requests from the stakeholders for the same feature model impact the configuration execution time?

**Procedure:** In the previous scenarios, the number of stakeholder requests were set to a constant of 20. However, in this third scenario, the number of features were set to a constant of 2000 in all of the feature models, but the number of requests changed in the different experiments. Therefore, different feature models all with 2000 features were generated and modularized. Subsequently, the configuration algorithm was run on these modular feature models with varying number of stakeholder requests. The average running time was recorded.

**Observation:** As it can be seen in Figure 8.a, the increase in the number of requests negatively affects the monolithic feature model, while the modular feature models stay relatively robust against the increase in the number of requests.

**Interpretation:** From the observation, we can concluded that modularity of feature models keeps the configuration time robust to changes in the number of requests from the stakeholders. This can be due to the fact that the requests will be divided between the different modules for satisfaction in the modular feature models, which decreases the execution time for configuration. On the other hand in monolithic feature models, all of the requests are assigned to the one module in the monolithic design, which causes an increase in the execution time.

**Scenario 4**

**Purpose:** In Scenario 4, the goal was to analyze the execution time performance of the configuration algorithm given varying number of features and feature modules. In other words, would the running time of the algorithm be impacted as the number of features and feature modules increase in modular feature model or not.

**Procedure** As shown in Figure 8.b, we increased both the number of modules and the number of available features in different experiments. The number of stakeholder requests were kept at a constant of 20. Different feature models with varying number of features were generated and modularized. The configuration algorithm was run on these feature models and the average time was calculated.

**Observation:** It is interesting to see that although the number of features and feature modules have gradually grown, the configuration time has stayed relatively the same in the experiments; however, in the monolithic feature model, with the increase in the number of features, the execution time grows much higher.

**Interpretation:** The stability in the execution time of the configuration algorithm on modular feature models with varying number of features and feature modules and the increase in the execution time of monolithic feature models is an indication of the fact that modular feature models are not only useful for enhancing reuse and maintainability but are also effective in enhancing the feature model configuration execution time.

It is important to analyze why modularization impacts the configuration time and manages to tame it under different scenarios. From our point of view, the main reason behind this is because in the proposed configuration algorithm, that has been introduced in this paper, only those modules of the feature model are included in the configuration process, whose features are required to be present in the final feature model configuration. For instance, if a feature model has 10 modules, it is possible that the
features of only 5 of its modules will be included in the configuration process. Therefore, since this process removes some of the extra modules that are not relevant to the configuration process at hand, therefore, the configuration execution time will be much lower. On the other hand, within monolithic feature models since all of the features will be included in any configuration process and even the superfluous ones will not be ignored, the average running time of the configuration algorithm is higher than their modularized counterparts.

The other parameter that could play an important role in the configuration time of a modular feature model is the distribution of stakeholder requests among the available modules. For instance, if all of the stakeholder requests concern only one module, then the configuration time will be much smaller compared to the case when the stakeholder requests cover all of the available modules. It is worth mentioning that when the stakeholder requests include features from all of the feature model modules, then the configuration execution time of a modular feature model is equivalent to the configuration execution time of the equivalent monolithic feature model.

7. Discussions

Modular feature models offer various advantages to the domain analysts, some of which include the possibility of collaborative and distributed development, more efficient feature model evolution, enhanced feature reusability, and others. However, the practicality of the employment of modular feature models depends on the efficiency of the underlying representation formalism that is used to construct such models. Previous experience in representing feature models using Description Logics [8] has shown promising prospect in terms of benefiting from the standard reasoning capabilities of this formal logic. Furthermore, there has been an extensive investigation of the possibility of developing modular Description Logic knowledge bases that provides the opportunity to gain from such developed technologies in the area of feature models. Different modularity frameworks exist that include $\mathcal{E}$-connections [21], Interface-based Formalism (IBF) [22], Distributed Description Logic (DDL) [23], and Package-based Description Logics (P-DL) [24]. Each of these formalisms provide some functionality that can be used for different purposes. We believe that DDL can best represent the requirements of a modular feature model. This is because DDL allows the development of self-contained independent knowledge bases (individual feature modules) that can be interconnected through bridges that connect the modules, which represents in essence the structure of a modular feature model. However, the focus of the other modularity formalisms is not that well aligned with our purpose: 1) in IBF [22] the attention is more on knowledge encapsulation and inheritance, which are not the main concerns in modular feature models; 2) $\mathcal{E}$-connections [21] defines modularity in terms of the connection of modules through connecting rules whose domain is in one ontology and their
range in the other, and the basic assumption is that the domain of the modules are disjoint, which is not a condition that we want to enforce for modular feature models; 3) P-DL [24] revolves around the notion of concept import where any module can bring concepts from other modules into its structure. This is in contrast to modular feature models that require feature interaction and mapping which is best supported by the DDL framework.

Furthermore, the Distributed Description Logic framework provides a support suite, called Drago⁵, that supports visualization and reasoning over modular description logics represented in OWL. So, it is possible to define the feature modules and the sentinel as described before as ontology modules and make use of the visualization and reasoning capabilities of Drago. Figure 9 depicts the visualization of the connection of a feature module with the sentinel in Example 2. The module bridges are shown as lines connecting the feature modules, which makes the comprehension of the formation of the modular feature model easier.

The representation of modular feature models through the employment of the distributed description logic framework also provides the advantage of being able to create a topology of the feature model structure in terms of a (weighted) graph, called the module brief graph. The topology is developed by representing feature modules as nodes and module bridges as edges. This graph-based representation of modular feature models gives way to a lot of potential techniques from graph theory that can be used to analyze the properties of the modular feature model that can assist in the improvement of the quality of the feature model. This is similar to what is performed in software metrics design [25], for instance the cyclomatic complexity of a program, which is the control flow within that program represented in graph format.

The other important aspect of the proposed work of this paper is the development of feature model configurations from modular feature model descriptions and stakeholder requests. It was shown that a configuration can be developed by making a correspondence with the concept covering problem in DL, a problem similar to set covering. In [18] it has been proven that for a set covering problem, the solution grows logarithmically in the size of the set to be covered with respect to the minimal one. In this case, the BuildConfiguration algorithm, the complexity source is in the solution of the CAPs and the ordering performed in [19]. It is possible to show that BuildConfiguration can be solved in polynomial time. However, it should be noted that the solution provided by this algorithm may not be complete; however, it is always sound. The reason that BuildConfiguration does not guarantee a complete configuration is due to its greedy nature. Since covering problems are yet to be solved polynomially, and reaching an acceptable solution in a reasonable time is crucial for our problem, we resort to a polynomial greedy algorithm for BuildConfiguration. The algorithm returns a good configuration, which may not always be the best one.

It is worth noting that although the underlying formalism for representing modular feature models has been Description Logics in this paper, in reality, feature model modules can be simply developed using any feature modeling tool such as SPLIT [19], and then we automatically convert these feature modules into Description Logic format (OWL ontologies) without the need for the designers to be aware of this or be familiar with Description Logics in any way. Software product line designers will only interact with feature model design tools and the conversion to and from Description Logics is automatically handled behind the scenes. For our research, we have used SPLIT and Drago for implementation and testing purposes in terms of creating, manipulating, configuring and visualizing modular feature models introduced in this paper.

8. Related Work

Within the realm of software product family engineering, feature models have been viewed as configurable complex systems that can be adapted to produce suitable software products under different circumstances and requirements [26]. Therefore, feature model design is most suitable for complex and large-scale problem domains with numerous features and product options. This requires a structured representation for feature models to be able to manipulate and reason over the developed models. To this end, Mannion was the first to propose the adoption of propositional formula for formally representing software product lines [27]. This idea has been extended by creating a connection between feature models, grammars and propositional formula in [28]. Grammars have also been used in other work such as [3] where context-free grammars have been used to represent the semantics of cardinality-based feature models. This semantic interpretation of a feature model relates with an interpretation of the sentences recognized as valid by the context-free grammar. However, such techniques can have limitations in defining the integrity constraints of the feature models. For this reason, Czarnecki and Kim employ the Object Constraint Language (OCL) to express such additional restrictions [29]. In their approach, feature models are converted into UML diagrams where integrity constraints are enforced on them through OCL expressions.

Wang et al. have addressed the issue of the formal representation of feature models from a different angle. In their approach, they employ OWL description logic ontologies to model feature model structure and integrity constraints [8]. Their reliance on OWL DL for representing feature models opens up the possibility of employing already existing OWL reasoners to perform feature model configuration validation and verification checks. For instance in their experiments, FaCT++ [30] has been used to see whether a developed configuration is valid or not. Our work in this paper can be seen as a further exploration of the work by these authors. The major advantages of our work is that we provide a modular approach to modeling feature models, and also that we provide an algorithm to automatically build modular feature model configurations based on some initial stakeholder requests, while the work by the mentioned authors is only restricted to monolithic feature models and only perform configuration verification and not configuration development. Along the same lines, the work by Zaid et al. [31] employs description logic ontologies and Semantic Web rules expressed in SWRL [32] to interrelate segmented feature models but only supports for configuration verification.

Similar to the description logic representation where reasoning is performed on feature models using standard DL reasoners, feature model configurations can be verified using Logic Truth Maintenance Systems (LTMS) in their representation in the form of propositional formula [33, 34, 35]. Three of the most widely used methods in this area are Constraint Satisfaction Problem (CSP) solvers [36], propositional SATisfiability problem (SAT) solvers [28, 37], and the Binary Decision Diagrams
Hence, probabilistic feature models are ideal for representing configuration likelihood but not desirability. In many configurations of a feature model in the past, but they are not desirable for the given product description on hand. The features to be seen together in a software product rather than desirability, i.e., two features may have been seen together from existing software products in the form of Conditional Probability Tables (CPT); therefore, such tables reveal the tendency the features provide the possibility for defining hard and soft constraints [42]. The joint probability distributions are mined from existing software products in the form of Conditional Probability Tables (CPT); therefore, such tables reveal the tendency the features provide the possibility for defining hard and soft constraints [42]. The joint probability distributions are mined from existing software products in the form of Conditional Probability Tables (CPT); therefore, such tables reveal the tendency the features provide the possibility for defining hard and soft constraints [42].

Furthermore, Czarnecki et al. have developed probabilistic feature models where a set of joint probability distributions over the features provide the possibility for defining hard and soft constraints [42]. The joint probability distributions are mined from existing software products in the form of Conditional Probability Tables (CPT); therefore, such tables reveal the tendency the features provide the possibility for defining hard and soft constraints [42]. The joint probability distributions are mined from existing software products in the form of Conditional Probability Tables (CPT); therefore, such tables reveal the tendency the features provide the possibility for defining hard and soft constraints [42].

Recently, we have proposed the use of the fuzzy extension of $P(\mathcal{X})$, which is standard Propositional Logic extended with concrete domains [37] in order to capture the hard and soft constraints of the stakeholders for being able to effectively reason over the needs of the stakeholders. On this basis, we have developed a maximal covering specialization algorithm that creates a sound and complete specialization of a feature model based on stakeholders hard constraints, which is complemented by the maximal covering configuration algorithm that orders and creates a sound and complete configuration given the soft constraints of the stakeholders. The focus of the techniques developed in this paper is to achieve maximum desirability for the developed feature model configuration for the stakeholders. For this reason, stakeholders’ objectives take center place in the proposed methods. The proposed interactive feature configuration process guarantees the complete satisfaction of the hard constraints of the stakeholders by providing a consistent final configuration for cases where the stakeholders requests are consistent, and provides complete justification for why the hard constraints happened to be unsatisfiable if the requests are inconsistent themselves. $P(\mathcal{X})$ was suitable for dealing with one monolithic feature model where all the facts are known in advance and all features are fully specified. However, when dealing with multiple feature models (i.e., modules), we cannot expect that from the very beginning of the development process all feature model modules will be known and fully specified. Thus, the use of description logic and its open-world assumption supporting incomplete knowledge representation is more suitable for modular feature models.

9. Concluding Remarks

In this paper, we have proposed an approach based on the Distributed Description Logic framework to modularize software product line feature models. The proposed approach allows the independent-distributed development of individual feature modules by different teams of domain experts, and provides for the integration of these individual modules into a unique modular feature modular through the use of module bridges. The modular feature model representation enhances reusability and feature model evolution. Also, in order to develop appropriate modular feature model configurations based on the stakeholders’ requests and requirements, we have created a correspondence between feature model configuration and concept covering in Description Logics. Accordingly, we have developed a greedy algorithm that attempts to find the (near) best possible configuration of the feature model. The models and techniques provide a comprehensive framework for representing and configuring modular feature models in the Description Logic formalism. We have discussed and shown that modularity in feature models increases reusability and maintainability. It also enhances distributed and collaborative feature model design and evolution. Furthermore, our experiments have shown that as feature models become larger and more complicated, modular feature models tend to stabilize the feature model configuration execution time.

As the two case studies that we reported in the paper demonstrate, the proposed approach can be beneficial for collaboration in cases when teams need to integrate software modules with some overlapping parts or in cases when they are developing new product families where each module is about an independent sub-domain of the studied family. This can then produce many positive implication for different software engineering tasks such as system and business process integration, or leveraging legacy systems in the development of new product families. However, practical implications also open up many research questions to be explored in the future. For example, how can the proposed formalization be used in maintenance of product families. In some cases, a new sub-domain of features (i.e., new modules) needs to be developed and integrated into the existing product family. The question is then whether we need to reconfigure all already deployed product configurations, or it is better to just update them with the new feature modules, which only are to be configured accordingly. Another research question is to investigate relations between commonly used maintenance operators (e.g., update or add) and the proposed DDL-based formalization of modular feature models to automate various feature maintenance tasks.

Among others, we are interested in further pursuing the work in the paper in several main directions:

1. In this paper, we have only provided the formalism for representing modular feature models along with an appropriate algorithm that would take a modular feature model and a set of stakeholder requests as input and would generate a feature model configuration as output. One of the main directions that we intend to study in the future work is how can feature...
models be actually modularized. This is an important issue to study as it pertains to many important matters such as the suitable number of modules in a modular feature model, and the proper module for the placement of each feature. We will be exploring two options to address this either 1) by the design of a support methodology for developing modular feature models, or 2) through the development of an algorithm for automatically modularizing a monolithic feature model;

2. The module bridge graph is an abstract representation of the structure of the modular feature model that can be used to infer interesting properties of the feature model. Recently, there have been some attempts to devise appropriate feature model quality metrics to analyze the structural properties of feature models [43]. These metrics are essentially analogous to those devised as code or object-oriented design metrics [44]. We will focus on the development of such quality metrics based on the module bridge graph, which can include degrees of cohesion, coupling, maintainability, reusability and others;

3. The other aspect that we intend to carefully investigate in our future work is to empirically compare our modular feature model configuration algorithm, which is based on Description Logics with other existing feature model configuration algorithms that are based on Propositional Logic and SAT solvers. Currently the major advantage of our work over those existing methods is that it fully supports for configuration of modular feature models, whereas the other methods only work with monolithic feature models. However, we will need to also compare their performance in terms of running time.

Notes

1. In this section, ‘feature models’ and ‘modular feature models’ are used interchangeably.

2. We later discuss that developing a non-greedy algorithm is NP-hard.


References


