ARIFS: an Environment for Incomplete and Formal Specifications Reuse

Rebeca P. Díaz Redondo, José J. Pazos Arias, Ana Fernández Vilas and Belén Barragáns Martínez

Departamento de Enxeñería Telemática. Universidade de Vigo. 36200 Vigo. Spain
{rebeca,jose,avilas,belen}@det.uvigo.es

Abstract

The selection of appropriate components for satisfying a given requirement is a key problem in software reuse. Although this problem is remarkable in the reuse of software within same domain or application area, which is known as vertical reuse, it is more pronounced in horizontal reuse, that is, the reuse of software elements in different application areas. This paper describes how ARIFS tool (Approximate Retrieval of Incomplete and Formal Specifications) provides a suitable reusing environment to classify, retrieve and adapt formal and incomplete requirements specifications. Both classification and retrieval tasks are based on functional similarities according to structural closeness, which provides a suitable basis for horizontal reuse; and semantic closeness, which is more appropriated for vertical reuse. To this effect, we define four partial ordering relations among reusable components and different measures to quantify functional differences among them. By using these measures we are able to offer an approximate and efficient retrieval, without formal proofs, and to predict adaptation efforts to satisfy the required functional specification.

1 Introduction

Reusability is widely suggested to be a key to improve software development productivity and quality, but, unlike other engineering fields, software engineering has not yet developed into a mature discipline where software reuse is totally embedded. In fact, although software reuse has been practiced informally since programming was born (basically code), substantial quality and productivity payoff from reuse can only be achieved if reuse is conducted systematically and formally [14] to which there have been innumerable attempts. Besides that, reusing at early stages of the development process —like at the requirements specification stage— is accepted by many within the community as a desirable aim, because of the possibility

* Partially supported by PGIDT01PX132203PR project (Xunta de Galicia)

©2002 Published by Elsevier Science B. V.
of increasing the reuse benefits [10]. However, there is little evidence in the literature to suggest that software reuse at requirements specification stage is widely practiced. Our proposal [5,4] deals with this concern, offering a methodology to reuse high abstract level components: incomplete specifications —obtained from transient phase of an iterative and incremental requirements specification process—; and their verification results —obtained from a model checking algorithm. In this paper we focus on describing the formal basis and the methodology set up by ARIFS tool oriented to both vertical and horizontal reuse, and in [6] our proposal to reuse formal verification information in the same context is detailed.

The paper is organized as follows: next section summarizes existing works about managing reusable components; section 3 outlines the software development process where software reuse is going to be included; the methodology and objectives of ARIFS tool are explained in section 4; section 5 describes functional relationships among reusable components; the retrieval process is explained in sections 6 and 7; an example of application is detailed in section 8; in section 9 the selection and adaptation of reusable components is explained; and, finally, a brief summary and future work are exposed in section 10.

2 Related work

Organizing large collections of reusable components is one of the main lacks in software reuse, because providing efficient and effective mechanisms to classify and retrieve software elements from a repository is not an easy problem. Retrieving mechanisms usually rely on the same idea: establishing a profile or a set of component’s characterizing attributes which is used to classify and retrieve them from a repository. Whenever this profile is based on formal specifications, problems derived from natural language are avoided, and this formal description is used for establishing a specification matching. The typical process starts expressing the relation between two components by using a logical formula. Then, a theorem prover checks its validity, and only if the prover succeeds, the component is considered to be suitable. The vast number of proof tasks makes a practical implementation very hard, so in many of the following works this number is usually reduced applying other techniques.

In [16] the retrieval process of code components is based on the Larch/ML specification language for component description and the associated interactive Larch prover for retrieval. Formal proofs are restricted to a small subset of the repository, which is previously selected by using preconditions and postconditions.

REBOUND (REuse Based On UNDERstanding) tool [13] is also based in Larch language, although specification matching is based on the HOL prover, which is almost automated. In order to reduce formal proofs, different heuristics, based on the semantic of specifications, are applied. In this first step, a dynamic ordering is needed, which reduces the efficiency of the retrieval scheme.

In [3] a two-tiered hierarchy of the repository based on formal specifications using OSPL is proposed. The lower level is based on generality relationships; and
the higher one on similarity relationships, which are assessed by a clustering algo-

rithm. Firstly, it is selected the most suitable cluster, and secondly, a theorem prover

is used to finish the search; besides this, LOTOS [2] is used as a supplement to the

functional descriptions in order to take into account the architectural properties of

components.

NORA/NAMMR tool [15] is basically a filter pipeline trying to ensure a plug-in

compatibility. There are used signature matching filters, rejection filters (based on

model checking techniques), and, finally, confirmation filters (based on Setheo the-

orem prover). One of its main problems is the recursive specifications manage-

ment which are not supported by Setheo.

As applying theorem proving in the retrieval process is very difficult to au-

tomate, Fisher proposes [7] library browsing —using a special navigator— as an

alternative to library retrieval.

3 Context

In this section we briefly describe the software development process, SCTL-MUS

methodology [12], where the reusing environment is going to be included. This

methodology joins: on the one hand, the totally formalization of the process, comb-

ining different FDTs (model-oriented and property-oriented); and, on the other

hand, an incremental and iterative point of view.

In figure 1(a), the first phase of this methodology (Initial goals) is shown, where

a complete and consistent functional specification of the system is obtained from

user’s specification. In every iteration of this stage, the user identifies and specifies

a set of functional requirements which lead to a growth in the system functionality.

These requirements are verified in the current model or prototype to check: if the

model already satisfies the requirements; if it is not able to provide these functional

requirements nor in the current iteration neither in future ones (inconsistency); or, if

the system does not satisfy the requirements, but it is able to do it (incompleteness).

Formal description of functional requirements is made by using the many-

valued logic SCTL [12] (Simple Causal Temporal Logic) —in box labeled as SCTL

in figure 1(a). A generic causal requirement in SCTL follows the pattern Premise

⇒ ⊗ Consequence, which establishes a causing condition (premise); a temporal

operator determining the applicability of the cause (⇒⊗); and a condition which

is the effect (consequence). Apart from causation, SCTL is a six-valued logic,

even though it is only possible specifying three different values: possible or true

(1), non possible or false (0) and unspecified (1/2). This concept of unspecification is

specially useful to deal with both incomplete and inconsistent information obtained

by requirements capture, because although events will be true or false at the final

stage, in intermediate phases of the specification process it is possible that users

do not have enough information about them yet, so these events are unspecified in

these phases. Syntax of an SCTL requirement $R \in \mathcal{R}_{SCTL}$ is as follows:
where \( \langle \Rightarrow \square \rangle \) is the set of temporal operators; \( \theta \in \Theta := \{ \text{true} \mid \text{false} \mid \emptyset \} \) are propositional constants; and \( a \in \Lambda \) is the alphabet of events in a prototype. Temporal operators \( \{ \Rightarrow, \Rightarrow \bigcirc, \Rightarrow \bigcirc \} \) —referred to as simultaneously, previously and next— are used to reason about transition successors and predecessors of a given state by determining the order pattern between the state in which premise is formulated, and the states in the scope of the consequence.

\[
\langle \mathcal{R}_{\text{SCTL}} \rangle := R \land R \mid R \lor R \mid R \Rightarrow \square R \mid \neg R \mid a \in \Lambda \mid \theta \in \Theta
\]

\[
\langle \Rightarrow \square \rangle := \Rightarrow \mid \Rightarrow \bigcirc \mid \Rightarrow \bigcirc
\]

SCTL requirements are synthesized to obtain a model of the system by using MUS (Model of Unspecified States). This state-transition formalism allows prototyping and feedback with users—in box labeled as MUS in figure 1(a)—; and supports the consistency checking by using a model checking algorithm —in box labeled as Verification SCTL-MUS in figure 1(a). MUS graphs are based on typical labeled-transitions graph, but including another facility: unspecification of its elements.

In figure 1(b) an example of MUS graph is shown. This system evolves from one state into another when an event or an observable action from \( \Lambda = \{ a, b, c, d, e \} \) occurs. In the initial state, \( E_0 \), event \( a \) is specified as a possible one, that is, system \( g_1 \) evolves from this state into state \( E_1 \) whenever event \( a \) occurs. System \( g_1 \) evolves from \( E_0 \) into state \( E_2 \) through an event which has not been specified yet, which is denoted by \( a_{\text{unsp}} \). In subsequent iterations the user may specify this transition with a possible event from \( \Lambda \) with the exception of event \( a \), because MUS graphs are deterministic ones. In state \( E_2 \), event \( d \) is a non possible one, which is denoted by \( \neg d \), and, finally, state \( E_3 \) is a totally unspecified state because every event in \( \Lambda \) has not been specified in this state nor as a possible event neither a non possible one.

1 Unspecified events of a state are not represented, only \( a_{\text{unsp}} \) of \( E_0 \) because it implies an evolution.
The second stage of the development process, or *Refinements phase*, starts from the complete and consistent specification of the system requirements obtained in the previous stage, or *Initial goals*, and its main goal is to design the system architecture. This architecture is expressed by a constructive FDT (LOTOS [8]) which allows a description of the system components, their interactions and interfaces. Finally, *Maintenance stage* is turned into a development phase whose starting point is the MUS model of the current system.

4 ARIFS: methodology and objectives

Generally speaking, ARIFS tool provides a friendly environment to classify, retrieve and adapt reusable components in the requirements specification phase of the SCTL-MUS methodology. These reusable components gather both its functional specification, which is expressed by the set of SCTL requirements and modeled by the temporal evolution MUS graph, and an interface or *profile* information, which is automatically obtained from its functional characteristics to classify and retrieve it from the repository (section 5). Besides this, every reusable component stores verification information, that is, the set of properties which had been verified on the MUS graph and their verification results [6].

The main goal is reusing already developed MUS prototypes which are functionally close to the functionality required by the user in order to reduce, on the one hand, synthesis tasks to obtain the current prototype and, on the other hand, future verification tasks because of the verification results linked to each reusable component.

As the software process where the reuse environment is going to be included is totally formalized, we have opted by a formal specification of reusable components. Although in many of previous approaches these formal representations are only used as a pattern to recover low abstract level components (like code), in our approach, these formal specifications are just the content of the components, so we have a *content-oriented retrieval*. That is, each component is, simultaneously, index and content of the retrieval, and we can reuse high abstract level components.

As our proposal starts precisely from the necessity of minimizing verification tasks in the software development process, we do not apply formal verification in the retrieval process, which entails an exact retrieval. Instead of this, we propose an *approximate components retrieval* which allows selecting, in an efficient way, suitable components which are functionally close to the query. This approximate retrieval is based on the concept of *unspecification*, inherent to incomplete systems —which are obtained from a transient phase of the iterative and incremental development process—, that is, not everything is true or false, maybe non specified yet. So, although the recovered components do not match all the requirements of the query, it is possible matching them after making some changes. The convenience of reusing the recovered components is decided after analyzing the adaptation tasks needed in each case.

of the model.
We have also defined a two-step retrieval process: firstly a rough search phase, where a little set of suitable components is retrieved; and secondly, a more refined one, where these components are ordered depending on the adaptation efforts of each one to satisfy the functionality required by the query. The main reason of supporting a layered retrieval process is basically an efficiency one. Dividing the retrieval process in two phases, we merge the two most important tendencies in managing reusable components: static and dynamic ordering. The stiffness of the classification lattice and the big amount of information needed to reuse components in an efficient way are the main drawbacks of static management. Using dynamic management, on the other hand, implies the whole reorganization of the repository according to the functionality specified by the query, whenever a new query is proposed. In spite of its high precision, the retrieval tasks are much slower in this case than in a static ordering. Having the two-tiered retrieval process, we are able to avoid the main problem of dynamic ordering —the whole ordering of all components in the repository in each query— without losing the high accuracy that this mechanism offers.

5 Static classification of reusable components

Establishing functional relationships among components enables defining component hierarchies or lattices to classify and retrieve them in an proper way. In this section, the four functional relationships among components that we have identified are defined. Each one is characterized in terms of a function $\mathcal{O}$ that associates with every MUS graph $g$ a set $\mathcal{O}(g)$ which constitutes the observable behaviour of $g$. For every such $\mathcal{O}$, the equivalence relation $\equiv_{\mathcal{O}} \in G \times G$ is given by $g \equiv_{\mathcal{O}} g' \Leftrightarrow \mathcal{O}(g) = \mathcal{O}(g')$, and the preorder $\subseteq_{\mathcal{O}} \in G \times G$ by $g \subseteq_{\mathcal{O}} g' \Leftrightarrow \mathcal{O}(g) \subseteq \mathcal{O}(g')$, that is, $\subseteq_{\mathcal{O}}$ provides a partial order between equivalence classes or graph sets indistinguishable using $\mathcal{O}$-observations, so $(G, \subseteq_{\mathcal{O}})$ is a partially ordered set, or poset. A subset $G_1 \in G$ is called a chain if every two graphs in $G_1$ are $\mathcal{O}$-related\(^2\). Two graphs non $\mathcal{O}$-related but being in two different chains which share at least one graph are called potentially $\mathcal{O}$-related graphs.

We have defined four functions which offer four observable behaviours of a MUS graph $g$. Two of them, complete traces, denoted by $TC(g)$, and complete and non finite traces, denoted by $TC^\infty(g)$, offer semantic viewpoints of the graph. Although both of them are based on traditional complete trace semantics [1], they also take into account both true and false events in order to differentiate false events from unspecified ones.

In table 1, the results of applying these two functions to a MUS graph $g$ are shown. As it is shown in this example, the MUS graph $g$ has five different evolution ways, that is, it can evolve from initial state to a final one through event $b$ followed by a state where event $d$ is non possible; from initial state to a final one through events $a$ and $c$; from initial state to a final one through events $a$, $e$ and $c$; from initial state to a final one through event $a$, a number non determined of events $e$ and,

\(^2\) Two graphs $g$ and $g'$ are $\mathcal{O}$-related iff $g \subseteq_{\mathcal{O}} g'$ or $g' \subseteq_{\mathcal{O}} g$.  

finally, event $c$; and from initial state through event $a$ and a number infinite of events $e$. These sequences of possible and non possible events are stored in $TC^\infty(g)$ and $TC(g)$. We have to note that although trace $aeo$ is included in $a(e) + c$ trace, both of them are explicitly included in $TC^\infty(g)$ and $TC(g)$ because of efficient reasons in collating tasks. The main difference between the two observation criteria is the treatment of non finite evolution ways of the graph, in $TC^\infty(g)$ non finite traces are included —contribution $a(e) + c$ and $a(e) + c$— but in $TC(g)$ they are considered as being finite ones —contribution $aeo$ and $ae$ respectively.

The other two identified functions, number of evolutions, denoted by $NE(g)$, and number of non finite evolutions, denoted by $NE^\infty(g)$, offer structural viewpoints of the graph. In table 1 the results of applying these two functions to a MUS graph $g$ are also shown. $NE(g)$ and $NE^\infty(g)$ obtain the number of transitions that the system have to make whenever it evolves through every single evolution way. For instance when the system evolves from initial state to a final one through events $a$ and $c$, it implies the system has to make two transitions, being this number one of the contributions to $NE(g)$ and $NE^\infty(g)$, and, when the system evolves from initial state to a final one through event $a$, a number non determined of events $e$ and, finally, event $c$, it implies the system has to make at least three transitions, but they could be more because of the repetition of event $e$, which is expressed by $(3)+$ in $NE^\infty(g)$. The main difference between them is the treatment of non finite evolution ways of the graph, in $NE^\infty(g)$ non finite traces are included —contribution $(3)+$ from $a(e) + c$ trace, and $(2)+$ from $a(e) + c$ trace— but in $NE(g)$ they are considered as being finite ones —contribution $3$ and $2$ respectively.

These four results, $TC^\infty(g)$, $TC(g)$, $NE^\infty(g)$, and $NE(g)$, are automatically obtained from the MUS graph and they constitute the profile or set of characterizing attributes of a reusable component.

From each of these four functions —$TC^\infty$, $TC$, $NE^\infty$, and $NE$— a partial order ($\subseteq_{TC}$, $\subseteq_{NE^\infty}$, $\subseteq_{NE}$) and an equivalence relation ($\equiv_{TC}$, $=_{TC}$, $\equiv_{NE^\infty}$, $=_{NE}$) among MUS graph are defined. These relationships among reusable components allow organizing the repository in four different lattices, one for each partial ordering. As a result of this, each reusable component ($C$) is classified in the repository after finding its correct place in each lattice. That is, it is necessary looking for those components $O$-related\(^3\) to $C$\(^4\) such as $C$ is $O$-included on them, and those

\[\begin{align*}
TC(g) &= (b-d, a, ae, aec, aec) \\
TC^\infty(g) &= (b-d, a, a(e)+, aec, a(e) + c) \\
NE(g) &= (1, 2, 2, 3, 3) \\
NE^\infty(g) &= (1, 2, (2)+, 3, (3)+)
\end{align*}\]

Table 1

Example of the four observable behaviours of a MUS graph

\[^{3}\text{Where } O \in \{TC^\infty, TC, NE^\infty, NE\}.\]

\[^{4}\text{Two components } C \text{ and } C' \text{ are } O\text{-related (}C \subseteq_{O} C' \text{ or } C' \subseteq_{O} C\text{) iff their MUS graphs } g \text{ and } g'\]

\[\begin{align*}
TC(g) &= (b-d, a, ae, aec, aec) \\
TC^\infty(g) &= (b-d, a, a(e)+, aec, a(e) + c) \\
NE(g) &= (1, 2, 2, 3, 3) \\
NE^\infty(g) &= (1, 2, (2)+, 3, (3)+)
\end{align*}\]
components \( \mathcal{O} \)-related to \( C \) such as they are \( \mathcal{O} \)-included on \( C \). In order to eliminate superfluous reusable components connections, anti-symmetric property\(^5\) of \( \mathcal{O} \) is applied. An example of one of this lattice is shown in figure 2(a), where several reusable components are ordered according to \( NE \) criteria.

Applying the four defined relationships among graphs, we are able to distinguish between semantic similarities, which reflects if two reusable component have similar functional features (considering sequences of events), that is, if they act alike; and structural similarities, which reflects if two reusable component have similar representation of their MUS graph without considering events linked to each transition (only the skeleton of the graph), that is, if they look alike. Semantic closeness provides a suitable frame to vertical reuse, that is, the reuse of specifications within the same domain or application area. Components of these applications may partially share the alphabet of events of the graphs, even though adaptation tasks can be also needed. Structural closeness, on the other hand, provides a suitable frame to horizontal reuse, that is, the reuse of specifications in different applications, so components of these applications hardly share the alphabet of events and it would be necessary to do mapping tasks. As difference to previous works [9] which manage reusable components of code, in our case structural information is really close to semantic one. At this level of abstraction, a direct relationship between them is supported because MUS prototypes are automatically obtained from the set of SCTL requirements, therefore, it is impossible having two structurally different prototypes from the same functional requirements.

---

\(^5\) \( C_1 \subseteq_\mathcal{O} C_2 \) and \( C_2 \subseteq_\mathcal{O} C_3 \), implies \( C_1 \subseteq_\mathcal{O} C_3 \).
6 Retrieval process: rough search

In order to reuse already developed MUS prototypes which are functionally close to the functionality specified by the query, we propose replacing the box labeled as MUS in figure 1(a) by the sequence of tasks shown in figure 2(b). The retrieval process starts whenever the user expresses the required functionality by a set of SCTL requirements. From this set of SCTL requirements a set of search patterns is automatically obtained which is used to retrieve the most suitable component from the repository. These search patterns reflect the functional characteristics expressed by the query and they are the result of applying $TC^\infty$, $TC$, $NE^\infty$ and $NE$ functions to the set of SCTL requirements [4].

In the first phase of the retrieval process, we take advantage of static classification schemes of the repository to recover, in a quick way, a set of components functionally close to the query. The underlying idea of this rough search is that the closer two reusable components are classified in the repository, the more functional similarities they have. Four classification distances are used in this stage, one for each partial ordering criteria defined in section 5.

**Definition 6.1** Let $g$ and $g'$ be two potentially $O$-related MUS graphs. Then, the classification distance between $g$ and $g'$, denoted by $d_c(g, g')$, is the number of transitions which is necessary to skip on chains from the first one to the second one. The classification distance between two non potentially $O$-related MUS graph is defined as $d_c(g, g') = \infty$.

**Definition 6.2** Let $C$ be a reusable component which has been classified in a lattice according to an $O$ criteria. The set of predecessor components of $C$, denoted by $P(C) = \{C_i\}_{i=1}^n$, is such as $C_i \subseteq C$ and $d_c(C, C_i) = 1$; the set of successor components of $C$, denoted by $S(C) = \{C_i\}_{i=1}^m$, is such as $C \subseteq C_i$ and $d_c(C, C_i) = 1$; and, finally, the set of equivalent components of $C$, denoted by $E(C) = \{C_i\}_{i=1}^l$, is such as $C =_O C_i$.

This functional proximity is assessed according to structural and semantic similarities. Structural similarities predominate in $NE$ and $NE^\infty$ criteria, but, meanwhile, the first one does not take into account behaviour loops, the second one, on the contrary, considers them. As consequence, applying one criteria or the other only depends on if the query in turn specifies behaviour loops:

- If the query specifies behaviour loops, structural retrieval is based on $NE^\infty$ criteria because this criteria let us recover reusable components with a high structural similarity to the given functionality. $NE$ relationship, in this case, does not offer more accurate results and, in this way, we are able to improve retrieval efficiency without penalizing final results.
- If the query does not specify any behaviour loop, structural retrieval is based on $NE$ criteria. In this case, $NE^\infty$ does not offer a better information about reusable components than $NE$, and as comparing reusable components according to $NE^\infty$ is more complex, we are saving resources without affecting the quality of the results.
Semantic characteristics predominate on $TC$ and $TC^\infty$ criteria, but, meanwhile the first one does not take into account behaviour loops, the second one considers them. So, as we have previously described, applying one criteria or the other will be depend on the characteristics of the query. If the functionality specified by the query expresses behaviour loops, $TC^\infty$ is the criteria selected, in other case, $TC$ is the more suitable criteria. The reasons are directly extrapolated from the reasons above.

In both situations, structural and semantic closeness, ARIFS returns those reusable components successor, predecessor or equivalents to the query, according to the selected criteria ($NE$ or $NE^\infty$, and $TC$ or $TC^\infty$).

7 Retrieval process: accurate retrieval

In this second phase, we start from two different set of reusable components: those which are structurally close to the query, according to $NE$ or $NE^\infty$ criteria; and those which are semantically close to the query, according to $TC$ or $TC^\infty$ criteria. It is supposed that the components recovered in the first stage are structurally and semantically closer to the query than the remaining components in the repository, but the criteria used to recover them from the repository only allows us to maintain precedence relationships. Therefore, we have probably recovered reusable components which have notable structural and semantic differences in spite of keeping the same partial relationship to the query. The main goal of this second phase is refining the search and assessing how is the difference between these components and the query, that is, predicting the adaptation efforts to satisfy the functionality of the query; and, consequently, deciding which of them is going to be the most suitable reusable component. Each set has a different selection process —structural refinement process is detailed in section 7.1, and semantic one in section 7.2—, although both of them are consequence from a common idea: minimizing the adaptation efforts to the query.

7.1 Structural refinement process

Structural refinement process tries to find which of the retrieved reusable component has the closest skeleton to the one given by the query. In order to be able to quantify structural differences among components, we define two different measures: number of evolutions distance which is based on $NE$ criteria; and number of non finite evolutions distance which is based on $NE^\infty$ criteria. The decision of applying one or the other one only depends on the characteristics of the query, if the query has behaviour loops the number of evolutions distance is used, in other case, the number of non finite evolutions distance is used.

**Definition 7.1** Let $g$ and $g'$ be two MUS graphs, the number of evolution distance between them, denoted by $d_{NE}(g, g')$, is the Euclidean distance between two vectors that being contained in $NE(g)$ and $NE^\infty(g)$ respectively, they do not have any element in common, they have the same number of elements (in other case the smallest vector is completed with zeros), and their elements are sorted according
between. In order to know which of them is the closest one to the case, we can see that both applied to these three graphs: the first one, \( g_3 \), the second one, \( g_4 \), and the third one, \( g_5 \). The possibility of loops is not considered in number of evolutions distance (definition 7.1), so in order to take into account the functional loops specified in MUS criteria, we use the number of evolutions distance. Firstly, the \( NE \) function is applied to these three graphs:

\[
NE(g_3) = (2, 2, 1, 2, 2) \quad NE(g_4) = (2, 2, 2, 2, 3, 3) \quad NE(g_5) = (2, 2, 2, 2, 3, 3)
\]

and, then, we obtain the number of evolutions distance between \( g_3 \) and \( g_4 \) and between \( g_4 \) and \( g_7 \):

\[
d_{NE}(g_3, g_4) = d_{NE}((2, 2, 1, 2, 2), (2, 2, 2, 3, 3)) = \| (1, 2) - (3, 3) \| = \sqrt{5}
\]

\[
d_{NE}(g_4, g_7) = d_{NE}((2, 2, 2, 2, 3, 3), (2, 2, 2, 3, 3)) = \| (2) - (0) \| = 2
\]

therefore, as \( d_{NE}(g_4, g_7) < d_{NE}(g_3, g_7) \), \( g_4 \) is closer to \( g_7 \) than \( g_3 \) according to \( NE \) criteria.

The possibility of loops is not considered in number of evolutions distance (definition 7.1), so in order to take into account the functional loops specified in MUS graphs, another measure of structural differences is needed. Number of non finite evolutions distance is based on \( NE^\infty \) criteria, so it takes into account this characteristic. \( NE^\infty(g) \) can be studied as the composition of two different vectors: the first one, \( NE^\infty(g) \), is the result of applying \( NE^\infty \) criteria to every single evolution path specified in the graph which does not present any functionality loop; and the second one, \( NE^\infty_i(g) \), is the result of applying \( NE^\infty \) criteria to every single evol-

Fig. 3. \( NE \) and \( NE^\infty \) orderings

In figure 3(a) an example of lattice according to \( NE \) criteria is shown. In this case, we can see that both \( g_3 \) and \( g_6 \) are the predecessor and successor respectively of the graph \( g_7 \), so they satisfy \( d_c(g_3, g_7) = d_c(g_6, g_7) = 1 \), but \( g_3 \) and \( g_6 \) are clearly different. In order to know which of them is the closest one to \( g_7 \), according to \( NE \) criteria, we use the number of evolutions distance. Firstly, the \( NE \) function is applied to these three graphs:

\[
NE(g_3) = (2, 2, 1, 2, 2) \quad NE(g_4) = (2, 2, 2, 2, 3, 3) \quad NE(g_5) = (2, 2, 2, 2, 3, 3)
\]

and, then, we obtain the number of evolutions distance between \( g_3 \) and \( g_4 \) and between \( g_4 \) and \( g_7 \):

\[
d_{NE}(g_3, g_4) = d_{NE}((2, 2, 1, 2, 2), (2, 2, 2, 3, 3)) = \| (1, 2) - (3, 3) \| = \sqrt{5}
\]

\[
d_{NE}(g_4, g_7) = d_{NE}((2, 2, 2, 2, 3, 3), (2, 2, 2, 3, 3)) = \| (2) - (0) \| = 2
\]

therefore, as \( d_{NE}(g_4, g_7) < d_{NE}(g_3, g_7) \), \( g_4 \) is closer to \( g_7 \) than \( g_3 \) according to \( NE \) criteria.

The possibility of loops is not considered in number of evolutions distance (definition 7.1), so in order to take into account the functional loops specified in MUS graphs, another measure of structural differences is needed. Number of non finite evolutions distance is based on \( NE^\infty \) criteria, so it takes into account this characteristic. \( NE^\infty(g) \) can be studied as the composition of two different vectors: the first one, \( NE^\infty(g) \), is the result of applying \( NE^\infty \) criteria to every single evolution path specified in the graph which does not present any functionality loop; and the second one, \( NE^\infty_i(g) \), is the result of applying \( NE^\infty \) criteria to every single evo-
obtained from \( /CS /CU \) and subtracting one unit to the result.

For the closest graph to \( /CV \), criteria are obtained: between them, denoted by \( /CS /CU \), after eliminating those components which are identical in both vectors. Components of \( /CS /CU \) must be sorted according to increasing ordering.

Definition 7.2 Let \( /C6/BX \) be two MUS graphs and \( NE^\infty(g) \) and \( NE^\infty(g') \) the results of applying \( NE^\infty \) function to \( g \) and \( g' \) respectively. The structural difference vector, denoted by \( df_e(g, g') \), is obtained as follows:

- \( df_{e1} \) is composed of every component of \( NE^\infty_i(g) \) and every component of \( NE^\infty_i(g') \), after eliminating those components which are identical in both vectors. Components of \( df_{e1} \) must be sorted according to increasing ordering.
- \( df_{e2} \) is composed of every component of \( NE^\infty_i(g) \) and every component of \( NE^\infty_i(g') \), after eliminating those components which are identical in both vectors. Components of \( df_{e2} \) must be sorted according to increasing ordering.

where \( df_e(g, g') = (df_{e1}, df_{e2}) \).

Definition 7.3 Let \( g \) and \( g' \) be two MUS graphs, the number of non finite evolutions distance between them, denoted by \( d_{e\infty}(g, g') \), is the norm of the vector \( df_{e\infty} = (df_{e1}, df_{e2}') \), where \( df_{e1} \) is obtained as in definition 7.2, and \( df_{e2}' \) is obtained from \( df_{e2} \), after eliminating every single element with recursivity, \( (\cdot)^+ \), and subtracting one unit to the result.

In figure 3(b) an example of lattice according to \( NE^\infty \) criteria is shown. In this case, we can see that \( \{g_5, g_3\} \) and \( g_6 \) are predecessor and successor graphs respectively of the graph \( g_7 \), so they satisfy \( d_e(g_5, g_7) = d_e(g_3, g_7) = d_e(g_6, g_7) = 1 \), but they are clearly different. In order to know which of them is the closest one to \( g_7 \) according to \( NE^\infty \) criteria, we use the number of non finite evolutions distance. Firstly, the \( NE^\infty \) function is applied to these four graphs:

\[
NE^\infty(g_7) = (2, 2, 3, 2+, 3+) \quad NE^\infty(g_3) = (2, 2, 1, 2, 2) \\
NE^\infty(g_5) = (2, 2+) \quad NE^\infty(g_6) = (2, 2, 3, 2+, 2+, 3+) 
\]

secondly, structural difference vectors are obtained; thirdly, we obtain \( df_{e\infty} \) vector between \( S(g_7) \) and \( g_7 \); and between \( P(g_7) \) and \( g_7 \); and, finally, structural differences according to \( NE^\infty \) criteria are obtained:

\[
df_e(g_5, g_7) = (2, 3, 3+) \quad df_{e\infty}(g_5, g_7) = (2, 3, 2) \\
df_e(g_6, g_7) = (2+) \quad df_{e\infty}(g_6, g_7) = (1) \\
df_e(g_3, g_7) = (1, 2, 2, 3, 2+, 3+) \quad df_{e\infty}(g_3, g_7) = (1, 2, 2, 3, 1, 2) 
\]

\( d_{e\infty}(g_5, g_7) = \sqrt{13}, d_{e\infty}(g_6, g_7) = 1 \) and \( d_{e\infty}(g_3, g_7) = \sqrt{23} \), therefore, \( g_6 \) is the closest graph to \( g_7 \), that is, the structural adaptation effort of \( g_6 \) to \( g_7 \) according to \( NE^\infty \) criteria is the lowest one.

### 7.2 Semantic refinement process

Semantic refinement process tries to find which of the retrieved reusable components has the closest functional features to the given by the query. The refined
retrieval according to semantic closeness to the query takes into account the events which enable the transitions of the prototypes, because this information provides the semantic of the component, that is, its meaning. In order to quantify similarities and differences among reusable components, we have defined two basic functions: functional meet and functional difference.

**Definition 7.4** Given two MUS graphs $g$ and $g'$, the **functional meet** between them, denoted by $g \cap g'$, is obtained as the intersection between $TC(g)$ and $TC(g')$ or $TC^\infty(g)$ and $TC^\infty(g')$. If $g$ or $g'$ specifies some functional loop, then $g \cap g' = TC^\infty(g) \cap TC^\infty(g')$, in other case, $g \cap g' = TC(g) \cap TC(g')$.

$g \cap g'$ can be interpreted as the functional specifications (sequences of events) which are shared by both graphs and it constitutes a good starting point to assess functional similarities between two reusable components.

**Definition 7.5** Given two MUS graphs $g$ and $g'$, such as at least one of them specifies some functional loop, and satisfying $g \subseteq TC^\infty g'$, the **functional difference** between them, denoted by $g' \ominus g$, consists of those components of $TC^\infty(g')$ which are not included in $TC^\infty(g)$.

Given two MUS graphs $g$ and $g'$, such as no one of them specifies some functional loop and satisfying $g \subseteq TC g'$, the **functional difference** between them, denoted by $g' \ominus g$, consists of those components of $TC(g')$ which are not included in $TC(g)$.

$g' \ominus g$ can be interpreted as the sequences of events which have to be added to make the functional specification of $g$ the same as the functional specification of $g'$.

These two functions (definitions 7.4 and 7.5) are the basis to define five semantic refinement functions (table 2) which allow us to quantify and predict functional differences and similarities among graphs and, consequently, among reusable components.

**Definition 7.6** Let $C$ and $C'$ be two reusable components, the **functional consensus** between them, denoted by $\rho(C, C')$, is defined as $\rho(C, C') = g \cap g'$, where $g$ is the MUS graph linked to $C$ and $g'$ the MUS graph linked to $C'$. As $\rho(Q, C)$ can be interpreted as the common functionality between the current query, $Q$, and a component $C$, we use this functionality measurement to obtain the component $C_i$ which maximizes $\rho(Q, C_i)$.

Given two reusable components $C_1$ and $C_2$, the first one is closer to $Q$ than the second one, according to functional consensus, denoted by $\rho(Q, C_2) \subseteq \rho(Q, C_1)$, iff $\rho(Q, C_2) \subseteq \mathcal{O} \rho(Q, C_2)$.

**Definition 7.7** Let $C$ and $C'$ be two reusable components, the **functional deficit** of $C'$ regarding $C$, denoted by $\delta(C, C')$, is defined as $\delta(C, C') = g \ominus \rho(C, C') = g \ominus (g \cap g')$, where $g$ is the MUS graph linked to $C$ and $g'$ the MUS graph linked to $C'$. $\delta(Q, C_i)$ expresses the functional characteristics required by $Q$, the current

---

6 The relationship $\mathcal{O}$ will be $TC^\infty$ or $TC$ depending on if the query has functional loops or no, respectively.
query, which are not specified by $C_i$, that is, those lack of functional requirements of $C_i$.

Given two reusable components $C_1$ and $C_2$, the first one is closer to $Q$ according to functional deficit criteria, denoted by $\delta(Q, C_1) \subseteq \delta(Q, C_2)$, whenever $\delta(Q, C_1) \subseteq \delta(Q, C_2)$.

**Definition 7.8** Let $C$ and $C'$ be two reusable components, the **functional excess** of $C'$ regarding $C$, denoted by $\varepsilon(C, C')$, is defined as $\varepsilon(C, C') = g' \oplus \rho(C, C') = g' \oplus (g \sqcap g')$, where $g$ is the MUS graph linked to $C$ and $g'$ the MUS graph linked to $C'$. $\varepsilon(Q, C_i)$ expresses the functional characteristics specified by $C_i$ that are not required by the query $Q$, that is, those extra functional specifications of $C_i$.

Given two reusable components $C_1$ and $C_2$, the first one is closer to $Q$ according to functional excess criteria, denoted by $\varepsilon(Q, C_1) \subseteq \varepsilon(Q, C_2)$, whenever $\varepsilon(Q, C_1) \subseteq \varepsilon(Q, C_2)$.

**Definition 7.9** Let $C$ and $C'$ be two reusable components, the **functional adaptation** between them, denoted by $\Delta(C, C')$, is defined as $\Delta(C, C') = \delta(C, C') \cup \varepsilon(C, C')$, where $g$ is the MUS graph linked to $C$ and $g'$ the MUS graph linked to $C'$. $\Delta(Q, C_i)$ expresses the functional characteristics that being specified by the query $Q$ are not specified by $C_i$ and viceversa, that is, those functionality which must be added or eliminated in $C_i$ to make the functionality of $C_i$ the same to $C$.

Given two reusable components $C_1$ and $C_2$, the first one is closer to $Q$ according to functional adaptation, denoted by $\Delta(Q, C_1) \subseteq \Delta(Q, C_2)$, whenever $\Delta(Q, C_1) \subseteq \Delta(Q, C_2)$.

We are interested in retrieving the component $C$ which maximizes the functional consensus regarding the current query $Q$, and minimizes its functional adaptation:

**Definition 7.10** The **functional adjusting vector**, denoted by $\Theta(Q, C)$, takes into account both characteristics: $\Theta(Q, C) = (\rho(Q, C), \Delta(Q, C))$. Given two reusable components $C_1$ and $C_2$, the adaptation efforts of $C_1$ to make its functional characteristics equal to the current query $Q$ ones are less than the adaptation efforts of $C_2$, denoted by $\Theta(Q, C_1) \subseteq \Theta(Q, C_2)$, whenever:

$$(\Delta(Q, C_1) \subseteq \Delta(Q, C_2)) \lor ((\Delta(Q, C_1) =_{\circ} \Delta(Q, C_2)) \land (\rho(Q, C_2) \subseteq \rho(Q, C_1)))$$

7.2.1 Dynamic ordering according to the query

Given any refinement relation $\xi$ (defined in the table 2), it is possible reordering a set of reusable components $C = \{C_i\}_{i=1}^m$ regarding the current query $Q$, that is, ordering the results $\{\xi(Q, C_i)\}_{i=1}^m$ according to one of the partial ordering relations defined in 7.6, 7.7, 7.8, 7.9 or 7.10.

The graphical representation of a partial ordering relation can be simplified as follows: since the relation is understood to be reflexive, we can omit arrows from points back to themselves; since the relation is understood to be transitive, we can omit arrows between points that are connected by sequences of arrows; and when
the representation is so oriented that all arrowheads point in one direction, we can omit the arrowheads. A graphical representation of a partial ordering relation in which all arrowheads are understood to be pointing upward is also known as the Hasse diagram of the relation [11].

<table>
<thead>
<tr>
<th>Name</th>
<th>Notation</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Func. consensus</td>
<td>( \rho(C, C') )</td>
<td>( g \cap g' )</td>
</tr>
<tr>
<td>Func. deficit</td>
<td>( \delta(C, C') )</td>
<td>( g \ominus (g \cap g') )</td>
</tr>
<tr>
<td>Func. excess</td>
<td>( \varepsilon(C, C') )</td>
<td>( g' \ominus (g \cap g') )</td>
</tr>
<tr>
<td>Func. adaptation</td>
<td>( \Delta(C, C') )</td>
<td>( \delta(C, C') \cup \varepsilon(C, C') )</td>
</tr>
<tr>
<td>Func. adjusting vector</td>
<td>( \Theta(C, C') )</td>
<td>( (\rho(C, C'), \Delta(C, C')) )</td>
</tr>
</tbody>
</table>

Table 2
Semantic refinement functions

As a consequence, these new partial ordering relations can be represented by a Hasse diagram like, for instance, in figure above, where \( \xi(Q, C_1) \) and \( \xi(Q, C_6) \) are the superior values in the partial ordering relation defined by \( \xi \), and \( \xi(Q, C_7) \) the inferior one.

As it was explained in the previous subsections, these partial ordering relations can be reduced to a partial ordering relation defined by \( TC \) or \( TC^\infty \) function, whenever the query \( Q \) has some functional loop, or \( TC \) function, in other case. Because of these partial ordering relations, we can find a situation where there is not any relation among the retrieved components of the first phase of the retrieval process, so in this case, we must apply a total ordering relation among them to be able to decide among the potentially suitable components:

**Definition 7.11** Given a semantic refinement relation \( \xi \) of tabular 2, and given two reusable components \( C_1 \) and \( C_2 \) such as \( \xi(Q, C_1) \) and \( \xi(Q, C_2) \) are not \( TC \) or \( TC^\infty \)-related (according to the characteristics of \( Q \)), the first one is closer to the query \( Q \), according to \( \xi \) criteria, than the second one, iff \( z(\xi(Q, C_1)) \leq z(\xi(Q, C_2)) \). Cardinal function \( z(\xi(Q, C_i)) \) obtains the number of elements of \( \xi(Q, C_i) \).

### 8 Example of application

In this section, we show, by using a little example, all the steps of the retrieval process set up by ARIFS tool which are involved in the semantic and structural search. Starting from a set of SCTL requirements which constitutes a query \( Q \), we try to retrieve a reusable component from a given repository which minimizes the adaptation efforts, in order to avoid synthesis tasks and future formal verification ones. In this example, this set of SCTL requirements is as follows:
The first step of the retrieval process starts obtaining the search patterns of the query. As $Q$ has functional loops, the search pattern is expressed in terms of $TC^{\infty}$ function: $TC^{\infty}(Q) = (a(e)+, ac, aec, a(e) + c, bd)$ for semantic retrieval, and in terms of $NE^{\infty}$ one: $NE^{\infty}(Q) = ((2)+, 2, 3, (3)+, 2)$ for structural retrieval.

Assuming the repository consists of the six reusable components whose MUS graphs are shown in figure 4(a), the results of the rough search, according to $TC^{\infty}$ criteria, are both the set of predecessor reusable components $P(Q) = \{g_1, g_5\}$ and the set of successor ones $S(Q) = \{g_6\}$. Therefore, the set $C = P(Q) \cup S(Q)$ constitutes the starting point of the second phase of the retrieval process, that is, the semantic refined retrieval.

As we try to recover the component which minimizes the adaptation efforts to the query, the refined criteria must be the functional adjusting vector. Firstly, we obtain the $TC^{\infty}$ function applied to every reusable component in $C$:

$$TC^{\infty}(g_1) = (ac, bd)$$
$$TC^{\infty}(g_5) = (a(e)+, bd)$$
$$TC^{\infty}(g_6) = (ac, aec, a(e)+, a(e) + c, e(f)+, bd)$$

and, secondly, the vector $\Theta(Q, g_i)$ must be obtained for each one of these reusable components of $C$. The following tables store the intermediate steps to calculate these vectors and the final results.
to reduce the dynamic ordering according to the current query. Instead of reordering also the closest reusable component to the query, we are not able to conclude which of the reusable components needs less adaptation efforts to the query. Therefore, we use the functional adjusting vector relationship defined in 7.10 whose result is shown in figure 5(a). As with this partial ordering, we are not able to conclude which of the reusable components needs less adaptation efforts to the query. Therefore, $g_6$ is the reusable component which needs less adaptation efforts to $Q$.

<table>
<thead>
<tr>
<th>$g_i$</th>
<th>$\rho(Q, g_i)$</th>
<th>$\varepsilon(Q, g_i)$</th>
<th>$\delta(Q, g_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1$</td>
<td>$(ac, bd)$</td>
<td>$\emptyset$</td>
<td>$(a(c)+, a(ec, a(e) + c)$</td>
</tr>
<tr>
<td>$g_6$</td>
<td>$(ac, aec, a(e)+, a(e) + c, bd)$</td>
<td>$(e(f)+)$</td>
<td>$\emptyset$</td>
</tr>
<tr>
<td>$g_5$</td>
<td>$(a(e)+, bd)$</td>
<td>$\emptyset$</td>
<td>$(ac, acc, a(e) + c)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$g_i$</th>
<th>$\Delta(Q, g_i)$</th>
<th>$\Theta(Q, g_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1$</td>
<td>$(a(e)+, aec, a(e) + c)$</td>
<td>$((ac, bd), (a(c)+, aec, a(e) + c)$</td>
</tr>
<tr>
<td>$g_6$</td>
<td>$(e(f)+)$</td>
<td>$((ac, aec, a(e)+, a(e) + c, bd), (e(f)+))$</td>
</tr>
<tr>
<td>$g_5$</td>
<td>$(ac, acc, a(e) + c)$</td>
<td>$((a(e)+, bd), (ac, acc, a(e) + c)$</td>
</tr>
</tbody>
</table>

After obtaining the vectors $\{\Theta(Q, g_i) \ \forall g_i \in C\}$, we apply the partial ordering relationship defined in 7.11— is obtained (figure 5(b)). Therefore, $g_6$ is the reusable component which needs less adaptation efforts to $Q$.

![Fig. 5. Hasse diagrams defined by functional adjusting vector relationship](image)

The result of the rough structural search, according to $NE^\infty$ criteria, is shown in figure 4(b), where the set of predecessor reusable components $P(Q) = \{g_5, g_6\}$ and the set of successor ones $S(Q) = \{g_6\}$. Therefore, $C = P(Q) \cup S(Q)$ constitutes the starting point of the second phase of the retrieval process: the structural refined retrieval. In this second phase, we obtain the number of non finite evolutions distance between each component in $C$ and the query $Q$:

$$d_f(g_5, Q) = (2, 3, 3+) \quad d_{f\infty}(g_5, Q) = (2, 3, 2)$$
$$d_f(g_6, Q) = (2+) \quad d_{f\infty}(g_6, Q) = (1)$$
$$d_f(g_3, Q) = (1, 2, 2, 3, 2+, 3+) \quad d_{f\infty}(g_3, Q) = (1, 2, 2, 3, 1, 2)$$

As $d_{f\infty}(g_5, Q) = \sqrt{15}$, $d_{f\infty}(g_6, Q) = 1$ and $d_{f\infty}(g_3, Q) = \sqrt{23}$, therefore, $g_6$ is also the closest reusable component to the query $Q$ according to $NE^\infty$ criteria.

This little example shows that having a two-tiered retrieval process, we are able to reduce the dynamic ordering according to the current query. Instead of reordering
every single reusable component in the repository according to the functionality specified by the query, we are able to reduce this task, analyzing only the small subset retrieved in the first step of the retrieval process.

9 Selection and adaptation

Once the retrieval process has finished, the user has to select between two reusable components: the closest reusable component to the current query, from a structural viewpoint; or the closest reusable component to the query, from a semantic viewpoint. This decision must be done according to the adaptation efforts offered and predicted by ARIFS in both cases — structural distances or semantic differences.

Whenever the selected reusable component is the most semantically similar to the query, semantic adaptation tasks are needed. Firstly, increasing the functionality specified by the retrieved component \( g \) is needed in order to satisfy the functionality specified by the query \( Q \). ARIFS sets up this adaptation task starting from \( \delta(Q, g) \), which predicts this increasing effort. Secondly, it is necessary to eliminate the functionality specified by the retrieved component \( g \) which is not specified by the query \( Q \). This reduction task is predicted by \( \varepsilon(Q, g) \) which is used by ARIFS to adapt \( g \). Finally, the verification information linked to \( g \) is modified to maintain these verification results in effect after the adaptation process [6].

In the example of the previous section, we have retrieved the reusable component whose MUS graph is \( g_6 \) (figure 6(a)) and it is adapted to satisfy the requirements expressed by the query \( Q \). After eliminating the traces in \( \varepsilon(Q, g_6) = (e(f)+) \), the prototype obtained is shown in figure 6(b), which satisfies the requirements expressed by the query \( Q \).

Whenever the selected reusable component is the most structurally similar to the query, a mapping or renaming task is needed in order to adapt the semantic of the reusable component to the query’s one. This mapping is also extended to the verification information linked to the reusable component, that is, every single property which has been verified in the component is adapted to maintain its verification results in effect after the adaptation process.

10 Summary and future work

The work introduced in this paper focuses on reusing incomplete prototypes to reduce synthesis tasks in the requirements specification stage of a totally formalized,
incremental and iterative software development process. Not only synthesis tasks are reduced, but also formal verification ones, because each reusable component also stores previous formal verification results [6]. Therefore, software components from the requirements specification phase are reused, which although being more complex is also more interesting than reusing code components.

We propose a classification mechanism based on four different partial orderings among incomplete prototypes, and a two-tiered approximate retrieval where formal proofs are avoided. The first phase takes advantage of the static classification to recover a small set of potentially suitable reusable components. In the second phase, different metrics are used to predict structural adaptation efforts, and several semantic refinement functions are used to predict semantic adaptation efforts; and, consequently, to select the closest reusable component to the functionality expressed by the query.

In order to continue this proposal, we are working on the selection phase. Nowadays, the user has to decide between two reusable components, the closest one to the query according to structural similarities and the closest one to the query according to semantic similarities. At this point, the user only has the prediction of adaptation efforts, but these adaptation efforts are assessed in different terms. We propose unifying the adaptation efforts assessing the semantic adaptation tasks and the structural ones to provide this information to the user.

References


