Use Case Driven Extension of ProjectIT-RSL to Support Behavioral Concerns

David de Almeida Ferreira, Alberto Rodrigues da Silva
INESC-ID / Instituto Superior Técnico
Rua Alves Redol 9, 1000-029 Lisboa, Portugal
david.ferreira@inesc-id.pt, alberto.silva@acm.org

Abstract—Requirements Engineering is a system engineering discipline of paramount importance. Its primary deliverable is the requirements specification, a document that entails the detailed description of business-specific needs to which the target software system must comply to. Despite the advances brought by modeling techniques, the specification of software systems still consists mostly in manually writing down requirements in natural language. Resorting to natural language to convey requirements has some advantages, such as its expressiveness and the stakeholders proficiency at using it for communication purposes, when compared with semi-formal or formal languages. However, ad-hoc natural language has intrinsic characteristics that make it error-prone, some of which may even hinder the effort of complying with good requirement criteria, regardless the usage of an automated requirements tools. To address this problem, within the scope of the ProjectIT initiative, a requirements specification language named ProjectIT-RSL was developed, based on common linguistic patterns. Despite covering several structural aspects of the software system under specification (e.g., actors, components, entities, properties, and relationships), ProjectIT-RSL still lacks the support for capturing behavioral concerns.

This paper presents a Use Case Driven extension for ProjectIT-RSL, with the purpose of supporting the specification of functional requirements according to a controlled natural language approach. The proposed solution combines patterns from Use Case textual description best practices with Pseudocode specifications.

Keywords—Requirements Engineering; Specification Language; Controlled Natural Language; Use Cases; Pseudocode.

I. INTRODUCTION

The size and complexity of modern large-scale software systems demand for high-levels of abstraction, conceptual reasoning, and validation upfront. Namely, a clear definition of the real business-specific needs is required. Therefore, regardless of the achievements in both industry and academia, the development of software systems is still a rather challenging process. From all system engineering disciplines, it is argued that Requirements Engineering (RE) is one of the most important regarding this issue, since it supports the work of several other disciplines throughout the entire software product life-cycle [1]: it provides useful scope and status information to Project Management upstream and provides a stable basis for development downstream, namely for design and testing.

Typically, the development of a software system begins with Requirements Engineering (RE) [2]. RE deals with the early activities related with the process of discovering the purpose, scope, stakeholders, boundaries, and actors of the system being specified [3]. These early activities can be classified as the requirements development process, whereas the activities of evolving accepted requirements (e.g., dealing with change requests, impact analysis, tracing, and status-tracking) can be regarded as the requirements management process [4]. In short, RE is concerned with real-world goals for software systems functionality, and also how it can be precisely specified and maintained throughout the software development process [5].

However, often the importance of RE is underestimated, resulting in a large amount of rework: any attempt to start technical work beforehand, without a deep understanding of the target software system’s purpose, will certainly jeopardize the project outcome [6].

The main deliverable of RE is an artifact (usually called requirements document) that contains the detailed textual description of what the target system should do, or constraints on its behavior [7]. However, this form of specification (based on natural language) is ambiguous and, in many cases, unverifiable because of the lack of a standard machine-executable representation [7]. The main problem appears during the translation of natural language requirements into a formal or, at least, semi-formal computer model [8]. Despite these complementary approaches, natural language textual specifications are still the most suitable, fast, and preferred manner (by non-technical stakeholders) to contribute and validate requirements specifications [7].

Taking these facts into consideration, within the scope of the ProjectIT initiative [9], [10], a requirements specification language named ProjectIT-RSL [11] was developed. This initiative seeks to address some of the aforementioned problems, since it advocates that the software development process should be focused on higher-level activities (such as Project Management, Requirements Engineering, and Architectural Design), instead of repetitive and error-prone manual activities. Following this mindset, ProjectIT-RSL was designed to increase the rigor of requirements specifications to enable tool support, namely to provide user feedback.
during the specification process and the automatic generation of initial design drafts upon validated requirements.

Despite the concrete results achieved, there are still several open issues, namely the lack of support for the definition of software system behavior. To cope with real-world problems, the controlled natural language approach provided by ProjectIT-RSL needs to cover not only static aspects of the target system (e.g., domain model concepts) but also dynamic aspects related with functional requirements.

The remainder of this paper is organized as follows. Section II presents a review of related work regarding the specification of software system behavior. Section III presents the Use Case Driven extension of ProjectIT-RSL and introduce an illustrative example. Finally, Section IV concludes this paper with final thoughts and discussion.

II. RELATED WORK

Historically, the requirements specification process has consisted in creating a natural language description of what the target system should do or constraints about its behavior [4], [7]. As opposed to constructed, artificial, or any other kind of engineered languages (with specific purposes in mind), the term natural language refers to any language commonly used by humans in their daily communication. Its usage to support RE activities has already been studied for a long time [8]. However, this form of specification is both ambiguous and, in many cases, unverifiable because of the lack of a standard machine-executable representation [7].

To mitigate the undesirable effects of natural language, one can impose some restrictions to reduce its complexity while maintaining its naturalness. Controlled Natural Languages (CNL) are subsets of natural languages whose grammars and vocabularies have been engineered in order to reduce (or even eliminate) both ambiguity (to improve computational processing) and complexity (to improve readability). CNLs are typically designed for knowledge engineering, not addressing behavioral aspects. A thorough analysis on the state-of-the-art of CNLs is provided in [12]. Attempto Controlled English (ACE) [13] is a good representative of this category of textual languages, because it is a mature general-purpose CNL with a solid toolset and a wide literature body of knowledge¹. Furthermore, ACE possesses some interesting (and useful) characteristics, namely every ACE sentence is unambiguous, even if people may perceive the sentence as ambiguous in full English [14].

To address functional aspects, namely to specify complex behavior (implementation-independent algorithms and other abstract computations), one can resort to Pseudocode [15]. The high-level algorithmic descriptions achieved with Pseudocode could be integrated with the previous CNL approaches to address their weaknesses in terms of behavioral expressivity. However, Pseudocode lacks a widely accepted and well-defined standard (only some guidelines are defined [16]). We considered that PlusCal [17], [18] is a suitable representative for this approach. Despite having formal roots, PlusCal was designed having in mind some of the Pseudocode’s guidelines (e.g., simplicity), which in turn reflects on PlusCal’s design principles and syntax. PlusCal is an algorithm language that can be used to replace Pseudocode, for both sequential and concurrent algorithms [17]. Specifications in PlusCal can be automatically translated into another formal language enabling the verification of formal properties through model-checking.

Additionally, one can resort to semi-formal graphical approaches. This category of languages is well represented by UML, the general-purpose language endorsed by OMG and considered by many as the de facto standard for modeling and documenting object-oriented software. Usually, modelers use Activity and Sequence UML diagrams to specify software system behavior. However, the information conveyed by UML models is usually incomplete, informal, imprecise, and sometimes even inconsistent. These flaws are caused by the diagrams’ limitations. These diagrams cannot fully convey the details of a thorough specification.

Also, there are hybrid approaches such as Use Cases: UML Use Case diagrams can be complemented with their textual descriptions counterpart. Use Cases provide a natural approach to organize and describe the overall system functionality through descriptions, in abstract terms, of how actors interact with the system to accomplish their goals [19]. Although they are not suitable to describe functional requirements of all kinds of systems, they are adequate to specify general information systems [19]. Also, since they describe the system functionality from an actor’s perspective, they can be easier to read and understand by non-technical stakeholders than more formal or complex notations (formal methods or UML diagrams enhanced with OCL). However, although it might seem a trivial task, the relatively small number of concepts, allied with the lack of a rigid set of structuring rules (or compositional rules, when dealing with Use Case diagrams), makes people often struggle to write (or model) useful and adequate Use Cases.

Finally, there are some combined approaches. Natural Language Processing (NLP) approaches to automatically derive conceptual models from requirements texts are not new [8], and neither is the usage of formal methods to support consistency checking through mapping techniques [20]. Debnath et al. [21] propose a solution to integrate requirements written in natural language into the OMG’s Model-Driven Architecture. They defined manual derivation strategies to obtain conceptual models and formal specifications from textual requirements. Their approach follows an OCL-based transformation process to define Computer Independent Model (CIM) models from natural language oriented models. Mala et al. present research regarding Use Cases combined with NLP techniques. They present an ap-

¹http://attempto.ifi.uzh.ch/site/pubs/
A. Reusing and Structuring with Use Cases

Besides the metamodel previously mentioned, ProjectIT-RSL specifications also follow another metamodel (depicted in Figure 1) with different concerns. The former is related with the specifications’ content, whereas the latter addresses the best practices regarding requirements document’s structure. Our proposal in this paper focuses on providing an extension for the Functional Requirements section.

Given the match between our goals in terms of representation of functional requirements and the purpose of Use Cases specifications, we consider that, to avoid “reinventing the wheel”, Use Cases are the best underpinning for developing the ProjectIT-RSL extension. Use Cases employ a small set of straightforward concepts that are easily understood by non-technical stakeholders. We also benefit from a well-defined body of knowledge [28]–[30], namely how these specifications can be structured and reused in an easy to understand and industry-standard manner. Finally, bearing in mind the current ProjectIT-RSL support for domain modeling, by adopting Use Cases for conveying functional requirements, we become aligned with the early stages of the well-known ICONIX process [30]. The major pitfall of using Use Cases as a requirements specification technique is the lack of well-defined rules. Nevertheless, this problem can be addressed by providing a Use Case context-free grammar that is aligned with the metamodel presented in Figure 2.

This extension enhances the Functional Requirements section by providing a subsection named Use Cases, which contains all Use Case specifications of the (sub)system being considered. An illustrative example is presented in Listing 1. This Use Case describes a simple and common interaction with a web application, in which a registered user tries to recover a password. The textual description already provides syntax highlighting according to a context-free grammar aligned with the metamodel presented in Figure 2.
B. Complex Behavior Specification with Pseudocode

To cope with the description of complex software system behavior, we advocate the usage of Pseudocode-like descriptions. According to our proposal, these Pseudocode-like snippets act as adornments for Use Case steps. They only should be used when a fine-grain level of detail is required, namely to customize the intended action meaning or to entail the detailed specification of a complex behavior (not provided by built-in actions). The verbs used while describing detailed Use Case steps have linguistic properties (such as thematic relations) that allow us to consider them as extension points to which functions in Pseudocode can be hooked up (if the function presents the same number of arguments and constraints as the verb theme). Gottesdiener [19] provides a comprehensive list of suggested verbs for both informative Use Cases that provide information to actors (e.g., find, list, notify, select, view) and performative Use Cases that allow actors to handle complex tasks (e.g., approve, authorize, choose, send, submit, validate). By resorting to this hook up mechanism we gain an additional level of flexibility regarding the possibility of providing custom-made functional specifications for the specific needs of the project at hand.

Pseudocode often appears in the literature as a simplified programming language to describe algorithms. Regarding the Pseudocode syntax, we adopt PlusCal’s due to its benefits: we do not “reinvent the wheel” by designing a concrete syntax for Pseudocode, and we also automatically ensure compatibility with PlusCal’s formal verification tools. PlusCal’s documentation [17] already provides a context-free grammar that we reuse within the ProjectIT-RSL extension for specifying Pseudocode-like snippets.

Listing 1. Use Case example: “Recover Password”.

Following the example of Listing 1, we can illustrate the usage of Pseudocode to specify the behavior related with the verb “send” used within step 5 (“The system sends an e-mail with new password to the Anonymous User through the Mail Server”). According to VerbNet, the sentence’s main verb (“send”) has three roles (arg0 the “sender”, arg1 the “sent”, and arg2 the “sent-to”). While mapping these roles we notice that step 5 contains additional information provided by a prepositional sentence (“through the Mail Server”). This piece of information will not be considered while hooking up the verb arguments with the built-in actions, unless we introduce a Pseudocode extension. We can specify a custom-made Pseudocode snippet (illustrated in Listing 2) to support this additional verb role (arg3 the “by-means”).

Listing 2. “Send-To-By” Pseudocode specification example.

Thematic relation is a linguistic term used to express a syntax–semantic correlation, namely the roles that a noun phrase plays with respect to the action or state described by a sentence’s main verb.
C. Parsing of ProjectIT-RSL Extension

This extension of ProjectIT-RSL resorts to different parsing techniques, being the pipeline divided in two different stages. First, we make use of a structural template, derived from Use Case textual description best practices [28]–[30]. This derived template is processed with a context-free grammar similar to those used by programming languages. Next, after the structural information has been parsed, the process follows to the second stage of the pipeline. At this stage we make use of NLP techniques, namely shallow parsing\(^3\) of the pieces of information previously captured according to its context within the structural template. The extracted information, namely from detailed Use Case steps, will populate models aligned with the ProjectIT-RSL metamodel. Pseudocode is used to specify complex behavior not covered by built-in actions associated with the verbs used while describing detailed Use Case steps.

We resort to shallow parsing techniques to address complexity of natural language, which hinders the possibility of building a general-purpose representations of meaning from ad-hoc natural language text [32]. To ease the burden of full parsing, we only look for very specific kinds of information in detailed Use Case steps, namely to populate models according to the “Actor performs Operation on Entity” core structure emphasized by the ProjectIT-RSL metamodel. Our shallow parsing approach can be considered as Chunking [32, Ch. 7, p. 261], more specifically a regular expression cascaded chunker. The possibility of adding new parsing rules that express linguistic patterns, namely to extract specific and contextualized pieces of information from detailed Use Case steps, makes the language extensible.

IV. CONCLUSION

Requirements Engineering is a mature discipline, and throughout the last decades several approaches to improve its processes and artifacts have been proposed. Yet, there are still open issues to address, namely problems regarding communication and requirements quality. The ProjectIT initiative advocates that a greater attention should be given to the rigor and quality of requirements specifications. Namely, it emphasizes the need for tool support, because most RE techniques are manually applied, thus making them error-prone and less cost-effective when dealing with modern software systems, due to their size and complexity. Since requirements provide the foundation for most of subsequent technical work, these activities would certainly benefit from consistency checking of the requirements specifications. Also, unless suitable tool support is provided, to validate and generate an early design draft from requirements specifications, it will be difficult to seamlessly integrate RE with the Model-Driven Engineering paradigm.

Considering these facts, this paper presents an extension for the ProjectIT-RSL language, to address its current lack of support for the specification of behavioral aspects. To convey information about the software system behavior with ProjectIT-RSL, namely to specify functional requirements, we propose an integration of well-known techniques, more specifically textual Use Cases with Pseudocode specifications. The former is an industry-standard technique that provides a template for structuring functional requirements, thus reducing the complexity of natural language parsing with well-defined contexts. The latter provides a technology-agnostic approach to express complex behavior (the computations associated with the customized-actions, i.e., the verbs used within Use Case steps). This extension complements the current ProjectIT-RSL support for domain modeling and, although it is process-independent, it can be well supported by processes such as ICONIX or RUP.

As future work, in the short-term we plan to validate the ProjectIT-RSL language with real-world case studies. First, we are going to develop the required tool support by extending the current ProjectIT toolset, namely the parsing algorithms and text editor. In the long run, we plan to use the requirements models entailed within ProjectIT-RSL to infer or verify properties of the target system (e.g., model-checking or theorem proving) in order to ensure the overall quality of its specification.

REFERENCES


---

\(^3\)As opposed to full parsing, where one or more complete parsing tree are derived, shallow parsing techniques try to identify and extract information from relations between words.