Abstract

The use of Natural Language (NL) in software requirement specification (SRS) documents introduces inconsistency. Numerous tools and methods are available for managing requirements. However, there are few procedures and tools that provide support for analyzing inconsistency in SRS textual documents. Detecting inconsistencies in SRS document is a challenging task that has spark the interest of researchers. The primary methods used to identify inconsistencies in NL SRS document are reviews and inspections. However, the application of human labor is time consuming, ineffective and introduces difficulties.

This paper describes a framework for identifying inconsistency in SRS documents. The proposed tool incorporates a knowledge base system that integrates the semantic and syntactic analysis of requirements.

CCS Concepts:
- Software and its Engineering → Software Functional Properties
- Correctness → Consistency

Keywords
- Inspection, inconsistency
1. Introduction

The first stage of the software development life cycle (SDLC) is the requirement phase. This stage involves eliciting requirements from stakeholders in natural language (NL), analyzing requirements, negotiating and revising requirements before all stakeholders sign off on the final document. That is, the software requirement specification document, SRS [31]

Requirement elicitation, the process of gathering and combining systems functionalities from stakeholders is the first phase of the requirement stage and has a direct impact on the quality of the software produced [13] [14] [16]. It is imperative at this stage that intensive human communication occurs in order to gather the correct requirements [15]. Subsequently a software requirement specification document is developed. This document captures all the required aspects of the system including functional requirements. At one end of the pendulum, stakeholders should be able to specify their needs as it relates to the system under development. On the other end, there must be some automated inspection procedure in place to check the SRS document for inconsistencies. Because SRS is the fundamental building block that guides the development process and this document should be free of defects such as ambiguity, incorrectness and must be analyzed as it pertains to consistency, completeness and correctness before all stakeholders sign off in agreement with its content [17]. To add, according to authors Xowghi and Gervasi addressing consistencies in SRS document directly has an impact on reducing incompleteness and incorrectness [18]. Nelson and his co-authors discovered that software developers design systems that contain embedded errors because these systems are dependent on the system requirements [32]. Errors in requirements are expensive to fix, increase development time due to rework and result in maintenance issues. According to Nelson:

“A software error costing a mere $1 when caught early in the life cycle, cost $5 to correct at midpoint and $100 to correct later in the life cycle”

Based on Nelson’s statement it is imperative that early detection of inconsistency is done in order to reduce overall SDLC related cost as well as to improve product quality.

As the utilization of software continues to proliferate in today’s society, software quality is becoming increasingly a paramount issue. As a matter of fact, with the exponential increase in software applications, inspection has equally grown as a vital activity for the purpose of validating requirement specification documents. Software inspection has been in existence for over three decades to ensure specifications and domain properties are equivalent to the requirements thus, Broy and Denert (2002) rightly stated:

“Inspections are now 30 years old and they continue to improve software quality and maintainability, reduce time to delivery and lower development cost”

Software engineering literature consistently designates inspection and testing as two widely renowned activities for software quality improvement. However, while both methods are used for defect detection (such as inconsistencies, ambiguity, and incompleteness) and removal in software products, testing can only be done when the software has been implemented. On the other hand, inspection (manual or automated) is the only method that can be applied in the early stage of the software development to prevent rework of application requirements, design and code by finding defects and removing them [7] [8] [9] [10].

Even though inspections have been utilized for more than three decades, it is quiet alarming that software is still being released with defects [1]. For that reason, further innovative research is needed to discover more real-world, simple, effective and automatable inspection methods [12].

James Martin (1984) stated, “The root cause of 56 percent of all defects identified in software projects are introduced during the software requirement stage of the development life cycle” [6]. This may be because requirements are gathered through NL [3]. The possibility of dissimilarities in interpretation and understanding by various stakeholders can expeditiously lead to defects such as ambiguity, inconsistency, incorrectness, omission and superfluous information, just to name a few [1]. Hence, according to Alshazly et al., it is vital to detect these defects in the interest of conserving resources and the conformance to requirements. In addition, it is necessary to uncover defects to preserve software quality and reliability by removing bugs [1]. Kamalrudin also noted that one of the major contribution to imprecision or defects in software requirement specification document is the use of natural language [2]. According to Fabbrini (2001) et al, SRS documents tend to contain errors because of the use of NL when translating requirements from stakeholders [4].

Numerous documented research focus on identifying inconsistencies in software requirements whether the model use is formal (example Zed language / Z notation) or semi-formal (example NL). For instance, Grant et al, used Z notation specification to validate and verify functional requirements for safety critical systems [5]. However, most of the studies ignore or does not define an automated procedure for inspecting textual document for inconsistencies.

Therefore, our research set out to define a procedure to identify inconsistencies in software requirement specification that can be automated for future studies, with particular emphasis on the functional requirements. Although there are many definitions for Inconsistency, in the context of this work, we will utilize the description of Inconsistent as defined by Alshazly et al., to be:

*Any part of the functional requirements document that is inconstant with other related functional requirements of the same type, structure or with the problem that the SRS artifacts solves [1]. Inconstant with regards to use of words, terminology, and internal logics [25] [26] [27].*
This study is motivated by results of prior research whose limitations spark the need to develop a new methodology to detect inconsistency. Our work takes into account some properties of the studies reviewed as well as built on their limitation to create an improved approach.

2. Related Works

In this segment, we presents previous studies on identifying inconsistencies in SRS document.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Approach</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kamalrudin [2]</td>
<td>Develop an automated tool for detecting inconsistencies between textual SRS and use case (use case are generated from interactions list)</td>
<td>The evaluation of the study concluded that the researchers were able to trace natural language requirements to a set of abstract interactions. Interactions (use case) are a compilation of phases extracted from the natural language. Nonetheless, the tool was not able to exhaust all possible forms of inconsistencies, since traceability is difficult [20]. The tool requires more work to be done in order to fully address the need to reduce inconsistencies in requirements. For example, the need to have a library to extend the opportunity of locating consistent interactions.</td>
</tr>
<tr>
<td>XLinkit [19]</td>
<td>First order logics, object Z specifications, specification tests, model abstraction and model checking to verify requirements. In order to detect inconsistencies goal elaboration, order abduction and morphing of path (knowledge &amp; rule base)</td>
<td>The tool was able to manage consistency between software artefacts generated at each stage of the software development life cycle.</td>
</tr>
<tr>
<td>Sugimoto et al. [21]</td>
<td>Requirement Framework Model to detect inconsistency of SRS. Application of the Dempster and Shafer’s theory was applied to interpret the inconsistency.</td>
<td>Researchers were able to locate 2 inconsistent requirement sentences of 46 sentences</td>
</tr>
<tr>
<td>Koth et al [28]</td>
<td>Heuristics Approach with specification requirement and semantics</td>
<td>The use of semantic checker as an incremental evaluation approach efficiency was improved. That is, the methodology prevents the re-evaluation of the entire XML document when new semantics information is added about attributes, rather it uses the previous saved semantics computation to make comparison to the new modification.</td>
</tr>
</tbody>
</table>

Table 1: Summary of Approaches to Detecting Inconsistencies in Software Requirement Specification Document
In an effort to document the gap that exist in detecting inconsistencies in software requirements Kamalrundin noted a deficiency in the inconsistencies tools literatures, in terms of the need to assess different representation and check for inconsistencies between scenario and textual descriptions [2]. The researcher proposed a consistency management and tracing tool within the Eclipse-based Marama meta-development environment that is able to trace requirements back to their design representation, particularly the use case diagram [2]. In the authors’ analysis, the innovative tool should assist in correcting requirements during the translation process by identifying inconsistencies between the NL requirements and interactions. The proposed methodology incorporated four steps. (1) Collection of NL requirements (2) Analysis is performed on NL requirements using a database comprising of use case interactions (3) use case models are generated and (4) items are selected from the use case and compare to the NL [2]. While the author focused on scrutinizing inconsistencies of requirements, yet again another study only looks on keeping requirements consistent between informal (NL) and semi-formal (use case) requirements. Moreover, no mention was made on the protocols that govern identifying appropriate generated use case models to compare to NL.

Kozlenkov et al developed a knowledge based tool with the capability of ensuring consistency between different artefacts produced at different stages in the software development life cycle with main focus on large and complex systems. In addition, the instrument ensures that the system under development meets the functional requirements [19]. One of the significant shortcoming of the study is the use of formal specifications. In order for users to make use of formal specifications they need to have a detail knowledge of the modeling language or have the language explain to them continuously. In addition, while the researchers were busy checking for inconsistencies between models, there has been no method incorporated in their tool to first check for inconsistencies within the SRS document. Since all models are based on the SRS, which is the driving force of the system, it is very necessary and imperative that this document is thoroughly check for such aforementioned defects. So as to prevent any carrying over of defects (inconsistency) due to the use of NL.

Other researchers utilize techniques such as domain ontology and ontology Library Management System to detect and correct inconsistencies between NL and other analysis and design representation [23] [24].

In 1999 Sugimoto proposed a static technique to locate and interpret inconsistency of SRS documents [21]. The study defined inconsistencies in two forms; 1. Illegal use of words and 2. When two or more requirement sentences that define the same data structure differs. The authors implement their methodology my making use of the requirement frame model to detect inconsistencies of SRS [21]. The operation of the requirement frame model integrates the following concepts:

- Identifying objects (verbs) and object types (attributes)
- Define operations among objects and role of the operations (cases) where cases represent concepts about agents, objects, goals of operations.

Based on the concepts, three major frames were formed [21] [22].

- A noun level frame
A sentence level frame
A functional frame

In order to determine the feasibility of the technique the authors developed a text-base requirement language name X-JRDL that is based on the requirement frame model. The purpose of X-JRDL dictionary is to analyze and simplify sentences that is transferrable into CRD (conceptual requirement description). X-JRDL description (words such as pronouns, verbs and adjectives) is analyzed using three interpreters. Verbs are further categorized into different concepts to find illegal use of words which the researchers categorized as a form of inconsistencies. Subsequently the defined procedure was applied to a stock control system. Of the 46 sentences in the SRS two (2) inconsistencies were identified [21]. While giving credit to the researchers for developing such novel approach the study falls short in the following aspects:

- It is difficult to prove the successfulness of the approach if readers are unable to identify the actual number of inconsistencies in the document use to support the study.
- The dictionary was written in Japanese language, hence that limits the ability to accept the study in a general scenario.
- The dictionary excludes the use of noun instead pronouns are incorporated in the X-JRDL. This is undesirable since a SRS document and any translation thereof must be specific about the actors who should perform particular requirements as define in the SRS document. Not being specific about who or what should perform system functionality may lead to yet another form of inconsistency in the SRS.
- The methodology requires automation in order to further identify any anomaly in the proposed procedure and enhance the technique.

On the other hand, the results additionally provided insights into how the proposed method can be ameliorated for automation.

Koth et al., [28] developed a methodology based heuristic analysis approach coupled with requirement specifications and semantics. This approach was selected by the researchers because in comparison with its counterpart formal analysis, heuristic analysis does not require the structure of mathematical model for making decisions [29] and is capable of specifying the steps for achieving required goal [30]. In addition, this is the most commonly used technique for identifying consistencies or inconsistencies in requirements. The semantic incremental attribute evaluation technique was developed and used on Extensible Markup Language (XML) document. This approach introduces incremental facilities and evaluates the attributes associated to XML semantics by adding incremental strategy to XML semantic evaluator [28]. In order to check the consistency of documents repeatedly until a consistent document is produced the Propagate Algorithm was also employed [28].

Gnesi et al. [33] developed and implemented a tool so called Quality Analyzer for Requirement Specification (QuARS) for analyzing lexical and syntax in software requirement documents. The function of the tool is to extract structured information and metrics for detecting linguistic inaccuracies and defects that can result in ambiguity at the later phases of the software development process. The lexical component looks on a word
in a language (example strong) while the syntactic component concentrates on the arrangement of words in a sentence (example, the users should enter website with password. – This can be interpreted as the website is only accessible when a user enters a password or that a website that has a password is only accessible to the user. Because QuARS is limited to defect identification and readability an updated version of QuARS was released named QuARS Express by Bucchiarone et al that incorporate improvement on defect identification and readability analysis [34].

3. Proposed Approach

We are proposing a methodical approach to our new hybrid technique. Figure 1 shows our proposed approach. The process of identifying inconsistencies in the SRS document will begin with comparing the requirements of each bucket to the consistency attributes or rules stored in the knowledge base database from a semantic and syntactic point of view. The semantic looks on the meaning of words in a language. While the syntactic concentrates on the analysis of requirement statement constructs. Once a rule has been broken, the tag for that requirement will be recorded. Subsequently, a log file will summarize each rule that has been broken that is associated with a particular requirement.

---

**Figure 1: Proposed Approach**

Our proposed technique involves the following steps:

- Loading a textual SRS document to the system.
- Grouping related requirements into different buckets. The system should be capable of recognizing requirement ID such as 1.1.a. The goal of grouping relate items into different buckets is to facilitate effective and efficient analysis of the data as well as to reduce re-evaluation processing time. In addition, this will help to record analysis result such that all inconsistency rule pertaining to a requirement appears together.
- Tag each requirement in a bucket: - A suffix associated with a bucket will be added to the tag of each requirements. The goal is to refine the scope of re-evaluation of the SRS document after modification. For example, if a change is
made to a particular section of the SRS document, only that specified bucket will be evaluated after the requirement had been added to the bucket rather than all buckets. Also, tagging each requirement provides identification of consistency rule violation traceability from the machine learning knowledge based system to the initial requirement.

- **Parse Requirements:** - Requirement will be parsed from both a syntactic and semantic standpoint. As outlined earlier on the syntax component will focus on the construct of a sentence. Syntactic parsing will highlight the part of speech such as verb or noun of a word in a sentence as well as the role of each word. Meanwhile the semantic parsing of the document will identify specific words and phrases in sentences of the SRS document.

- In our approach we put forward the need for inclusion of a machine learning knowledge base system (MLKBS). So that the system will be able to draw upon the knowledge of human expert to identify inconsistency issues that normally require human competence. The knowledge base system (MLKBS) will be directly identifying and locating inconsistencies by making use of machine learning systems with Meta data for storing all the information related to requirements and rules.
  - The MLKBS incorporate a comprehensive Meta dictionary which is a collection of facts about the system’s domain within the database to check for verbs, nouns etc.
  - The MLKBS should be able to provide reasoning about information in the knowledge base by identify sentences and performing comparison amongst sentences syntax and semantics rules.
  - The system must be able to store values for check attributes to reduce validation time when modification is done to the SRS document.

The machine learning system will make use of all the stored data in Meta data for further data mining and automated reasoning.

```java
// Identifying Inconsistencies in SRS Document Algorithm
load_doc();
for each requirement
{
    group_related_req();
    add_to_bucket();
}
add_tag();
parses_req();
while (bucket >-1)
{
    select_bucket();
    analyze_sem_view();
    con_attributes();
    analyze_sen_view();
    con_attributes();
}
gen_log_file();
```

**Figure 2:** Algorithm for Identifying Inconsistency in SRS document
Number of inconsistency identify (NoI) = \( \sum_{i=1}^{n} b \)

Where \( b \) represents buckets
Where \( n \) represents number of buckets

Figure 3: Mapping a Requirement to Consistency Rules in the Knowledge Base System

Figure 3 illustrates how each requirement will be compared to the consistency rules stored in the library of the KBs.

4. Case Study

The case study below shows from an abstract level how the automated system should function. We assume the development of a website that allows users to make purchasing of items online. Our customer requires that only users with authorized access – user name and password are allowed to use the website. We are assuming that our customer means well when stating the needs of the business. Initially, we do not have a clear knowledge about the domain, table 2 and table 3 shows our beginning requirements and consistency attributes.

<table>
<thead>
<tr>
<th>Requirements ID</th>
<th>Requirement Description</th>
<th>Requirement tag Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Req 1.0</td>
<td>The user shall enter website with a user name and password</td>
<td>Req_1(&lt;b1&gt;)</td>
</tr>
<tr>
<td>Req 2.0</td>
<td>The software shall support database entry</td>
<td>Req_2(&lt;b1&gt;)</td>
</tr>
<tr>
<td>Req 3.0</td>
<td>Accessing the website allows purchasing</td>
<td>Req_3(&lt;b1&gt;)</td>
</tr>
<tr>
<td>Req 4.0</td>
<td>Users who do not have a user name and password are denied access to the website</td>
<td>Req_4(&lt;b1&gt;)</td>
</tr>
<tr>
<td>Req 5.0</td>
<td>Users whose user name and password appears in the database shall be authenticated</td>
<td>Req_5(&lt;b1&gt;)</td>
</tr>
<tr>
<td>Req 6.0</td>
<td>Password must be strong containing letters (upper and lower case), numbers and special characters</td>
<td>Req_6(&lt;b1&gt;)</td>
</tr>
</tbody>
</table>

Table 2: Beginning Online System Requirement
Where \( b_1 \) represents bucket one which is the suffix for each tag

<table>
<thead>
<tr>
<th>Requirement Tag Number</th>
<th>Consistency Attribute ID</th>
<th>Consistency Attributes (syntactic and semantic Rules)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Req_1(&lt;b_1&gt;)</td>
<td>CA1</td>
<td>If user enter user name and password then allow user access to the website</td>
</tr>
<tr>
<td>Req_2(&lt;b_1&gt;)</td>
<td>CA2</td>
<td>If database exist then user data shall be stores in the database</td>
</tr>
<tr>
<td>Req_3(&lt;b_1&gt;)</td>
<td>CA3</td>
<td>If user is authenticated then user will be allowed to purchase</td>
</tr>
<tr>
<td>Req_4(&lt;b_1&gt;)</td>
<td>CA4</td>
<td>If user does not have a user name or password then user will be prevented from entering the website</td>
</tr>
<tr>
<td>Req_5(&lt;b_1&gt;)</td>
<td>CA5</td>
<td>If user name and password exists in the database then the user is valid</td>
</tr>
<tr>
<td>Req_6(&lt;b_1&gt;)</td>
<td>CA6</td>
<td>If password contains upper and lower case letters, numbers and special characters then it is strong</td>
</tr>
</tbody>
</table>

Table 3: Consistency Attributes

For demonstration purpose, we are only using one bucket (\( b_1 \)). Each requirement will be compared to all the constraint attributes that exist in the knowledge base system.

<table>
<thead>
<tr>
<th>Requirement Tag Number</th>
<th>Consistency Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Req_1(&lt;b_1&gt;)</td>
<td>Y</td>
</tr>
<tr>
<td>Req_2(&lt;b_1&gt;)</td>
<td>N</td>
</tr>
<tr>
<td>Req_3(&lt;b_1&gt;)</td>
<td>N</td>
</tr>
<tr>
<td>Req_4(&lt;b_1&gt;)</td>
<td>N</td>
</tr>
<tr>
<td>Req_5(&lt;b_1&gt;)</td>
<td>N</td>
</tr>
<tr>
<td>Req_6(&lt;b_1&gt;)</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 4: Identifying Inconsistency Result Table

<table>
<thead>
<tr>
<th>Log File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Req_1(&lt;b_1&gt;) is inconsistency with CA2, CA3, CA4, CA5, CA6</td>
</tr>
<tr>
<td>Req_2(&lt;b_1&gt;) is inconsistency with CA1, CA3, CA4, CA5, CA6</td>
</tr>
<tr>
<td>Req_3(&lt;b_1&gt;) is inconsistency with CA1, CA2, CA4, CA5, CA6</td>
</tr>
<tr>
<td>Req_4(&lt;b_1&gt;) is inconsistency with CA1, CA2, CA3, CA5, CA6</td>
</tr>
<tr>
<td>Req_5(&lt;b_1&gt;) is inconsistency with CA2, CA3, CA4, CA5</td>
</tr>
<tr>
<td>Req_6(&lt;b_1&gt;) is inconsistency with CA1, CA2, CA3, CA4, CA5</td>
</tr>
</tbody>
</table>

Not in \( b_1 = 29 \)

Table 5: Resulting Log File
5. Conclusion And Future Work

We defined a new approach for identifying inconsistencies in SRS document. This approach has taken into consideration strengths of previous study such as Gnesi et al, Sugimoto and Koth et al. It builds on those systems techniques shortcomings and it incorporates the use of knowledge based system. Future work is required to build the system or find means to integrate systems together that will identify inconsistency in SRS document as outlined by this approach.

6. References


