Enhancing Understandability of Objective C Programs Using Naming Convention Checking Framework

Ruchuta Nundhapana and Twittie Senivongse

Abstract—Understandability is a software characteristic that helps ease software maintenance and evolution. When modifying or reusing software that is written by someone else, software developers often have difficulties in trying to understand what the existing software does and how. Such an issue is commonly found in software-developing organizations. This paper discusses an approach taken by an IT organization in Thailand which attempts to enforce coding standards within its iOS development team in order to promote software understandability and maintainability. Among coding standards, naming conventions are important but are most often violated. This paper presents the development of a naming convention checking framework that consists of tools to automatically detect naming convention violations in Objective C programs. The framework facilitates iOS developers in modifying the programs so that they adhere to the naming conventions. An experiment showed that the developers’ understanding in the programs that had been modified, as suggested by the naming convention checking framework, did improve at a statistical significance level of 0.05. This approach can enhance program understandability and can be applied to other software-developing organizations.

Index Terms—naming convention, Objective C, maintainability, understandability

I. INTRODUCTION

UNDERSTANDABILITY is a software characteristic that helps ease software maintenance and evolution. It is always the case that software has to undergo change to fix errors, to handle new or changed user requirements or software environments, or to prevent future problems. This requires software developers to inspect code of existing software and try to understand what functions it performs and how. According to Boehm’s quality model [1], code possesses understandability characteristic to the extent that its purpose is clear to the inspector. To make the purpose clear and understandable, there must be consistency, structuredness, conciseness, and legibility in the code. Understandability is a subattribute of maintainability.

In software maintenance and evolution, it is likely that software developers have to inspect and understand code that is written by someone else. The original developers may have been transferred to other software projects, have changed positions or jobs, or have retired. It is common that software that is in use today within organizations was developed long time ago and all details might have been forgotten. Software developers have to carefully study the code in order to perform maintenance tasks.

It is often the case that software within organizations may be written using different styles and conventions. This is because organization-wide coding guidelines may not be in place, or developers have different levels of experience or are not strict with coding conventions of the computer languages of use. Software developers should be concerned with naming identifiers. Naming variables, constants, methods, functions, and classes, for example, should follow the organization’s guidelines or coding standards of the languages. Identifiers should convey meaning of what a program does and what data are used or produced. Having various naming styles for identifiers makes it difficult for software developers, who may themselves use different styles, to understand the code written by other developers.

This paper presents a case of an IT organization in Thailand. While beginning to move to Swift development, its iOS application development team have been maintaining a number of Objective C applications, adding new business requirements or modifying them when the operating system is upgraded. Initially, the team did not enforce any coding standards and they usually had to spend time trying to understand the code written by former team members. In some cases, the developers found that it was difficult to understand existing code and decided to develop the whole application anew. However, redevelopment took time and some original requirements might have been missed out, making the newly developed applications incomplete. Therefore, the iOS development team recently began to enforce coding standards. The team gathered Objective C coding guidelines that are recommended by Apple [2] and by other sources [3], [4], [5], with an emphasis on naming convention as well as the use of magic numbers [6], [7] and literal strings [6], [8]. Also adding their own guidelines, the team established an Objective C naming guideline that was used in the development of a naming convention checking framework. Using the tools in the framework, the team could automatically check where naming conventions were violated in the existing Objective C applications. The team...
could revise the code to adhere to the guideline in order to improve understandability and ease future maintenance.

This paper is organized as follows. Section II discusses related work. Sections III and IV present the Objective C naming guideline and the development of the naming convention checking framework. An evaluation of the framework is shown in Section V and the paper concludes in Section VI.

II. RELATED WORK

Related work was reviewed in two aspects, i.e. coding conventions and automatic checking tools.

On coding conventions, Smit et al. [6] suggested that coding conventions may have impact on maintainability of software. For example, the use of magic numbers (i.e. literal values that appear in a program) and hard coded strings could make the code difficult to read, understand, and maintain. They conducted a survey of software engineers to identify the relative importance of 71 coding conventions to maintainability and measured the convention adherence of four open-source Java projects. The result showed that the most common violations were related to the use of magic numbers and multiple literal strings as well as naming. Butler et al. [9] argued that automatic checking of naming conventions was limited to checking of typography. They proposed a naming convention checking library for Java, and match them with identifier naming conventions, i.e. Camel, Pascal, Hungarian, Underline, and Capital. The result showed that Camel was used the most frequently in these languages, and Java projects had the highest consistency in the use of naming conventions, followed by C and C++ projects.

On automatic coding convention checking tools, Objective Clean [11] can check coding styles of Objective C programs. A developer first has to take a survey to define the rules about the coding styles that are to be applied to a project. The rules are about usage of a space in a statement, method parameter prefix, brace, and empty line only, i.e. it is not for checking naming convention. At the end of the survey, a configuration file is created, and the tool can be used to set up the configuration file within a project. Then developing the project on the Xcode IDE can adhere to the rules. At build time, if the code violates one of the rules, then Xcode will throw a build error and identify the offending line. Another tool called Faux Pas [12] inspects iOS or Mac app’s Xcode projects and warns about possible bugs as well as about maintainability and style issues. With regard to naming convention, its Unidiomatic Accessor Naming rule produces a warning if the name of a getter method starts with “get”, while the Identifier Naming rule allows enforcing custom naming guidelines for different kinds of identifiers via regular expressions.

Like Objective Clean [11], this paper presents a coding convention checking tools and framework that can set up convention rules in an Xcode project and, during code building, can identify the locations within the code which offend the rules. Unlike Objective Clean and other tools, this paper focuses on comprehensive checking of identifier naming and the use of magic numbers and literal strings as they are the most violated conventions [6].

III. OBJECTIVE C NAMING GUIDELINE

In the case of an IT organization in Thailand, its iOS development team compiled an Objective C naming guideline shown in Table I which would be used as a standard in the team. The conventions in the guideline were taken mainly from the coding guidelines for Apples’ Cocoa framework [2] which recommend general naming conventions, how to name classes, methods, functions, properties, variables, and constants, as well as acceptable abbreviations and acronyms. The guideline lists correct naming and wrong naming as examples of recommendations and violations respectively. Also, naming conventions recommended by other sources [3], [4], [5] were included. In addition, the team themselves added four conventions to this list. Among those four were the conventions about magic numbers and literal strings. A magic number is a numeric literal value buried in the code instead and should be avoided (except -1, 0, 1, and 2) [6], [7]. For example, the value 56.0 appears out of the blue in area = width * 56.0, and it is unclear what it means. Literal strings is a series of characters enclosed in double quotes [8], e.g. mail.sender = “abc@gmail.com”. Especially multiple occurrences of the same string in the code would require change in all locations if the string pattern has to change [6]. It is better to replace magic numbers and literal strings with named constants.

IV. NAMING CONVENTION CHECKING FRAMEWORK

The overview of the naming convention checking framework for checking adherence to naming conventions of any Objective C program in an Xcode project is depicted in Fig. 1. Steps in the framework are as follows.

A. Extract Identifier Names

The framework provides a “naming list library” (i.e. RNNamingListObject.h and RNNamingListObject.m) which an iOS development team member has to import into the project and call the library by using the command [RNNamingListObject startGetListName] in the file AppDelegate.m and run. The library calls Objective C runtime methods of Cocoa (Touch) framework to get identifier names (i.e. class, method, function, variable, property, and constant names) from the program. The result is a naming list in a text file, where each entry in the file shows [identifier name, type of identifier]. This naming list is used later by the Objective C convention checker tool.

B. Set Configuration

The rest of the framework is supported by an OS X application called the “Objective C convention checker.” The developer can configure the tool by specifying 1) prefix of identifiers which the team allow to use when naming identifiers in a program, 2) naming conventions (in Table I)
that the developer wants to check violations, and 3) abbreviations that are defined by the team or by Apple [2] and allowed in a program.

C. Extract Magic Numbers and Literal Strings

Using the Objective C convention checker, the developer selects the program and the tool uses the Word Segment API of Python framework to segment the code into words. Then the tool uses the class NSRegularExpression of the Foundation framework to match the code with the regular expressions for magic number (except -1, 0, 1, 2) in Table II and for literal string in Table III. Magic numbers and literal strings that are found in the program are stored in an SQLite database.

D. Check Naming Conventions

Using the Objective C convention checker, the developer uploads the naming list file (from Section A) and, for each identifier name, the tool checks the general conventions first. Then it checks other naming conventions, depending on the type of the identifier name (i.e. class, method, function, variable, property, or constant names).

<table>
<thead>
<tr>
<th>Table I: Objective C Naming Guideline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
</tr>
<tr>
<td>--------------</td>
</tr>
<tr>
<td>General</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Class name</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Method/Function</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>- Delegate Method</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>- Private Method</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Function</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Property/Variable</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Instance variable</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Magic Number</td>
</tr>
<tr>
<td>Literal String</td>
</tr>
</tbody>
</table>
In naming convention checking, the Word Segment API is used to segment each identifier name into an array of words. Checking of naming conventions in Table I can be done based on the following four categories of checking.

1) Check number of words
This is to check if an identifier name contains a number of words according to the guideline. For example, an algorithm to check if an identifier name contains multiple words is shown below.

**Convention:** Name should consist of multiple words.

**Input:** Name (array of words)

**Output:** Result (yes or no)

1: Declare integer variable count
2: Read name
3: Set count to number of words
4: If count > 1
5:  Print “name is valid” and return yes
6: Else
7:  Print “name is violating” and return no

2) Check string or character
This is to check if an identifier name contains or omits a string or character according to the guideline. For example, an algorithm to check if an instance variable name starts with an underscore is shown below.

**Convention:** Instance variable name should have underscore as a prefix.

**Input:** Instance Variable name (array of words)

**Output:** Result (yes or no)

1: Declare a string variable char
2: Read instance variable name
3: Set count to number of words
4: If count = 1
5:  Print “name is valid” and return yes
6: Else
7:  Print “name is violating” and return no
3: Set char to first character of name
4: If char is equal to underscore
5: Print “name is valid” and return yes
6: Else
7: Print “name is violating” and return no

3) Check string pattern using regular expression

This is to check if an identifier name contains a string pattern according to the guideline. For example, an algorithm to check if a class name starts with a prefix that is configured to use by the team is shown below.  

Convention: Class name should have prefix.

Input: Class name (array of words) and prefix
Output: Result (yes or no)
1: Declare a string variable name
2: Declare a string variable regex pattern
3: Read class name
4: Read prefix
5: Set name to the class name
6: Set regex pattern with prefix to 
   "(?<={prefix})[A-Z][a-z].*"
7: If name matches regex pattern
8: Print “name is valid” and return yes
9: Else
10: Print “name is violating” and return no

4) Check meaning and type of word

This is to check if an identifier name consists of meaningful words and the words are of the types according to the guideline. For example, to check if a method name starts with a verb followed by a noun, the tool uses the REST Words API to obtain information about the first and second word of the method name by appending the URL https://wordsapiv1.p.mashape.com/words/ with the requested word. If the requested word has meaning, the Words API returns information, including the part of speech. In some cases, the tool has to first obtain the present simple form of the requested word from the REST WebKnox Word API by requesting the URL https://webknox-words.p.mashape.com/words/{word}/simplePresent and specifying the requested {word}. After that, the part of speech of the present simple form is obtained from the Words API. The algorithm is shown below.

Convention: Method name should start with a verb followed by a noun

Input: Method name (array of words)
Output: Result (yes or no)
1: Declare a string variable word
2: Read method name
3: Set firstWord to the first word of the name
4: Set secondWord to the second word of the name
5: Add firstWord and secondWord to wordArray
6: For each word in wordArray
   7: Invoke WordsAPI URL passing word as argument
   8: Get part of speech of word from response from WordsAPI
   9: If word == firstWord
   10: If part of speech of word is not equal to noun
   11: print “name is violating” and return no
   12: If word is the last word
   13: Else
   14: If part of speech of word is equal to noun
   15: If word is the last word
   16: Print “name is valid” and return yes
   17: Else
   18: Print “name is violating” and return no

   All naming violations that are found are stored in the SQLite database.

E. Generate Shell Script

Using the Objective C convention checker, the developer generates a shell script from the naming, magic number, and literal string violation list in the SQLite database. The shell script is shown below.

KEYWORDS="@["{violating identifier name}"
find "$SRCROOT" \(-name "*.h" -or -name "*.m" \) -print0 | xargs -0 egrep --with-filename --line-number --only-matching "($KEYWORDS).*" | perl -p -e "s/($KEYWORDS3)/ warning: {description of violation}/"

F. Identify Naming Convention Violations

The developer has to open the program and add the shell script from Section E in the Run Script menu of Xcode. When the shell script is executed on the program, the locations of the identifier names that violate the naming convention guideline are identified.

G. User Interface of Objective C Convention Checker

Some screen shots of the Objective C convention checker are shown in this section. Fig. 2 is the main input screen with a menu for the developer to follow the framework. Fig. 3 shows an example of two shell scripts that are added to Run Script for the violations regarding the properties window and dUsage. Fig. 4 shows highlights on the violations and warning messages in the program.

V. EVALUATION

To evaluate if the naming convention checking framework could enhance program understandability, the iOS development team conducted an experiment by asking four developers to study Objective C programs as detailed in Table IV.
For the significance level of 0.05, \( t \) calculate was 7.595. At the significance level of 0.05, \( t \) calculate was greater than \( t \) critical, \( H_0 \) was rejected and \( H_1 \) was accepted. The team concluded that the framework was effective and revision of identifier names in the programs as suggested by the framework could save maintenance time and improve program understandability.

Each pair of developers inspected two programs on Xcode and each program had two versions, i.e. the original version before using the framework and the revised version after revision as suggested by the framework. They studied the two versions in different order (i.e. Original before Revised or Revised before Original) to reduce bias in understanding. After studying each version, they ran that version of the program once to see how it worked. Then, they answered a test of 30 questions, such as those in Fig. 5, which assessed their understanding in that version of the program. The paired t-test was used to test the following hypotheses:

\[
\begin{align*}
H_0 & : \mu_1 - \mu_2 = 0 \\
H_1 & : \mu_1 - \mu_2 > 0
\end{align*}
\]

where \( \mu_1 \) is the average of time spent in answering all questions for the original program correctly, and \( \mu_2 \) is the average of time spent in answering all questions for the revised program correctly.

Given the experimental result in Table V, \( t \) calculate was 2.353. At the significance level of 0.05, \( t \) calculate > 1.963, \( H_0 \) was rejected and \( H_1 \) was accepted. The team concluded that the framework could save maintenance time and improve program understandability.