TECHNICAL DEBT IDENTIFICATION USING TEXT CLASSIFICATION

PROJECT REPORT

Submitted by

Kaliappan

David Sundaraj

Prashanth Lidwin Jessuva

submitted to the Faculty of

INFORMATION SCIENCE AND TECHNOLOGY

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

MONTH YEAR

ANNA UNIVERSITY

CHENNAI - 600 025

BONA FIDE CERTIFICATE

Certified that this project report titled Technical Debt Identification Using Text Classification is the bona fide work of Kaliappan, David, Prashanth who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE: CHENNAI   **\<NAME OF GUIDE\>**

DATE: 31/11/2018   **\<DESIGNATION\>**

                                              PROJECT GUIDE

                                              DEPARTMENT OF IST, CEG

                                              ANNA UNIVERSITY

                                              CHENNAI 600025

COUNTERSIGNED

Dr. SASWATI MUKHERJEE

HEAD OF THE DEPARTMENT

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600025

ABSTRACT

The issue of Technical debt has been a subject of controversy for sometime now, Technical debt as such is not harmful for a software project but it is essential to identify such technical debts in the project as soon as possible because the later you find the more the debt increases and at last it becomes unbearable to pay back similar to the financial debts in our real life.

There have been many methods suggested to identify technical debts in the project like Code Smells, Source Code Analysis, Comment analysis etc., In this project we have come with a new and novel way to identify these technical debts and it is by going through the bug reports of that particular software. On applying text processing to the bug reports we have classified a bug report as either a technical report or a non-technical report. Thus the developers and testers can concentrate in detail on the bugs that were marked as technical.

For this project we have concentrated on the Chromium Project, it is a open source project from Google. The bug can be filed by anyone using a software called Monorail. We have scraped 700 of these issues for now as our dataset and given this to software engineering experts who classified these bugs as either technical or non-technical. This is the dataset using which we have experimented with various Text Classification methods. In essential we are using NLP methods to analyse and understand the issues and then classify it as either Technical or Non-Technical.

Keywords: Technical Debt, Chromium Project, Code Smells, NLP.

ACKNOWLEDGEMENT

Acknowledgement should be brief and should not exceed one page when typed in one and a half line spacing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

$-, \neg, \sim$	Negation operator
$+, \vee, \cup$	Disjunction operator
$X, \wedge$	Conjunction operator
$\rightarrow$	Conditional operator
$\leftrightarrow$	Biconditional operator
$\diamond$	Future tense modal operator
α	Action

CHAPTER 1

INTRODUCTION

This chapter is about the introduction of your work.

1.1 FORMAT OF THE THESIS

A few guidelines for writing the thesis are given below. These are taken from Ph.d regulations of Anna University [1].

1.1.1 Table of Contents

The Table of contents should list all captions following it as well as any caption which precedes it. The title page, Certificate and Acknowledgment will not find a place among the items listed in the Table of Contents but the page numbers of which are in lower case Roman letters. One and a half line spacing should be adopted.

1.1.2 List of Tables and Figures

The list should use exactly the same captions as they appear above the Tables and Figures in the text. One and a half line spacing should be adopted.

1.1.3 List of Symbols and Abbreviations

One and a half line spacing should be adopted for typing the matter under this head. Standard symbols, abbreviations, etc. should be used.

1.1.4 Chapters

The chapters may be broadly divided into 3 parts (i) Introductory chapter, (ii) Chapters developing the main theme of the Thesis and (iii) Results, Discussion and Conclusion. The main text shall be divided into several chapters and each chapter may be further divided into several divisions and sub-divisions.

- Each chapter should be given an appropriate title.

- Tables and Figures in a chapter should be placed in the immediate vicinity of the reference where they are cited.

- Footnotes should be used sparingly. They should be typed single space and placed directly underneath in the very same page which refers to the material they annotate.

1.1.5 Appendices

Appendices are provided to give supplementary information, which if included in the main text may serve as a distraction and cloud the central theme under discussion.

- Appendices should be named alphabetically starting from Appendix A, Appendix B, etc.

- Appendices, Tables and references appearing in appendices should be numbered and referred to at appropriate places just as in the case of chapters.

- Appendices shall carry the title of the work reported and the same title shall be included in the Table of Contents page.

1.1.6 References

Any works of other researchers, if used either directly or indirectly, the origin of the material thus referred to at appropriate places in the thesis should be indicated. Use of bibtex is recommended.

1.1.7 Citation

A few examples for citing articles, books and websites are given in Chapter 2.

1.1.8 Tables and Figures

A Table or Figure including caption should be accommodated within the prescribed margin limits and appear on the page following the page where their first reference is made.

- Tables and Figures on half page or less in length may appear on the same page along with the text. However, they should be separated from the text both above and below by triple spacing.

- Tables and Figures on half page or less in length may appear on the same page along with the text. However, they should be separated from the text both above and below by triple spacing.

- All Tables and Figures should be prepared on the same paper or material used for the preparation of the rest of the Thesis.

- Two or more small Tables or Figures may be grouped if necessary in a single page.

1.2 TYPING INSTRUCTIONS

Some guidelines for typing are provided below.

1.2.1 General

One and a half line spacing should be used for typing the general text. The general text shall be typed in Font Style Times New Roman and Font Size 13.

1.2.2 Chapter

The word CHAPTER without punctuation should be centered 50 mm down from the top of the page. Two spaces below, the title of the chapter should be typed centrally in capital letters. The text should commence 4 spaces below this title, the first letter of the text starting 20 mm inside from the left hand margin. The division and sub-division captions along with their numberings should be left justified. The typed material directly below division or sub-division heading should commence 2 spaces below it and should be offset 20 mm from the left hand margin. Within a division or sub-division paragraphs are permitted.

Even paragraph should commence 3 spaces below the last line of the preceding paragraph, the first letter in the paragraph being offset from the left hand margin by 20 mm.

1.3   NUMBERING INSTRUCTIONS

1.3.1   Page Numbering

All page numbers (whether it be in Roman or Arabic numbers) should be typed without punctuation on the upper right hand corner 20 mm from the top with the last digit in line with the right hand margin. The preliminary pages of the Thesis (such as Title page, Acknowledgement, Table of Contents, etc.) should be numbered in lower case Roman numerals. The title page will be numbered as (i) but this should not be typed. The page immediately following the title page shall be numbered as (ii) and it should appear at the top right hand corner as already specified. Pages of main text, starting with Chapter 1 should be consecutively numbered using Arabic numerals.

1.3.2   Numbering of Chapters, Divisions and Subdivisions

The numbering of chapters, divisions and sub-divisions should be done using Arabic numerals only and further decimal notation should be used for numbering the divisions and sub-divisions within a chapter. For example sub-division 4 under division 3 belonging to chapter 2 should be numbered as 2.3.4. The caption for the sub-division should immediately follow the number assigned to it. Refer below to the subsection and subsubsections. Every chapter beginning with the first chapter should be serially numbered using Arabic numerals.

A subdivision of the subsection

Use `subsubsubsection` in Latex to indicate a subdivision. The numbering should automatically appear in the Table of Contents.

1.3.3   Numbering of Tables and Figures

Tables and Figures appearing anywhere in the Thesis should bear appropriate numbers. The rule for assigning such numbers is illustrated by an example. Thus, if a Figure in Chapter 3, happens to be the fourth then assign 3.4 to that Figure. Identical rules apply for Tables except that the word Figure is replaced by the word Table. If Figures (or Tables) appear in appendices then Figure 3 in Appendix Section 2 will be designated as Figure A 2.3. If a table to be continued into the next page this may be done, but no line should be drawn underneath an unfinished Table. The top line of the Table continued into the next page should, for example read Table 2.1 (continued) placed centrally and underlined.

1.3.4   Numbering of Equations

Equations appearing in each Chapter or Appendix should be numbered serially, the numbering should commence afresh for each Chapter or Appendix.

CHAPTER 2

LITERATURE SURVEY/RELATED WORK

In order to understand about technical debt and the current methods of identifying this in a project there was an extensive Literature Review done, as part of the Literature Review many IEEE and ACM papers were read so that there is a complete understanding about the existing methods and the places where the method can be improved.

2.1   THE WYCASH PORTFOLIO MANAGEMENT SYSTEM.

In the development of a software named WyCash+ for Wyatt Technology, there was a problem of a new feature fitting poorly in existing architecture, by using a makeshift method the new feature was accommodated and released. Later this makeshift method was replaced with the fully functional feature. Cunningham had first used the analogy of technical debt in this research paper.

2.1.1   Citation of Journals

Method for citing journals can be seen in References [2], [3] and [4].

2.1.2   Citation of Proceedings

Proceedings should be cited as given in the References [5] and [6].

2.1.3   Citation of Websites

The websites should be cited as shown in this Reference [1].

2.1.4 Citation of Books

Books should be cited as shown in References [7] and [8]. The edited book that consists of several articles should be cited as shown in the Reference [9].

2.1.5 Citation of Thesis

Reports and thesis should be referred as given in the Reference [10].

2.1.6 Citation of Patents/ Newspapers/ Media

Check the Reference [11] for referring the patent information and [12] for citing newspapers or other media related information.

2.2 SECTION 2

2.2.1 Subsection 1

2.2.2 Subsection 2

2.3 SECTION 3

2.3.1   Subsection 1

2.3.2   Subsection 2

CHAPTER 3

DESIGN OF YOUR WORK

Your actual work should come in here. Include system architecture, modules, etc., with diagrams and provide an explanation for each of these.

3.1   VIEW OF TABLES

Sample view is shown in Table 3.1 and an example is given in Table 3.2.

Table 3.1: Example 1

$c1$	$c2$
r1	r2
r1	r2

Table 3.2: Example 2

x	y
x	*not y*

3.2   VIEW OF FIGURES

Examples of pictures are shown in Figure 3.1 and Figure 3.2.
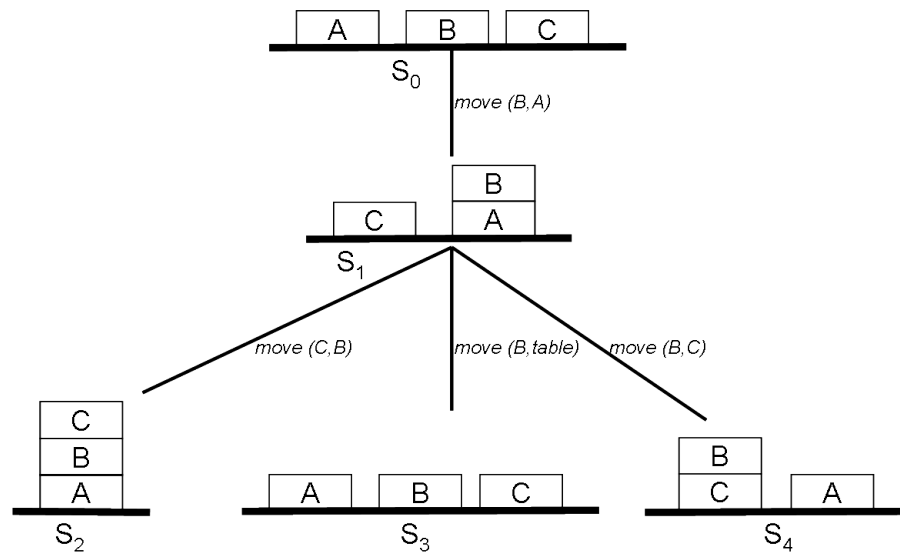
3.3   VIEW OF EQUATIONS
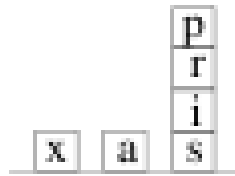
Figure 3.1: Example 1

Figure 3.2: Example 2

The sum of squares of *a* and *b* are calculated as shown below:

$$(a+b)^2 = a^2 + b^2 + 2ab \qquad (3.1)$$

From Equation 3.1, the data is obtained.

CHAPTER 4

IMPLEMENTATION OF YOUR WORK

The implementation details of your work should be mentioned in this chapter.

4.1   ALGORITHM 1

1: Get the number of variables *num*
2: Start with an empty list x [ ]
3: **for** each *n* of *num* **do**
4:      Get the x
5:      Get the t
6:      Get the w
7: **end for**
8: **if** $xl == s$ **then**
9:      terminate with the failure message **failed**
10: **else**
11:     $z = x$
12: **end if**

CHAPTER 5

RESULTS AND PERFORMANCE ANALYSIS

This chapter should provide the details of results and the analysis of your work. Depending on the type of project, there may not be analysis. In such cases, mention the title as "Results" or "Testing and Results".

CHAPTER 6

CONCLUSION AND FUTURE WORK

Conclude your work and provide an explanation for future enhancements.

APPENDIX A

TOPIC 1

A.1 SECTION 1

A.2 SECTION 2

APPENDIX B

TOPIC 2

B.1   **SECTION 1**

B.2   **SECTION 2**

REFERENCES

[1] Anna University PhD-Regulations 2015. `https://cfr.annauniv.edu/research/regulation/PhD-Regualtion-2015.pdf`. Accessed: 20 March 2015.

[2] K Alishahi, F Marvasti, V A Aref, and P Pad. Bounds on the sum capacity of synchronous binary cdma channels. *Journal of Chemical Education*, 55:3577–3593, 2009.

[3] T G Conley and D W Galeson. Nativity and wealth in mid-nineteenth century. *Journal of Economic History*, 58:468–493, 1998.

[4] S Waldron. Generalized welch bound equality sequences are tight frames. *IEEE Transactions on Information Theory*, 49:2017–2309, 2008.

[5] Richard E Fikes and Nils J Nilsson. STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. In *Proceedings of the 2Nd International Joint Conference on Artificial Intelligence*, IJCAI'71, pages 608–620, 1971.

[6] Weiguo Fan, Michael D Gordon, and Praveen Pathak. Personalization of Search Engine Services for Effective Retrieval and Knowledge Management. In *Proceedings of the Twenty First International Conference on Information Systems*, ICIS '00, pages 20–34, 2000.

[7] D H Holt. *Management Principles and Practices*. Prentice-Hall, Sydney, 1997.

[8] Philippe Aghion and Steven Durlauf, editors. *Handbook of Economic Growth*, volume 1. Elsevier, 1 edition, 2005.

[9] Dan Riley. *Industrial relations in Australian education / edited by Dan Riley*. Social Science Press [Wentworth Falls, N.S.W.], 1992.

[10] J P Hos. *Mechanochemically synthesized nanomaterials for intermediate temperature solid oxide fuel cell membranes*. PhD thesis, University of Western Australia, 2005.

[11] A H Cookson. Particle trap for compressed gas insulated transmission systems, 1985. US Patent 4554399.

[12] J Ionesco. Federal Election: New Chip in Politics. *The Advertiser*, page 10, 2010.