Final Year Project

Time Series Based Summarization

Mak Yen Wei

Bachelor of Computer Science

(Data Science)

Jan 2023

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

90 percent of engineering is searching on google! Bugs solving being one of the crucial task for a developer. While it's very important, solving bugs is tedious when you have to look through forums after forums, stackoverflow after stackoverflow just to solve a simple linear equation problem in python. It is very common to see developers spending hours just to find the exact solution on stackoverflow, This can happen in possible two situation.

Solving bugs in the software world is not straight forward at all, there are way too many co-dependent libraries and packages that are used in the software world across langauges, frameworks, and platforms. At times, it's a fresh issue that occured based on the latest release of a particular software. On the other hand, it could be a typo in the code that is causing the problem. Then, there are times where the problem is caused by the local environment of the developer. There is a non-ending list of possibilities that can cause the problem.

Therefor the aim of this research is to solve these two problems using text processing methods, rankings methods and lastly summarization models to ease the process of finding the exact solution to the problem for developers.

1.1 Problem Statement

Throughout this research work, we aim to solve 2 of the major issues outlined. The first being that developers are not able to find the exact solution to the problem. The second being that developers are not using the right keywords to search for the solution.

First being that the developer is not using the right keywords to search for the solution. This can lead to miss-interpretation of the problem and the solution queried by the forum. Since there are millions of developers using stackoverflow, you can expect the amount of questions and answers to be very similar to each other when the keywords are vauge and not specific.

Second being that the developer is not able to find the exact solution to the problem. The software world now is very much complicated, every language, packages or libraries are co-dependent to each other, this creates a problem where it's very time-consuming to find the root solution to the problem. Developers ended up going into an endless rabbit hole just to find dependencies and dependencies of dependencies to solve the problem. This can be very frustrating and time consuming.