**UNVERISITY OF WATERLOO**

Faculty of Mathematics

**Code Architecture and Code Optimization**

Blueprint Software Systems Inc

90 Eglington Avenue E

Ontario, ON, M4P 1A6

Prepared by

Jiaheng Zhao

3A Computer Science

ID: 20613526

November 11 2017

# Memorandum of Submittal

To: Alex Utkin, Senior Software Developer

From: Jiaheng Zhao, Software Developer

Date: Nov 11, 2017

Re: Work Report: Code Architecture and Code Optimization

As we agreed, I have prepared the enclosed report, “Code Architecture and Code Optimization” for my 3A work report and for the NGSTAR team. This report, the first of four work reports that the Co-operative Education Program requires that I successfully complete as part of my BCS Co-op degree requirements, has not received academic credit.

The NGSTAR team are one of the developer team that is responsible for maintain the performance of Legacy Product Silverlight and implement new features in brand new product Storyteller. My job is to design, implement, maintain and support the back-end and front-end software solutions of our team. I have processed and extracted data from Social Media both manually and automatically using scripts. This report aims to analyze the importance and impact of social media in product analytics.

The Faculty of Mathematics requests that you evaluate this report for command of topic and technical content/analysis. Following your assessment, the report, together with your evaluation by qualified work report markers. The combined marks determine whether the report will receive credit and whether it will be considered for an award.

Thank you for your assistance in preparing this report.

Jiaheng Zhao

# Table of Contents

[Memorandum of Submittal i](#_Toc491425004)

[Table of Contents ii](#_Toc491425005)

[Executive Summary iii](#_Toc491425006)

[1. Introduction 1](#_Toc491425007)

[2. Analysis 2](#_Toc491425008)

[2.1. What is Code Optimization? 2](#_Toc491425009)

[2.2. Importance of Processing Social Media Data 2](#_Toc491425010)

[2.2.1. Process Data Manually 3](#_Toc491425011)

[2.2.2. Process Data Automatically 4](#_Toc491425012)

[2.3. Limitation with Current Data Processing Methods 5](#_Toc491425013)

[2.3.1. Data Quality and Data Quantities 5](#_Toc491425014)

[2.3.2. Efficiency and Productivity 6](#_Toc491425015)

[2.4. Potential Improvements 7](#_Toc491425016)

[3. Conclusions 8](#_Toc491425017)

[References 9](#_Toc491425018)

# Executive Summary

The purpose of this report is to analyze how code architecture and code optimization can improve software performance , code readability and code maintainability. Two different approaches of handling social media data are discussed based on their effectiveness and limitations. Feasible improvements on quantitative and qualitative analysis of social media are provided.

In the report’s analysis, it discusses the importance of processing social media data manually in the early stage of a product release. It points out the necessity and drawbacks of automating issue tracking as well as accounting for the late stage of product release. The limitation and inaccuracy of current data processing methodologies are investigated, and corresponding improvements are advised.

The conclusion of this report emphasizes the essential elements of extracting useful information throughout the product release, using different procedures in different stages. In addition, it highlights the impacts of social media in product quality analytics.

# Introduction

Blueprint sys Inc is a global leader in requirements definition solutions for software business analysts. The products – Blueprint Storyteller for Agile, Blueprint Automate for DevOps and Blueprint RegTech for Compliance resolve many of the time-consuming, costly, and error-prone challenges that IT organizations face when driving digital transformation for their businesses. The develop team is implementing new features that help customer analysis impact, historical version control and review experience. The project is built on thousands of files that is written in C# in back-end and angular in front-end. To modify code, debug defects or implement new features inside the project require a well-designed code architecture, consistent coding standards as well as clear, concisely codes.

This report will discuss the great influences with how the well-format codes can help developers navigate, refactor and delete unnecessary blocks. In addition, how an elegant architecture can assist developers implement new features.

# Analysis

## What is Social Media?

Social media is defined as websites and applications that enable users to create and share content or to participate in social networking. In this work report, the scope of social media is limited to three websites Crackberry, Reddit and Twitter.

**Crackberry:** The most popular BlackBerry site in the net, featured with forums discuss BlackBerry devices, apps and accessories, and moderated by BlackBerry authorities.

**Reddit:** A social news aggregation, web content rating, discussion website. In this report, reddit is restrict to BlackBerry subreddit, meaning a reddit site focus on BlackBerry.

**Twitter:** A online news and social networking service where users post and interact with messages.

## Importance of Processing Social Media Data

Before talking about processing data from social media, it is important to understand the responsibility of product quality analytics team. Product analytics managers combine quantitative and qualitative analysis to inform and influence important product decisions as well as building and maintaining reports and metrics to monitor the performance of product and understand customer experience. Therefore, reliable and timely data is essential for successful product analytics. Social media provides data inputted by end users, so the relevance of data is guaranteed. Since users are very attached to their phones, there is a high likelihood they will post to social media) Thus, the proportion of users covered is adequate. Therefore, social media fulfills all the requirements of being a source of data for product analytics purpose.

### Process Data Manually

The most straight forward way to interpret social media data is to read them. Although it is the most primitive way, it captures the greatest amount of information. In the early stage of product launch, any potential issue of the device remains unknown. There are always some unforeseen defects by the designer and the manufacture. Therefore, an actual human is required to go through the forums.

During the early launch of a product, a forum user reported a major concern regarding the product quality of their device. The thread quickly attracted massive attention from many other users on that site. People started to wonder is this incident just a coincidence or a significant quality issue. Product quality analytics team spotted this issue the day after it was posted, and compiled it into a report in the same week. BlackBerry was able to take necessary measures as early as possible, which create more time to fully investigate the source of the problem.

Manual reading of social media plays a significant role in product analytics in terms of identifying issue and gathering feedback. A person can identify can categorized an issue based three questions. What is the issue, where does it happen and how does it affect product’s operation? It is very hard for computers to imitate such procedures to identify an issue. The more issue identified manually, the easier to process similar issue later automatically. In addition to identify issue, gathering user feedback helps the developers to find missing features, improve future design and understand users’ requirements.

### Process Data Automatically

The possibility of finding new problems of a product tends to decrease as its launch time increase. Common and severe issues have likely been identified in the manual data processing stage, because they were mentioned multiple times by different users. With the majority of issues categorized, the next step is to measure its quantity. It will be a tedious and error-prone task for human, but a suitable task for computers.

If a post is complaining about an issue, certain words are more likely to appear than others. For example, if some users are complaining about overheating of a phone, they tend to say the words like ‘overheating’, ‘warm’ and ‘hot’. However, if the search was performed with those keywords, there will be a lot of noise like ‘Today is very hot’. In order to reduce this type of noise, another set of keywords like ‘phone’, ‘device’ and ‘KEYone’ should be included in the search as well. There are different methods to achieve such search, a technique called regular expression is used due to its powerful ability to flexibly capture similar phrases.

Regular expression, or regex in short, is defined as a sequence of characters that define a search pattern. This search pattern is used by a string searching algorithm for finding specific pattern in text.

The regex is very helpful in two aspects, which are formatting phrases and applying negation. A phrase formed by certain predicates and adjectives can be recognized as an indication of an issue. Although the important parts are the predicates and adjectives, placeholders for other words are needed to construct a meaningful sentence. The number of placeholders is limited so that the predicate and adjectives are in the same sentence. Regex supports an easy way to create placeholders, and user are free to choose the length of placeholders. The placeholder can be restricted to a set of characters, which is very helpful in some circumstances.

Words like ‘not’, ‘never’ and ‘rarely’ can change the meaning of a sentence entirely. Regex provides the ability to filter phrases that are preceded or followed by certain words. Therefore, if only high sentiment phrases are wanted, regex accepts words and phrases like ‘good’, ‘excellent’ and ‘not bad’, but reject ‘bad’, ‘poor’ and ‘not good’. This is a significant advantage of regex over the regular keyword based searching. The sequence of words appearing in a sentence decides the meaning of the sentence.

## Limitation with Current Data Processing Methods

### Data Quality and Data Quantities

To measure the severity of an issue, one of the common criteria is the number of users affected. Data regarding to affected user percentage is collected entirely by scripts, which counts the number of matches of defined search patterns. The limitations of this approach can be described in two aspects.

The first one is the complexity of English language. The number of synonyms and antonyms of a word is large, so some words are missed due to the limitation of creator’s vocabulary. Moreover, a problem can be described in many ways in different sentence structures. Therefore, it is very hard to guarantee a search pattern can find all issue presented in the social media.

The nature of social media creates another source of inaccuracy in quantitative measure. People tend to comment and report issue that are already popular. Also, most social media users only read contents that are presented in the front page, so data tends to be polarized. Furthermore, fraudulent information is hard to avoid as proof is not required to claim an issue on the internet.

### Efficiency and Productivity

Even with good skim and scan skill when reading social media, the time required to go through all the content is very long. The amount of information that needs to be processed is too large for a human. Constant reading of large text tends to bore the readers and slow their reading speed. Important issues may be missed when the readers decide to skip a large chunk of text. It is very hard to maintain a high level of reading for an extended period of time.

Building an accurate search pattern in regex requires a large amount of time. Every word need to be chose carefully, and tested against test sets. Frequently, a change in the search pattern will increase the accuracy for certain phrases, but decrease the accuracy for others. Patience and experimentation are important for an accurate search pattern, so the productivity of generating search pattern is restricted.

## Potential Improvements

With the help of machine learning, words that are most relevant can be grouped together. Instead of thinking about all the English word represent hot, a group of words generated by machine learning modules such as pythons Word2Vec can be applied in the regex. The time needed for creating regex for each category will be reduced, and the percentage of relevant information that are captured will increase.

# Conclusion

To sum up, social media is very important part of product quality analytics, despite that data originated from social media could be biased and false. The benefits of reading social media data manually in the early stage of product release outweighs its inefficiency. Automated data processing system is necessary as the data grows quickly to the scale where human reading will be an infeasible task. Powerful tools like regular expression should be used in creating search patterns. The accuracy of search pattern should be tested thoroughly based on the number of matches and the relevance of matches.

The inefficient and time consuming nature of manual data processing is unsatisfactory. Constant and high volume reading is prone to mistakes, as human will get tired and have trouble focusing. The automated approach is the correct path, but suffers from data inaccuracy and lacking productivity. Extra time and effort is needed to create a high quality search patterns, and this process is costly.

With the help of more advanced concepts and algorithms like machine learning and neural network, the speed of issue categories identification and tracking can be increased. Using complex system will require more tweaking and effort at the start, but returns a greater reward in the long run.

# References

CrackBerry.com (2007). Advertise with CrackBerry.com. Retrieved August 24, 2017, from <https://crackberry.com/advertise>.

Reddit (2005). Reddit wiki. Retrieved August 24, 2017, from <https://www.reddit.com/wiki/index>

Twitter (2006). Twitter About. Retrieved August 24, from <https://about.twitter.com/>

Regular-Expressions.info. What Regular Expressions Are Exactly – Terminology. Retrieved August 24, 2017, from http://www.regular-expressions.info/tutorial.html