

Project plan for degree project

Template version 1.6 – June 12, 2020

COURSE CODE (PA2534): MASTERARBETE I PROGRAMVARUTEKNIK

April 10, 2021

Title	Data-driven requirement engineering: Applying NLP techniques to react bug reports provided by the stakeholders as feedback.
-------	---

1 Introduction

Now a days most of the products and companies are increasingly adopting the culture of software-intensity which makes the opportunity of having products on the Internet and companies can easily learn and collect feedbacks about the products over the internet easily [1],[2]. In the development phase of a software product the goal of requirement engineering is to identify the demand of the stakeholders, documenting and classifying the demand of the requirements, negotiating with the requirements, and managing them and the make assumptions about the requirements. Requirement Engineering focuses mostly on involving capturing user's needs, making a release plan of the needs of the users and getting their feedback [6]. A large part of requirement engineering is concerned with the demand of the users, what their needs are. A survey showed that 90% of the release engineers believe that the customer feedbacks are most important while evaluating the system as succeed or failed [10].

Now a days, users are demanding more advance features which they may ask for by seeing the same features in another app or sometimes from their own demand. The success of a software product mostly depends on the acceptance of the users and their feedback plays an important role while the product will mark as succeed or failed [6]. That is why taking the feedbacks of the users should be taken seriously and modifying the software based on their demand.

While providing the feedback the users provide their feedback in a general way. Like facing problems while using the software which might be like system failure, connectivity issues or the features were not functioning as it should be. In these cases, the users send a report or provide a report about the problem they are using. Besides sometimes the reports or feedback provided by the users contains only the user experience like how they are feeling while using the application which are not beneficial to the developer team to update the system. Some reports might be incomplete. Overall, the feedbacks and reports contain mixture of different types of data.

But in maximum cases all the reports and feedbacks are stored in the same place which are not good enough to take steps based on those. Despite of a large number of significant data is being collected through user feedbacks, companies and software developers do not have enough knowledge about how their products are being used by the users, which features of the system should be modified for a better experience of the users, which features should be added which will be appreciated by the users to use in the future[3].

Failure of dealing with the feedback data provided by the users can cause failure of the software product which also may be cause of deduction of users. But all the feedback provided by the users are not usable for the software development. That is why the data should be filtered as

well as categorized and should be stored based on the categories. If the data can be stored based on the usage categories for development will be causing less time consuming and less costing. In this way the developers can easily take steps like fixing bugs in the application, analyzing and adding new features which is requested by the users and at the same time can ignore which are ignorable.

2 Aim and objectives

The aim of this thesis is to provide an analysis methodology for reacting and response back fast to the bug reports from a dataset.

To be detailed we will filter the bug reports from a large set of data which is collected as the user feedback. Besides filtering the bug reports, we will rank the bug reports for release plan based on user demand so that the requirement engineers can take necessary steps without wasting so much time in filtering the bug reports from the dataset and making a release plan for the software development.

To reach the aim of this thesis collecting a dataset is most important. A dataset storing with all types of feedbacks from the users will be filtered with a data analysis tool. For filtering the bug reports from the data set with the tool we will consider natural language key words which are used by the users while providing feedback or requesting for fixing a bug in the software. After filtering them we will rank them by searching most common key words in the system. Then by seeing the ranking the requirement engineers can easily make a release plan for the software.

The steps to reach our goal will be:

- Collecting the dataset.
- Choosing a data analysis tool for filtering.
- Feeding the dataset in the analysis tool.
- Filtering the dataset by providing natural language keywords (all possible keywords) which are used for reporting bugs in the software.
- Apart the filtered data from the main the dataset by making another sub-dataset.
- Rank the dataset based on the usage most common keywords.
- Remove the unnecessary data from sub-dataset (if required).
- Make a release plan with the final sub-dataset.

3 Research questions

For reaching out the goal of this thesis the research questions are:

1. What information contained in user feedback are useful for bug reports?
2. which tool and techniques are reported to be most effective in overcoming the problems in bug report classification?
3. which tool and techniques are reported to be most effective in overcoming the problems in bug report classification?
4. What algorithms will be most effective at filtering the user feedbacks and separating unrelated user feedbacks from the bug reports?
5. How overcoming the bug report classification filtration problem would help in better release planning?

4 Research method

Typically, in data-driven requirement engineering feedback is collected through customer interviews, questionnaires, and customer surveys [5]. In this context, Chen et al. recommend collecting the customer data as well as product data. The customer might be like demographic, psychographic, behavioral data etc. and the product data might be like operation, performance, responsiveness etc. [1].

To reach the goal of this thesis it is required to know which tools are available out there as well as we need to find out the data source. To compete with this, we need to do “Literature review” to find out the resources of the datasets which can be reliable in terms of this thesis. To find out the resources in this thesis search will be conducted in different databases like IEEE Xplore, Google scholar, Scopus, ACM digital library and many more. So, this literature review will be performed to collect the data for our experiment.

After collecting the data from various resources by doing a systematic “literature review”, an “Experiment” can be one of the best ways to be chosen as the method for reaching our goal for this thesis. Since our main goal of this thesis is to organize the data and filtering the bug reports from the dataset. By doing experiment it is possible to different types of experiments with the data like filtering the data, removing the unnecessary data, categorizing the data etc. We will use a data analysis tool (like R) to analyze the dataset. A data analysis tool will helpful to analyze the dataset and sorting out the bug reports from the dataset. That is why “Experiment” is the best match for this thesis.

As of other research methodologies like “Survey” will not be enough to get the data which we need for this project. Because by performing “Survey” we might not be able to gather sufficient data, and besides we might not be able to manage the participants as we required. Besides “Ethnographies” is used for investigating or to learn about culture of a particular setting or environment which is not a perfect match for this thesis. “Case study” is basically used for studying a given case for a particular situation and it is not as wider as we required for this thesis.

After the above discussion, it is quite clear that a “Literature review” for collecting the dataset and to reach our main objective “Experiment” is the best match.

5 Expected outcomes

Users are providing their feedbacks on the online platforms rapidly. Every day the companies are getting a huge amount of data from these feedbacks. These feedbacks consist of various types of data like their expectation, their experience, what they will more likely to use in the future releases, bug reports etc. The main goal of our project is to rectify the bug reports from these feedbacks and response back to the bug as early as possible.

After this thesis, the developers will be able to filter the data from the dataset to come up with the bug reports and can easily response back to the reports.

This thesis will provide methodologies of filtering the bug reports from a dataset where various types of data will be stored which were provided by the users as feedback. Filtering those bug reports and making those reports usable for the requirement engineers is the main expected outcome of this thesis.

After this research the practitioners will be able to organize the data. They will be able to fix the bugs without wasting much time on eliminating waste from the data as well as they can identify new feature requests from the customers. It will also be possible to remove the garbage data from the dataset which are not required.

6 Time and activity plan

This is a time-efficient analysis methodology which will help the requirement engineers to react and response fast to the bug reports which are provided as feedback by the users in the software while using the software. Its working is based on the following steps which are summarized in the below table which also including the timeline of each major activity:

SL No	Research phase	Objectives of the phase	Due Date	Duration
1	Selecting topic and making a draft on the topic.	<ul style="list-style-type: none">• Selecting a topic with having a research gap.• Writing a draft on the topic including research questions, research methods and expected.• Consulting with the supervisor on the draft.• Re-shaping the topic title and others based on the discussion.	31st January, 2021	1 Week
2	Background re-research, literature review and thesis proposal writing	<ul style="list-style-type: none">• Search for the available extensive relevant literatures.• Finding out the research gap.• Modify research questions based on the discussion.• Revise research methodology based on the discussion (if needed).• Add expected outcome, risk management and others.	6th February, 2021	1 Week
3	Thesis proposal submission	<ul style="list-style-type: none">• Submit the completed proposal.	7th February, 2021	
4	Research design planning	<ul style="list-style-type: none">• Design the research template.• Choose the tool to be used for analyzing the dataset.	14th February, 2021	1 Week
5	Dataset collection and selecting dataset	<ul style="list-style-type: none">• Collecting datasets from different resources.• Studying the datasets.• Selecting a dataset from the collected datasets which fits with the aim most.	21st February, 2021	1 Week

6.1 Supervision plan

While doing the supervision plan with the supervisor, it is quite difficult and unlikely to the previous years because of the Covid-19 situation. Doing face-to-face meetings has become tough for the regulations. For this reason, until the situation becomes okay, we need to do the meetings virtually. Sometimes if possible, meeting with the supervisor can be done but obviously by taking an appointment.

In the time and activity plan table I have stated that before every submission I will provide a

SL No	Research phase	Objectives of the phase	Due Date	Duration
6	Filtering the dataset for coming up with the bug reports	<ul style="list-style-type: none"> • Finding out the keywords usually used for reporting bugs. • Feeding the dataset in the tool. • Filtering the dataset with the keywords. • Save the sub-dataset found after filtering. 	7th March, 2021	2 Weeks
7	Ranking the bug reports	<ul style="list-style-type: none"> • Feed the sub-dataset in the data analysis tool. • Rank the reports from the most common to fewer common keywords found after filtering. 	21st March, 2021	2 Weeks
8	Redoing step 6 and 7 for cross checking	<ul style="list-style-type: none"> • Repeating step 6 and step 7. • Cross check with the two results. • Repeat step 6 and 7 again (If two results have significant difference). 	28st March, 2021	1 Week
9	Analysis gained data and removing unnecessary data	<ul style="list-style-type: none"> • Analysis the final data manually. • Remove the unnecessary data (If needed). 	4th April, 2021	1 Week
10	Storing the result	<ul style="list-style-type: none"> • Store the result for further usage. 	4th April, 2021	
11	Writing thesis draft	<ul style="list-style-type: none"> • Thesis draft using the results. • Check with the research question if the result is satisfying the research questions. • Sent the thesis draft to the supervisor. • Discuss on the problem and issues found on the draft. 	25th April, 2021	3 Weeks
12	Completing and submitting thesis draft	<ul style="list-style-type: none"> • Modifying the draft report based on the discussion with the supervisor. • Submit the thesis draft. 	9th May, 2021	2 Weeks
13	Submission of opposition report	<ul style="list-style-type: none"> • Analysis opposition thesis project. • Find out the work gaps, limitations. • Create a report. • Submit a opposition report. 	23rd May, 2021	2 Weeks
14	Thesis presentation	<ul style="list-style-type: none"> • Creating a thesis presentation. • Submitting the presentation. 	28th May, 2021	5 Days
15	Submission of Final Thesis	<ul style="list-style-type: none"> • Recheck the final thesis paper. • Submitting the final thesis. 	28th May, 2021	5 Days

Table 1: Time and activity plan

completed draft to the supervisor at least one week earlier. Within this one week the supervisor will be able to go through the drafts and arrange a feedback meeting to provide feedback on the reports.

Besides the reporting while analyzing the data I will provide a progress report every week so that the supervisor could suggest me if I went through wrong way. Also, sometimes it may happen that an issue came up during working which may cause interruption on the ongoing thesis. In this case I can immediately send an email to the supervisor and ask for help.

7 Limitations and risk management

In this thesis one dataset will be used for analysis which is a limitation. If we could use multiple dataset and process them than we could get a better output from the thesis. Also, analyzing the same dataset with multiple data analyzing tools would provide us a better output.

The tool will be selected and using for the data analysis might be failed to provide any output from the dataset. In this it will be hard to get rid of from the situation. In this case using another tool would be a solution.

Communication problem might be raising due to worldwide covid-19 situation. Communicating virtually with the supervisor might be a little bit tough as face-to-face communication makes the meetings more effective. To mitigate this risk the meetings should be done more often and student should be more responsible, more productive and more attentive while doing the feedback meetings with the supervisor.

It may also cause that the selected dataset is not having enough resource and for that we might not be able to reach our goal. In this we can use other datasets also we can perform an on-line survey to get a better output. But doing an online survey would be more time-consuming. Therefore, we should collect multiple datasets in the very beginning for backup plans.

References

- [1] Fabijan, A., Olsson, H. H., & Bosch, J. (2015, June). Customer feedback and data collection techniques in software R&D: a literature review. In *International Conference of Software Business* (pp. 139-153). Springer, Cham.
- [2] Maalej, W., Nayebi, M., & Ruhe, G. (2019, May). Data-driven requirements engineering- An update. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 289-290). IEEE.
- [3] Stanik, C., & Maalej, W. (2019, September). Requirements intelligence with openreq analytics. In *2019 IEEE 27th International Requirements Engineering Conference (RE)* (pp. 482-483). IEEE.
- [4] Fabijan, A., Dmitriev, P., Olsson, H. H., & Bosch, J. (2017, May). The evolution of continuous experimentation in software product development: from data to a data-driven organization at scale. In *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)* (pp. 770-780). IEEE.
- [5] Slegten, K. J. (2018). *Data-Driven Requirements Engineering* (Master's thesis).
- [6] Maalej, W., Nayebi, M., Johann, T., & Ruhe, G. (2015). Toward data-driven requirements engineering. *IEEE Software*, 33(1), 48-54.

- [7] Pagano, D., & Maalej, W. (2013, July). User feedback in the appstore: An empirical study. In 2013 21st IEEE international requirements engineering conference (RE) (pp. 125-134). IEEE.
- [8] Palomba, F., Linares-Vásquez, M., Bavota, G., Oliveto, R., Di Penta, M., Poshyvanyk, D., & De Lucia, A. (2015, September). User reviews matter! tracking crowdsourced reviews to support evolution of successful apps. In 2015 IEEE international conference on software maintenance and evolution (ICSME) (pp. 291-300). IEEE.
- [9] Bettenburg, N., Just, S., Schröter, A., Weiss, C., Premraj, R., & Zimmermann, T. (2008, November). What makes a good bug report?. In Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering (pp. 308-318).
- [10] Nayebi, M., Farahi, H., & Ruhe, G. (2017, November). Which version should be released to app store?. In 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 324-333). IEEE.