Report on Final PhD Examination

5 February 2014

Thesis title:

The Position of Quality Assurance Contributors in Free/Libre Open Source Software Communities

PhD Candidate:

Adina Barham

Examining committee:

Jonathan Lewis, Toshimichi Otsubo, Yoshiki Kurata, Sulayman Sowe

博士学位請求論文審査報告

2014 年 2 月 5 日

論文題目：

『The Position of Quality Assurance Contributors in Free/Libre Open Source Software Communities』

博士学位請求者：

バーハム・アディナ

論文審査委員：

ジョナサン・ルイス、大坪 俊通、倉田 良樹、スライマン・ソー

STRUCTURE OF THE THESIS

The structure of the thesis is as follows.

1 Introduction

1.1 Significance and Contributions of this Thesis

1.2 Research Questions

1.3 Outline of the Research Project

1.4 Thesis Structure

2 Literature Review

2.1 Free/Libre Open Source Software

2.2 Software Quality Assurance and FLOSS

2.3 Communities and FLOSS development

2.4 Summary

Bibliography

THE THESIS

This study investigates the relationship between software quality assurance (QA) and Free/Libre Open Source Software (FLOSS). More specifically, it investigates the impact of QA adoption on the structure of FLOSS communities. QA is defined for the purposes of this thesis as: Testing, contributing code to automated testing tools or any test related activity, triaging bugs or any activity performed on the projects issue tracker, and participating on the QA dedicated communication channels.

The thesis addresses two main research questions: Is QA a separate layer in FLOSS communities? and What are the communication patterns between QA members as well as with other project participants?

The research starts with a preliminary study of the top 100 FLOSS projects listed on Ohloh.net, which establishes that more than a quarter of these projects include some form of QA in their development. Following this, a preliminary case study is made of Mozilla, which is a large and mature community with a long history in developing successful FLOSS products such as Firefox and Thunderbird. QA mailing list data and issue tracker data are retrieved, cleaned and analysed using first simple statistics and then social network analysis techniques. A number of hypotheses regarding the position of QA activities within FLOSS communities are formed based on the findings.

Following this, four other case studies are carried out. The projects studied differ from Mozilla in size and history and are: Ubuntu, Plone, KDE, and LibreOffice. The hypotheses proposed based on the Mozilla case study are examined and adjusted in the light of the findings for those case studies.

The findings of the first phase suggest that in the case of Mozilla a smaller percentage of peripheral members (occasional contributors) are active on the mailing list as opposed to the issue tracker where a larger number of members have only one non-repeated act of communication. Activity on the QA mailing lists seems to be independent from activity on the issue tracker presenting peaks that are not directly related to time progression. Furthermore, a small group of people seems to be highly active in comparison to the rest of the community on both issue tracker and mailing lists. Social network analysis of the Mozilla community shows a large group of people spanning both issue tracker and mailing lists. However, almost two thirds of the

connections are created by single acts of communication from one member to another. Furthermore, the existence of a highly active small group of people is also supported by the social network analysis of the Mozilla community. As regards information flow within the Mozilla community, the risk of a small group of people brokering information seems to be very low. As far as the communication carried out only on the QA lists, the patterns seem to display similarities to the whole network communication but on a smaller scale, i.e. one group of people working together where a small subgroup is highly active. However, the analysis of the Mozilla community revealed that issue tracker data and QA mailing lists' data is insufficient to determine whether QA represents a separate layer in the community.

In the second phase of research, in addition to retrieving communication data conducted on issue trackers and QA mailing lists, all mailing lists associated with each project were retrieved as far as possible. Furthermore, a list of community members contributing code to the project was downloaded from Ohloh, a public directory of FLOSS and contributors. This list was used for both data cleaning and contributor layer identification. The analysis of these four case studies included general statistics methods and social network analysis techniques applied in a similar manner as for the Mozilla preliminary case study. The findings seem to validate some of the hypotheses proposed in the first phase of the research. For example, all four communities contained a large group containing most of the projects' participants that spanned mailing lists and issue trackers. Furthermore, all communities had a small number of participants with a higher than average activity that displayed strong connections among themselves and were connected to a large number of members. However, some case studies displayed particularities. For example, while in the Mozilla, Plone and KDE cases the QA contributors seemed to merge with other layers in the case of Ubuntu the QA group tended to be more separate. In other words, a smaller number of members contributing to QA activities in the latter two communities were contributing with other activities to the project while members contributing with QA activities in Mozilla, Plone and KDE seemed to bring a variety of contributions to the project by submitting code and participating to a large number of non-QA mailing lists. These findings may suggest that in some cases less technically knowledgeable individuals are finding a new way to contribute to FLOSS development aside from documentation and localisation. Further study can be conducted in this direction considering the targeted user base for these projects. Ubuntu and LibreOffice may display a more segregated and rigorous QA due to the fact that the intended end-users for these software products are not necessarily

technically "savvy".

With respect to the first research question, the five communities studied have dedicated communication channels, wikis, and other resources for providing QA related information. However, only the Ubuntu and LibreOffice communities displayed a somewhat separate QA layer where a large percentage of its members do not appear to be contributing code or communicating on other mailing lists. In the other communities a much smaller percentage of users were performing exclusively QA related activities; instead, they had multiple roles within the project. However, it is possible that non-QA activities were performed in different time frames than the ones in which the contributors were part of the QA team. Further study is required to clarify this point.

With respect to the second research question, all communities' graphs displayed a large group of people spanning both mailing lists and issue tracker. Previous research has suggested that FLOSS networks contain a small number of individuals with significantly higher connections than the network's average (called hubs). The case studies supported this argument, as they found a very small number of individuals with a high degree compared to the network's average. Similarly, in the QA teams studied a small number of vertices had a higher degree than average. Within these teams, participants did not direct all communication efforts to one or a few members who then conveyed that information to members of other groups. Instead, QA team members seemed to communicate not only among themselves but also directly with members of other groups. Furthermore, fewer than 1% of community members had a higher than average betweenness centrality value, and those values were not particularly high, which suggests that information flow in the networks was not particularly vulnerable to a small number of individuals leaving the community.

EVALUATION

This thesis is the first major academic study on the important topic of QA and FLOSS communities. It expands FLOSS research to cover the recent emergence of QA practices, while situating the new developments in the context of existing FLOSS concepts such as communities, contributions and freedom. Its findings will be useful not only for other FLOSS researchers but also for FLOSS communities seeking to improve their QA practices.

This thesis is also notable for its methodology. Many previous studies have made reference to Mockus et al's study of Mozilla Methodology, but this is the first to take Mockus et al's work and use it systematically as a basis for studying other FLOSS projects.

The thesis's ambitious generation of 17 hypotheses from a preliminary study of Mozilla, which it then tests on four more cases, offers a model on how to carry out quantitative research on FLOSS communities. It also offers a useful foundation for other FLOSS researchers to apply to other projects and activities. It might be objected that it was not necessary to investigate all the hypotheses for all four projects, as the results were easily predicted for some hypotheses and cases. However, the thesis followed Mockus et al in seeking to make its process as transparent as possible, and did not assume that a similar result in five case studies was enough to exclude the hypotheses from future analysis.

On the other hand, the thesis should have made it clearer why it excluded code quality from its analysis. Pressman and Sommerville defined types of QA activities and distinguished white box and black box activities; the former have to be performed by developers. This thesis, however, focuses on QA activities that can be performed by non-specialists and hence does not cover code quality.

The thesis could also have benefitted from a stronger general sociological perspective. In particular, it could profitably have addressed the question of whether and how the FLOSS onion model (with its periphery and center, and layers between) differs in practice from the pyramid hierarchical model used more widely in sociology.

SUMMARY OF EXAMINERS' CONCLUSIONS

On 18 December 2013 we examined Ms. Adina Barham regarding her PhD thesis "The Position of Quality Assurance Contributors in Free/Libre Open Source Software Communities." Ms. Barham satisfactorily answered all our questions regarding her thesis.

We therefore conclude that Ms. Barham has achieved the requisite level of academic achievement and ability to be awarded the degree of PhD in Social Sciences from this University.

23 December 2013
Examining committee:
Jonathan Lewis
Toshimichi Otsubo
Yoshiki Kurata
Sulayman Sowe

学位請求論文最終試験の結果の要旨

　2013 年 12 月 18 日、学位請求論文提出者バーハム・アディナ氏についての最終試験を行った。本試験において、審査委員が提出論文『The Position of Quality Assurance Contributors in Free/Libre Open Source Software Communities』について、逐一疑問点について説明を求めたのに対し、バーハム氏はいずれも十分な説明を与えた。

　以上により、審査委員一同はバーハム・アディナ氏が一橋大学博士（社会学）の学位を授与されるのに必要な研究業績および学力を有するものと認定した。

2014 年 2 月 5 日
審査委員：ジョナサン・ルイス
大坪 俊通
田 良樹
スライマン・ソー