

# Reflection

Research Methods in Software Engineering and Computer Science

Sarah Telford

a1810750

## 1 What was your research project?

My research project was determining if including code segments in StackOverflow questions increases the likelihood of solutions being provided. Throughout code development programmers are going to encounter problems that they cannot solve, however, online discussion forms are a fantastic source to find solutions and ask for advice.

StackOverflow was selected as the focus for this study because it is the most popular programming discussion form and is used by individuals of all skill levels. Given the popularity of StackOverflow amongst the programming community users, it is important that users can utilise this platform to get insightful and useful responses.

Due to the large quantity of questions on StackOverflow, 21 million, I decided to place restrictions on where I collected data. Python was selected as the language for focus in this study. I decided to focus on Python questions because this is one of the most popular programming languages with StackOverflow reporting 44.1% of users program using Python<sup>1</sup>. I believe that using Python as my focused also allowed me to gather a more representative sample of the population. This is because this language is more widely used, whereas more specialised languages may not receive questions from a variety of users.

## 2 What did you learn about doing research as part of your project that you didn't know before?

Through the process of completing this report, I developed several different research skills. I started this project by developing a research methodology that I intended to follow throughout each stage of the project. This methodology outlined what I intended to complete each week, what data I needed to gather, how the data was going to be analysed, and what information and data I would need to collect from other studies. However, I realised that planning this project on a weekly schedule was not the most effective or efficient method for completing the project. This was because I needed to be adaptable and flexible in my project due to other commitments and workloads. Furthermore, I found that trying to stick to a rigid schedule meant that I would be rushing through certain tasks and stages to cross them off my list. After a few weeks of trying to stick to this schedule, I created a more adaptable outline of project stages to allow me to continue to work through the assignment while also enabling flexibility to work on other things.

Undertaking such a large research project was a new experience for me. Previously, research projects I have completed were smaller and did not require as much time or effort. Furthermore, since this project was self-driven, with little input on what needed to be completed, figuring out what to work on next or include in my report was a new experience that took

me some trial and error. However, I found this process interesting, and it taught me a lot about the stages of developing a research paper.

### **3 What would you do differently next time?**

Assuming that I was conducting research in the same field, there are several changes I would make to my research process.

The main change I would make would be to give myself more time in the data collection and analysis stage. I spent a long time looking at past papers and conducting research for my paper I had to rush through the data collection and analysis stage. This is evident by the small sample size collected and the basic analysis of the results. Related to this, I would aim to collect more data and conduct a statistical analysis of the results. More data would ensure that results are representative of the population and statistical analysis would provide better insights into the correlation between data.

Furthermore, I would aim to look at more past research papers. In my report, there were only 3 research papers referenced. The papers were used because they supported my finding and were related to the area that I was researching. By finding more related studies I would be able to better support my claims and data.

### **4 What is your advice to someone who is going to work on a similar project?**

If someone was interested in further developing the research conducted in this study or was interested in looking at the correlation between code segments in questions and the solvability of a StackOverflow post, I would suggest a few different pieces of advice. Firstly, I would advise that they do not have

to look specifically at Python-related questions and instead expand the sample to include any or all programming languages. By lifting the restriction on where the sample is taken from researchers will be able to increase the representativeness of the data and be able to generate more accurate results.

Secondly, I would suggest that future researchers increase the sample size. The sample used in the study was small, meaning that outlier data impacted the results more significantly. However, this issue could be mitigated with the inclusion of more data. Ideally, future researchers would be able to create a program to mine information from StackOverflow. This will allow them to gather large quantities of data quickly. Furthermore, this would remove any human error in the collection of data process.

My final piece of advice would be for researchers to conduct more analysis of the data. The analysis conducted in my paper was basic and did not look at the statistical significance of the data. One method that could be used to test statistical significance would be to generate a T-test. This would inform research about if there is a statistical significance in the mean of the two categories, containing code and not containing code.

Overall, there are several things that future researchers could do to improve this study. However, I believe the most important thing would be to give yourself enough time to complete data collection and analysis.

## **REFERENCES**

- [1] Stack Overflow Developer Survey 2020. (2022). Retrieved 5 April 2022, from <https://insights.stackoverflow.com/survey/2020>