StreetComplete user testing

Introduction

Hi! As part of my thesis, I did a little user testing of StreetComplete a few months ago. The testing was focused on the overall use of StreetComplete, I didn't interfere with the workflow of the participants - I wanted to find out how they use the application and where some of the weak parts of the mapping workflow were.

StreetComplete v56.1 was tested together with StreetMeasure v1.3. I am sorry for the late feedback, but many screens should still be the same. I had 24 participants - sixteen of them in the range of 20–26 years old (mostly university students) and seven of them in the range of 27–64 years old (people in productive age). Further demographic details are attached below (*Fig. 1*). Testing took place in three Czech cities (Brno, Olomouc, and Šlapanice) in the timeframe of 28th February to 1st of April 2024. As a testing device, the OnePlus Nord 2 5G was used.

	woman	man
sex	9	15
	left hand	**right hand**
dominant hand	2	22
	yes	**no**
formal geography education	17	7
	yes	**no**
experience with contributing to OSM	15	9
experience with StreetComplete	5	10
advanced OSM contributor?	6	9

Fig. 1: Demography of participants

The course of testing

Before each testing session, the participant was introduced to the world of OpenStreetMap - what its is purpose, how they can contribute, etc. Nothing about StreetComplete was mentioned. Then, I gave them a mobile phone with StreetComplete in the default state (cleared all the app data/cache). They read through introductory screens in the app and then I gave them two simple tasks:

1. Log in to the app with the following credentials.
2. During the next 10 minutes, map as many tasks as possible.

Participants could move through the city freely, they could fill in only data, that they found in the terrain. During this testing part, participants' voice was recorded for further analysis - they could give me instant feedback on StreetComplete functions they liked or disliked. I was also taking some notes, mostly about participants' behavior in the app (missclicks, etc.). After ten minutes of mapping, the participant filled in a questionnaire.

Results

The evaluation was divided into two parts. Firstly, the questionnaire responses were analyzed, and secondly, the audio recordings were transcribed and evaluated, along with my notes. Here are the main findings.

Questionnaire

First half of the questionnaire consisted of five claims. Participants gave each claim rating based on agreement with the claim.

1. I found the app intuitive to use
2. I found some of the tasks difficult
3. The default answers (values) were understandable
4. The pictures for the tasks were sufficiently instructive
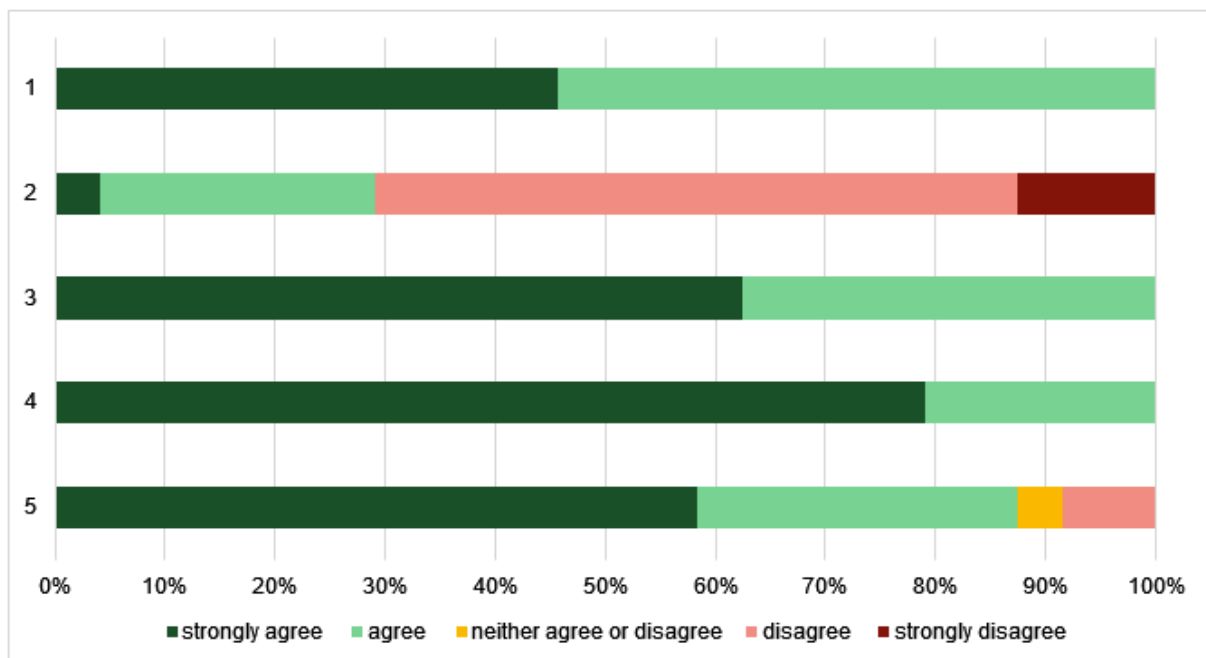5. I liked the base map

Fig. 2: Distribution of responses for five claims

Based on the evaluation, a graph was drawn up showing the proportions of responses for each claim (*Fig. 2*). Overall, participants rated the application very positively. A higher level of dissatisfaction can be observed for claim 2, where almost 30 % of participants tend to agree that some tasks within the application are too difficult. In the case of claim 5, two participants did not like the base map, one participant was undecided.

The second half of the questionnaire contained open-ended questions where participants could express their opinions on the functionality and usability of the application. Based on responses to the first question, participants experienced confusion while using the application due to several issues: inaccurate GPS positioning (6 responses), inability to add uncharted locations to the map (2 responses), inability to display all tasks at once (1 response), and the complexity of some presets (1 response). Other participants either did not fill in the answer or did not mind using the app.

The second question was focused on the tasks themselves, and what the participants liked or disliked about them. The most commonly praised aspects were the simplicity and clarity of the tasks (9 responses), along with their practicality and interestingness (3 responses). On the negative side, the main criticism focused on the complexity of certain tasks (5 responses), such as measuring the width of the road (3 responses).

The third question was focused on the evaluation of the base map. Participants were asked about the level of detail, i.e. whether the map lacked information important for orientation or task performance. For 14 participants the base map was sufficiently detailed for the purposes of the application, for 7 participants the base map was not detailed enough. The most frequent complaints were that there were missing points of interest (3 responses), that the base map did not correspond to reality (1 response), or that there was no option to switch to aerial imagery (1 response). Unfortunately, most respondents did not provided a more descriptive answer.

The fourth question was open for any comments. A majority of positive evaluations of the app further highlighted its interactivity, fun, and potential. On the other hand, some criticized the unfamiliarity of the app among wide public, the inaccurate display of the GPS location, or the lack of tutorials for more complex tasks.

The fifth and final question asked if the respondent would use the app in their free time. The aim was to find out the motivation of the respondents. Sixteen participants answered positively in the questionnaire. Most often they mentioned the usefulness of the data collected (7 responses), they consider this activity as a "useful killing of time". They also found using the app fun (3 responses), interesting (1 response), or suitable for educational purposes (1 response). Two participants also highlighted discovering interesting places they had not noticed before in their neighborhood.

Audio recordings and notes

Participants did not usually have a problem with the app itself, but rather with the tasks they were solving. The first assignment, i.e. logging into the OpenStreetMap application with an account, was completed by all participants without any major problems. Some hesitated in the login process, where they had to confirm the authorization of the application. The second part of the application testing, the assignment "During the next 10 minutes, map as many tasks as possible" was completed by all participants. Everyone completed (uploaded) several tasks.

The most problematic task was "How wide is the road here?", where the contributor is required to fill in the width of the road in meters. Ten participants encountered the task, and half of them did not even begin to solve it after reading the instructions. In total, five participants completed it, one calculated the width of the road from the steps, one estimated the width, and the remaining used the supplementary photo tool. Participants did not have confidence in the photo tool measurements, and two participants failed to complete the measurements. Both beginner and advanced OpenStreetMap contributors had difficulty completing the task. Some of the participants did not notice, that they could use photo tool.

Eight participants had difficulty completing tasks involving editing opening hours. They did not know how to add different opening hours for different days of the week, how to add a lunch break, whether to mark days when closed, or how to edit existing opening hours. However, after a few attempts, most found the correct editing method and managed to fill in the opening hours of businesses and shops correctly. Some

participants chose not to fill in some opening times - either because of the complexity of filling in or the difficulty of finding out opening times (when they were not displayed when entering the business).

Eight participants also had difficulty identifying the type of building. The task "What kind of building is this?" offers numerous choices, which are also grouped into different thematic groups. This made it difficult for participants to quickly find the correct building type as they were forced to click through all the different choice groups offered. About half of the participants abandoned the task because they were not sure of the correctness of their choice or did not find the desired building type in the preferences at all. Situations where the ground floor of the building had ground-floor shops or services and the other floors had offices or residential space also proved problematic - participants did not know which building type to choose. Participants were also oblivious to the fact that only the main building type (group) could be selected, e.g. commercial building - there was no need to select a more detailed building type (e.g. shops), which would have made the decision easier for some. Some participants said they would appreciate the ability to search by building type, as they did for other tasks.

If the contributor is not satisfied with the default values, the "Uh..." menu can be opened for most tasks. This offers e.g. "Can't say", "There is something else", "It does not exist" and other answers adapted to the individual task. However, four participants did not notice this menu at all, and so if they did not know how to respond to a completed task, they simply closed it. The fact that the menu is sometimes too hidden is also mentioned by one advanced contributor. However, if a participant did open the menu, they were often hesitant about which answer to choose. Seven participants struggled with this. Some of the responses were very similar, with participants most often hesitating between "It does not exist" (the object itself) and "Can't say" (the attribute itself). However, this is most likely only a problem with Czech translation, since the English version of the menu is more descriptive.

Negative reactions were caused by the display of the tasks themselves on the map. Five participants complained that it is not possible to see all available tasks at once, as some are only shown as a dot (instead of a bubble that can be clicked on). This does not allow the contributor to plan well the route through the area, and to perform tasks overlaid on other tasks (they cannot be selected). This feature of the app has been criticized by newbies, beginners without much knowledge of OpenStreetMap, and advanced contributors. However, I can see, that in the newest version of the app, this issue is already resolved.

The last more broadly perceived negative feature was the display of task notes. During testing, participants encountered only notes created by other StreetComplete users. It was always an automatically generated note that was created from the "Uh..." menu. Participants did not know how to complete the note tasks. Some also seemed to overlook the question "Can you contribute anything to this note?" and focused directly on the text of the note. Moreover, the text was always half in English and half in Czech. Most participants skipped this type of task. Two performed it by taking a picture of the feature the note was targeting.

Other problems identified were encountered by individuals - elements split into multiple parts confused some beginners (e.g. one road mapped as multiple short lines), the difficulty of adding non-existing elements (e.g. a missing shop), how to correctly fill in the name of a business (if it has more than one name) or which object the task refers to (a poorly described task to find out the height of the curb). Most of these problems are related to the specifics of OpenStreetMap and require more advanced knowledge of OpenStreetMap to perform them correctly.

Some problems also resulted from the lack of detail in the base map - it was not easy to determine the location of the feature to which the task refers. Participants oriented themselves by street names, address locations, or building shapes (e.g. a shop was located in the corner of a building). In some instances, the base map was inaccurate because it was generated from outdated OpenStreetMap data.

Although this chapter contains negative evaluations of parts of the application, on the whole, the positive ones prevailed. Errors from participants' inattention are not mentioned.

Conclusion

Overall satisfaction with the app is high, with participants giving positive ratings to the appearance, functionality, usefulness of the data collected, and enthusiasm for using it. They also mentioned the potential of the app for meaningful spending of their free time or as a complementary activity when walking around the city.

During the testing itself, there were tasks where participants did not know what to do. The application contains only a brief introduction that describes the basic functioning and principles. For individual tasks, more detailed descriptions or hints are mostly missing. In addition, users seem to be overwhelmed by the number of available tasks when they first open the app (the app itself prompts users to turn off some tasks after a while, but only two participants noticed this). Moreover, they often solve the more complex ones first, which may be due to the absence of simpler ones in the mapped area or the priority of displaying individual tasks. To improve the user experience and the quality of entered data, the following functionalities or improvements are suggested:

- Tasks should be tiered according to difficulty. If the user is a beginner (which can be determined from the OpenStreetMap account), only the easiest tasks should be displayed first - especially those that are answered with yes x no or by selecting a few preferences in the form of images. As the number of changes increases, more difficult tasks would be unlocked for the contributor.
- More difficult tasks should have either a tutorial (triggered e.g. after unlocking a set of tasks) or an accessible help that could be opened at any time during the completion of the task (and would contain a detailed description of the problem).
- If there is a choice of multiple values (e.g. determining building kind), a search field should be available.
- The "How wide is the road here?" task should be hidden by default. Its solving is problematic even for advanced contributors and is prone to filling in incorrect values.
- The items in the "Uh..." menu should be more descriptive. As an example, the option "It does not exist..." would be more appropriate whereas "The object does not exist" would be more appropriate. However, this might be problematic only with the Czech translation. The "Uh..." option should be mentioned when the application is first launched, as some participants were unaware of its existence or did not take it into account.

I would like to thank all the contributors for creating such a useful editor. I have been using it since the first public version of the app. I apologize for the late publication of the test results. The original diploma thesis can be found [here](#) (Czech).