|  |  |
| --- | --- |
| The main issue I had with the paper is related to the correlation of structured navigation and actual productivity. I was very confused about that until reading section 6.2.1 (page 7) when the paper referred to the result of Robilard et al. [9]. That helped but still I am not entirely convinced about the generality of this correlation and how strong it is. I would like to understand this correlation better. I think that an early discussion about the principles behind it and an illustration on how you compute a score (based on edit sessions on 5 distinct categories) is of great help. I was also confused about Blaze. From the description from Section 6.2 (paragraph before 6.2.1) I was under the impression that Blaze measures the length of edit sessions. I could not understand how Blaze identifies that one edit session is over. Considering the example from the paper, why "D" is the end and not "E"? Is the execution of a task in a different categories (e.g., build/test) !  the event that defines the end of an edit session? How accurate is this? I was also very curious to understand how to compose these sessions to define a score? Does it make sense to show examples of sectioned bars (/sequences of edit sessions) for two developers in the paper to illustrate how these two developers perform? Is it the case that only the sections that goes to the "structured category" bucket counts? What happens if the bars are very fragmented? |  |
| I am wondering that developers could perform completely different based on the task they receive and based on how they work. I see that authors mentioned this issue in the second paragraph from Threats to Validity (Section 7). I imagine this can be a serious threat to the use of such scoring metric. |  |
| I would like to see a detailed description of the 15 attributes (features) mentioned in Section 6.2.1 |  |
| Five categories mentioned in Section 6.2.1 do not correspond to those mentioned in Figure 8 | In fig – Structured, in sec6.2.1 – other actions |
| I don't find much motivation in the paper that explains why structured search is better than unstructured search. The usage levels of the two are astonishingly different in Fig 8. Are developers really that out of touch with how much time they are spending on this? |  |
| The Releated Work section is very disruptive to the flow of the paper. The Monitoring Practice Studies part appears to be completely irrelevant at this point in the paper - even after reading the paper the work seems marginally relevant. It should either be moved to the end of the paper, or the monitoring practice studies section should just be removed and the Gamification section trimmed down. |  |
| "Results, though inconclusive, show... " -> how can they be both inconclusive and show something? |  |
| I don't understand why you equate a "team goal" to "helping their team in a competition" - was explicitly asked? Otherwise, it seems like an unjustified inference. |  |
| "We conclude that a facility to share..." -> it seems like you should site [16] here. |  |
| I was surprised to see you segment by country and not by age. | Can ignore this comment |
| - "Figure 2 shows over 90% of respondents are comfortable with sharing either anonymous or non-anonymous information with their team members." -> Figure 2 shows this 90% result for "Selected People". Please correct. | there is some confusion on the data representation (color code) |
| Section 5: - who hosts the data? What is done with the data post-survey? | Do we need to address this? |
|  |  |
|  |  |
|  |  |
|  |  |