

Review #813A

Overall merit

2. Weak reject

Paper summary

This paper proposes an approach to curate relevant web page access for developers when they program using an IDE. The information is then saved for later consumption. The authors developed a prototype and evaluated it through two lab user studies, one for link curation and the other for link-supported code changes.

Strengths

- The tool is based on a human-in-the-loop approach.
- The discussion of the related work is clear and contextualized.

Weaknesses

- The user study design is potentially biased. While the discussion is insightful, it only represents one particular scenario.
- The ethical issues are insufficiently discussed.

Comments for authors

Significance and novelty

- The idea of capturing implicit links is appealing. Rather than searching the information related to the source code from scratch, we can use the previous effort devoted to establishing those links. I especially link the human-in-the-loop approach, which is going to guarantee the relevance of the curated link to the developers. At the same time, since the relevance is judged subjectively, it is unclear what is the relevance to future developers.
- The links constitute a new set of artifacts relevant to the source code and related knowledge. Their existence implies additional maintenance costs. How the links are maintained when either the code or the website is updated and who is reasonable for

maintaining those links? Those questions are important to answer before this tool can provide important practical values.

- Both the above two points were discussed in the discussion section after the user study. It is not sufficient in my opinion. Those are important issues that should be carefully considered in the design of the tool in the first place.
- Since SALI monitors developers when they navigate web pages, privacy is a big concern. Without careful consideration and proper handling of this concern, I doubt the developers would be motivated to use this tool. This is not only limited to what information is shared but also what information is recorded (even without sharing) and how secure they are. Since the authors wrote "SALI keeps track of the web page access frequency, recency, and duration.", it seems that serious private information is recorded. What mechanisms were used to guarantee the security of those data and how the privacy practice is communicated with the users?

Soundness

- It is unclear how the TF-IDF score is calculated. What constitutes the web page N? How the nature and size of the collection of web pages are going to impact this measure?
- In Section IV.A, please explain how those tasks are designed. While the authors suggested that the task represents realistic change, no evidence is provided to support this claim. They all have three or four steps which might represent a certain degree of difficulty. The authors later also discussed how they modified the code base so that the tasks can be finished in a similar amount of time. It is unknown how the tool might support code changes for easier or harder tasks.
- The authors didn't include the condition that the users are not instructed to curate the links during the first tasks. The observations discussed in RQ1 are insightful, but only applicable when the participants were clearly instructed to curate the links. Is this a realistic request during their daily coding activities?
- During the inductive coding, the authors wrote: "for themes, two external researchers, who had no prior involvement in the project, generated themes based on the codes obtained in the second step. " Please explain the background of the two external researchers and why they have the expertise to perform this step.
- Please justify the step during which the authors performed the selection of the links from the better solutions in the consumption study. First of all, is it always the case that the better solution would correspond to useful links? Secondly, Who would be responsible for the link quality? This is related to the maintenance concern I mentioned earlier. In practice, the links people created have different qualities. Moreover, in the consumption study, the author collected a set of links from 16 participants. Is this

possible in real life? I would expect only a small portion of code that would be actively worked on by many developers from the team.

- The data from two participants in the consumption tasks were discarded because they didn't use the curated links. I doubt this is a valid decision. They used the same interface as the others and their action represents how people might interact with the tool once deployed. The results of this study, therefore, are biased toward only those who actively use the tool.

Verifiability and Transparency

- How frequently does SALI perform the link recommendation? Are there any methods used to distinguish code change from active editing?

- Please clarify if the study has been approved by the ethical board of the authors' organization.

Presentation

- Most of the presentation is clear. The paper is well organized and I like how the authors presented and discussed their results.

Question for authors

1. What mechanisms were used to guarantee the security of the user data and how the privacy practice is communicated with the users?

2. How the TF-IDF score is calculated. What constitutes the web page N? How the nature and size of the collection of web pages are going to impact this measure?

3. Please explain the background of the two external researchers and why they have the expertise to perform the inductive coding. 4. Please justify the step during which the authors performed the selection of the links from the better solutions in the consumption study.

* * * * *

Overall merit

4. Accept

Paper summary

This paper describes SALI, a tool for helping developers record links to webpages they visit while programming. When previously visited webpages are found to be relevant to ongoing code edits, they are recommended back to developers for them to record. Recording the link saves the link, which is then displayed to subsequent developers working with the source code. To generate the recommendations of relevant webpages to record, a link recommender was designed, using TF/IDF text similarity scores to determine which web pages are relevant to which lines of code. To evaluate SALI, two separate user studies were conducted. 10 developers first used SALI to complete two programming tasks, using SALI to curate links as they worked. A second set of 10 developers then did two further tasks, using the best code solution from the previous study and a randomly selected set of curated links from one of the participants. The results revealed that developers used SALI to curate 4 links per task on average, felt that it was easy to use, had a substantial amount of overlap in the links they chose, and generally chose curations based on a recent web page. In addition, developers with SALI were 2.4 times faster locating the most relevant webpage, but this did not result in significantly improved overall task performance.

Strengths

- * Offers a practical developer tool
- * Comprehensive evaluation of use, encompassing both creating and consuming links
- * Interesting findings about how developers choose to curate and consume links

Weaknesses

- * Results did not demonstrate improvement in task performance

Comments for authors

Novelty

A number of prior systems have offered approaches for helping link artifacts together, including linking code to web context. SALI differs in letting developers control when links are created, where the tool first suggests links to include and developers can then explicitly accept or reject these recommendations.

Significance

Developers make extensive use of web resources, as found in a number of studies including the evaluation. Helping developers better record these web resources for later use might help developers to be more productive. Some aspects of the problem might benefit from additional motivation. Is it really so burdensome to record a link by simply writing a comment? Why embed a link and not some relevant information from the webpage?

But overall the paper answers many of the key questions about its value through the extensive two user studies. In particular, the user study establishes that (1) developers do find links they want to record, (2) that these are usually related to something they just saw, (3) developers without links felt lost, and (4) developers felt it would be particularly useful for onboarding.

The biggest limitation in the motivation is the lack of evidence that the tool ultimately helps improve task performance. While the study demonstrates that curations help developers find a relevant web page considerably faster, this difference ultimately does not translate into a measurable benefit in overall task performance.

Many of the observations about how developers choose to curate and then use curated websites offer insight into developer practice. It was interesting to see that there was a lot of overlap in the curations developers wanted to do and almost exclusively wanted to associate links with lines of code rather than with whole files. And it was interesting that developers are most likely to want to record something that they just saw. However, some of these findings might differ for other tasks or in a real world context.

Given the challenges in ensuring that an external file with links referencing a specific line of code is always updated and kept up to date when any external tool modifies the file, is it really worth it to keep the links separate from the source code? An alternative design might be to embed the links in the source itself, but use an IDE plugin to hide or display these links differently.

Soundness

The tool is effectively designed, drawing on ideas from similar tools in the past but differentiating itself through the focus on interactive curation.

The evaluation is particularly well-conceived, making use of a two phase design to observe both curation and consumption of links. What was the rationale for taking the best study solution and then separately mapping each of the curated link back to this solution? This is artificial, in that participants would normally have the curated links along with the code these were associated with. Was the goal here to control for task progress, so that participants in the second task would all start at the same point?

Verifiability and Transparency

The paper describes the key details of the tool itself as well as the study design. The paper includes a replication package with study materials.
How is the browser history data obtained from the browser without violating browser security restrictions?

Presentation

The paper is easy to follow and well organized. The paper makes effective use of figures in describing key aspects of the technique and study. Fig 2 does a particularly effective job compactly but precisely summarizing the study design. The results section is well organized to highlight key findings.