

CSE504 Homework 1

Haoran Cai ¹

January 11, 2016

¹ Department of Statistics
University of Washington, Seattle, WA

Write and update code documentation

Problem overview

Writing code documentation is important for both software developers and the success of softwares. The document here ¹refers to source code documentation, API documentation as well as algorithm documentation. For developers, it can help original author organize thoughts and contributors work more efficiently. For the software itself, clear and up to date documentation can greatly attract people to use and contribute. However, writing documentation for the software is not the most interesting thing that software developers like to do. Because it takes long time to write in a quality that other people can easily understand. Furthermore, a more challenging question is when the code itself keeps getting updated, the documentation may not. As a result, the software may either lack detailed human understandable documentation or the documentation is out of date. We can even see these issues in very popular open source software projects. When the number of software contributors getting large or the development cycle getting longer, it is very hard to efficiently control the quality of the document and force the documentation up to date. ²

Force developers write documentation

To encourage software developers writing high quality documentation, I propose to use or develop a ³documentation style checking tool every time when the code gets committed into the code base. Because almost every software development requires version control (git, SVN), so this necessary extra checking can make sure only the high quality code can be committed. This documentation checking will for example check if there is ⁴enough documentation in each function, class. Furthermore, it will ⁵evaluate the documentation style. To lower the burden of the developers, this checking ⁶has many levels and the code will be automatically assigned an importance level based on the structure of the whole software, higher level indicates that this part of code needs to write more detailed documentation. ⁷

Summary of Comments on CSE504 Homework 1

Page: 1

1 Number: 1 Author: mernst Subject: Highlight Date: 1/12/2016 6:25:18 AM

Your mentioning many (but not all) types of documentation, but it would be better for you to focus on one particular type. It is unlikely that you will build to solve all problems with all types of documentation, and by thinking about all of them you're depriving yourself of the specifics and details that will enable you to formulate a concrete idea. That is, think about a specific problem and focus on it rather than a general class that contains multiple related but distinct problems.

1 Number: 2 Author: mernst Subject: Highlight Date: 1/12/2016 6:26:07 AM

This paragraph is laying out important problems, but it is a bit wordy. I feel that you can get to the point more quickly rather than using generalities or repetition that don't really provide new evidence for your motivation.

1 Number: 3 Author: mernst Subject: Highlight Date: 1/12/2016 6:27:24 AM

Such tools already exist that force developers to write documentation. How would yours be different? One important issue is getting developers to write quality documentation. A tool can force them to write something but I can't ensure that that something will be helpful to others in the future. In other words, I feel that quality is more important than quantity, and the semantics are more important than the syntax of the documentation.

1 Number: 4 Author: mernst Subject: Highlight Date: 1/12/2016 6:28:09 AM

What is the definition of "enough"? Can you at least describe this in an intuitive way or what you have in mind? Language can make it more formal, but it needs to be concrete rather than vague even now.

1 Number: 5 Author: mernst Subject: Highlight Date: 1/12/2016 6:29:07 AM

What does this mean? Can you give examples of good and bad style that you have in mind? If you're thinking about some specific problem you've encountered in the past, then give those details. That will help readers understand where you're coming from.

1 Number: 6 Author: mernst Subject: Highlight Date: 1/12/2016 6:29:53 AM

What are they? Be specific. Someone could write this text without having ever really thought about the problem, and some reading the text doesn't learn anything about the reader's thoughts.

1 Number: 7 Author: mernst Subject: Highlight Date: 1/12/2016 6:30:39 AM

Overall, I wasn't able to get a specific idea of what you have in mind. This needs to be much more specific, with concrete problems and concrete (even if far-fetched idealistic, and even if you don't know how to implement them) examples of what the tool would output.

Make documentation up to date

To make the large documentation ¹up to date, I propose two ways to solve this issue. One way currently widely used is to make a better visualization of the documentation. For example, Sphinx will automatically extract the code documentation from source code and allow users to convert those information into different formats that easier for developers and other people to view. For example, a popular choice is to convert the documentation into html format, people can read and track all the documentation in a browser, thus it is easier for them to understand the whole structure of the code base. ²Visualization can definitely remind developers to update required and useful documentation, however a trickier problem is that when code base gets larger and larger, developers may even forget what part of documentation needs to be updated. This frustration may even make developers and contributors not willing to write high quality documentation. Combining the tool I mentioned in last section, I purpose to construct a ³"documentation network" that track the dependency of each documentation. When a high quality documentation passes the test and gets successfully committed into code base, we either create a new node to represent this documentation or update the node if this documentation has been previously committed. The edges in this graph ⁴obviously represents the dependency of documentations. For example, in the API documentation, if one class inherits another class, then these two chunk of documentation (nodes) may have an edge to connect them. When the code base gets growing, this network will also automatically expand. When developer wants to commit a code or documentation, the tool will automatically remind him what other parts of documentation ⁵needed to be updated by querying the "documentation network".

1 Number: 1 Author: mernst Subject: Highlight Date: 1/12/2016 6:31:22 AM

Please start up a defining what the problem is. You have given a specific difficulty that a programmer faces. Before you jump into a solution, you need to make the specific problem clear.

1 Number: 2 Author: mernst Subject: Highlight Date: 1/12/2016 6:33:31 AM

I think of "visualization" as being about different graphical representations. Just changing the formatting from plaintext HTML does not seem like a very interesting transformation. What is the specific, important problem that this transformation is solving? Also, how is this transformation related to problems such as forgetting to update, which the latter part of this sentence mentions? I'm not sure how the ideas in the sentence or in this paragraph hang together.

1 Number: 3 Author: mernst Subject: Highlight Date: 1/12/2016 6:34:14 AM

What is this? It's confusing to readers if you make up a term but don't give any examples of definitions. Show a picture, and explain a little scenario of how developer uses it to solve the developers problems.

1 Number: 4 Author: mernst Subject: Highlight Date: 1/12/2016 6:34:31 AM

This isn't obvious to me. I don't even know what the nodes are.

1 Number: 5 Author: mernst Subject: Highlight Date: 1/12/2016 6:35:36 AM

Was it necessary to query the documentation at work? If the developer has edited procedure foo, then it's pretty straightforward (and purely syntactic check) to tell the developer to update the documentation for procedure foo. Maybe you can be more specific about the problem.