Assignment 1

Huan Sun

Apart from those already covered in the assignment, other difficulties I often encounter during software development are:

- When I pose a query, too much information can be found on the Web, and it takes a lot of time to finally get what I really want. There are two kinds of queries I often pose in a search engine during software development: (1) Queries on how to do something, e.g., "How to traverse a dictionary in C#" and "How to make/implement/install X" ; (2) Queries for debugging, e.g., "assertion failure in Y" and a Java exception. Some queries are easily to resolve by looking at the first few pages returned by Google, say "How to traverse a dictionary in C#". But very often, especially when dealing with some unexpected exceptions and failures, I feel hard to quickly get useful information to solve my problem. For example, discussion forums such as StackOverflow are usually what one resorts to during this process. For one query, one might get many responses, each of which might also have their own comments and follow-ups. I think in these cases, a summarization tool in discussion forums which can present succinct answers to a user query might be of great help. A good summarization tool might need to decide what information the user is asking, differentiate the key aspects in responses such as "Why the user is getting this error " (reason) and "What the user needs to do" (action) and list them separately. Therefore, I think more research efforts could be devoted to developing such tools.

- It is hard to read and understand code snippets written by other people, especially those not properly commented. In a large component, while it is easy to understand each instruction (e.g., by checking API documentation or usage of basic operators), one can lose the big picture on what the entire code chunk is doing. It might be too demanding or time consuming for programmers to write all the comments and annotations manually. I am thinking automatically annotating a code snippet might be useful: Based on the description for each instruction in the code snippet, generate a high-level description about what the code is doing. More exciting would be if a tool can automatically segment a long code snippet based on the functions (or, what they try to achieve) and then annotate each segmentation. Such annotations could also help search for such snippets in the future. I need to investigate the state-of-art research results in this aspect, though.

- Proactively detecting some defects in the code can be difficult. Without any exceptions or error reports, it is hard even to realize there is something wrong in the code before sufficient testing with

many inputs and desired outputs. I am wondering whether there are some analysis tools that can automatically detect "abnormal" usages of variables or give some warnings when using a method. For example, there is some discussion about the error related to "Object.Equals" in StackOverflow (`http://stackoverflow.com/questions/2219047/net-dictionaries-have-same-keys-and-values-but-arent-equal/2219052\#2219052`). Actually in the method usage covered by MSDN tutorials, there are remarks specially dealt with the error (`https://msdn.microsoft.com/en-us/library/bsc2ak47.aspx`). With that said, if a tool can automatically recognize there are pitfalls in using a certain method (by referring to a tutorial or documentation of that method), it might be good to give some warning to the programmers. This might save programmers a lot of time, compared with tracing down what might be going wrong later.